Spatiotemporal and Temporal Forecasting of Ambient Air Pollution Levels through Data-intensive Hybrid Artificial Neural Network Models

PhD Thesis

Sheen Mclean S. Cabaneros

Department of Mechanical and Aerospace Engineering

University of Strathclyde, Glasgow

October 27, 2020

Jointly-authored Publications

The following outputs were produced during the course of this PhD programme:

- Cabaneros, S.M., Calautit, J.K., and Hughes, B.R. (2020), "Short- and Long-term Forecasting of Ambient Air Pollution Levels using Wavelet-based Non-linear Autoregressive Artificial Neural Networks with Exogenous Inputs," *International Journal of Environmental Impacts: Management, Mitigation and Recovery.* **3(2)**:143-54.

- Cabaneros, S.M., Calautit, J.K., and Hughes, B.R. (2020), "Spatial Estimation of Outdoor NO_2 Levels in Central London Using Deep Neural Networks and a Wavelet Decomposition Technique," *Ecological Modelling,* **424**:109017.

- Cabaneros, S.M., Calautit, J.K., and Hughes, B.R. (2019), "A Review of Artificial Neural Network Models for Ambient Air Pollution Prediction," *Environmental Modelling & Software,* **119**:285–304.

- Cabaneros, S.M., Calautit, J.K., and Hughes, B.R. (2017), "Hybrid Artificial Neural Network Models for Prediction of Urban Roadside PM_{10}, $PM_{2.5}$ and NO_2 Pollution 24 to 48 Hours in Advance," *Book of Abstracts - 12th Conference on Sustainable Development of Energy, Water and Environment Systems*, Dubrovnik, Croatia.

- Cabaneros, S.M., Calautit, J.K., and Hughes, B.R. (2017), "Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO_2 Pollution," *Energy Procedia,* **142**:3524–30.

Acknowledgements

Special thanks to my viva examiners, Dr. Nicolas Kelly of Strathclyde University, Glasgow, and Dr. Fan Wang of Herriot-Watt University, Edinburgh, for their valuable comments and constructive remarks on this thesis. I am also grateful to Dr. John Kaiser Calautit of the University of Nottingham for his inputs and guidance during the early stages of my doctoral study.

I must thank Dr. Alice G. Comahig, my undergraduate thesis adviser, for first introducing me to the interconnected world of ANNs. My first baby, my thesis utilising a neural network known as ADALINE, would be so proud of seeing her younger and more sophisticated siblings 11 years later.

I could not have completed this thesis without the guidance of my supervisor, Prof. Ben Hughes. His personal anecdotes and unfiltered words of advice have helped me survive my entire PhD journey and encouraged me whenever I face article rejections. Thanks also to my colleagues, Victoria, Tom, and Diana, and good mates, Bowy, Bia and Daniel, without whom my PhD life would have been much less fun and eventful. Finally, I am extremely grateful to my parents for their support and encouragement all these years.

Abstract

Outdoor air pollution (AP) is a serious public threat which has been linked to severe respiratory and cardiovascular illnesses, and premature deaths especially among those residing in highly urbanised cities. As such, there is a need to develop early-warning and risk management tools to alleviate its effects. The main objective of this research is to develop AP forecasting models based on Artificial Neural Networks (ANNs) according to an identified model-building protocol from existing related works. Plain, hybrid and ensemble ANN model architectures were developed to estimate the temporal and spatiotemporal variability of hourly NO_2 levels in several locations in the Greater London area. Wavelet decomposition was integrated with Multilayer Perceptron (MLP) and Long Short-term Memory (LSTM) models to address the issue of high variability of AP data and improve the estimation of peak AP levels. Block-splitting and cross-validation procedures have been adapted to validate the models based on Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Willmott's index of agreement (IA). The results of the proposed models present better performance than those from the benchmark models. For instance, the proposed wavelet-based hybrid approach provided 39.15% and 28.58% reductions in RMSE and MAE indices, respectively, on the performance of the benchmark MLP model results for the temporal forecasting of NO_2 levels. The same approach reduced the RMSE and MAE indices of the benchmark LSTM model results by 12.45% and 20.08%, respectively, for the spatiotemporal estimation of NO_2 levels in one site at Central London. The proposed hybrid deep learning approach offers great potential to be operational in providing air pollution forecasts in areas without a reliable database. The model-building protocol adapted in this thesis can also be applied to studies using measurements from other sites.

Acronyms and Nomenclature

List of Acronyms

AAR	Average Recall Rates (%)
AARE	Average Absolute Relative Error
AI	Artificial Intelligence
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
AOD	Aerosol Optical Depth
AP	Air Pollution
AURN	Automatic Urban and Rural Monitoring Network
BIC	Bayesian Information Criterion
BP	Back-Propagation
BP	Barometric Pressure ($\bar{m}$)
BRP	Binary Recursive Partitioning
CA	Correlation Analysis
CART	Classification and Regression Trees
CMAQ	Community Multiscale Air Quality

Acronyms and Nomenclature

CNN	Convolutional Neural Network
COPD	Chronic Obstructive Pulmonary Disease
CV	Cross-Validation
DA	Data Augmentation
DEFRA	Department for Environment, Food and Rural Affairs
DWT	Discrete Wavelet Transformation
EEMD	Ensemble Empirical Mode Decomposition
ELM	Extreme Learning Machine
EU	European Union
FC	Fully-Connected ANN model
FDMS	Filter Dynamics Measurement System
FL	Fuzzy Logic
GA	Genetic Algorithm
GRNN	General Regression Neural Network
IA	Index of Agreement or Willmott's Index
ICA	Imperialist Competitive Algorithm
IDW	Inverse Distance Weighting
IEEE	Institute of Electrical and Electronics Engineers
LIN	Linear Interpolation
LM	Levenberg-Marquardt
LME	Linear Mixed Effects

Acronyms and Nomenclature

LR	Learning Regression
LSTM	Long Short-term Memory units
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MSAD	Mean Sum of Absolute Deviations
MSE	Mean Squared Error
NARX-ANN	Non-linear Autoregressive ANN model with Exogenous Inputs
NCA	Neighbourhood Component Analysis
NF	Neuro-Fuzzy
NMBE	Normalized Mean Bias Error
NN	Nearest Neighbour
NRMSE	Normalised Root Mean Squared Error
OECD	Organisation for Economic Co-operation and Development
PCA	Principal Component Analysis
PSO	Particle Swarm Optimisation
R	Pearson correlation coefficient
RBFN	Radial Basis Function Network
RBM	Restricted Boltzmann Machine
RCP	Royal College of Physicians
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network

Acronyms and Nomenclature

SAE	Stacked Autoencoder
SFS	Sequential Forward Selection
SOM	Self-organizing Map
SSE	Sum of Squared Error
SVM	Support Vector Machine
T	Temperature (°C)
TEOM	Tapered Element Oscillating Microbalance
TfL	Transport for London
TSAD	Total Sum of Absolute Deviations
TSP	Total Suspended Particles
UAM	Urban Airshed Model
WD	Wind Direction (°)
WHO	World Health Organization
WNN	Ward Neural Network
WRF/Chem	Weather Research and Forecasting Model with Chemistry
WS	Wind Speed (m/s)
WT	Wavelet Transformation

Nomenclature

$\hat{y}$	Estimated value of the actual predictand value Y
λ	Regularisation parameter for all weights when implementing the NCA technique

Acronyms and Nomenclature

$\mathbf{b}$	Vector containing all bias factors of nodes in either the hidden or output layer of an ANN
$\mathbf{w}$	Weighting vector of the nodes in either the hidden or output layer of an ANN
$\mathbf{w}^*$	Vector containing both network weights and biases
$\mathbf{x}_p$	Vector containing a specific predictor
$\mathrm{Ref}\,(x)$	Reference point for x
$\tilde{F}$	Approximation of the error function F
b_0	Intercept or the sample mean of the actual predictand value Y
b_i	Coefficient of the selected predictor $\mathbf{x}_i$
C_i	Region produced by child node i which does not overlap with another region
$d_w\,(\mathbf{x}_i, \mathbf{x}_j)$	Weighted distance between two sample vectors $\mathbf{x}_i$ and $\mathbf{x}_j$
E	Network error
e	Euler's number
F	Error function
f	Transfer function
L	Loss function that measures the disagreement between the predicted value of the randomised regression model
l_o	Output layer
P	Total number of predictors
Q	Total number of network training samples
T	Set of training samples $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$

Acronyms and Nomenclature

T^{-i} Set T without the point $(\mathbf{x}_i, \mathbf{y}_i)$

w_p Weight assigned to the p-th predictor

$w_{ji}^{(l)}$ Weighting factor from node i from layer $(l-1)$ to node j of layer l

Y Actual value of the predictand or target variable

BLM   London Bloomsbury Road site

CAM   Camden Kerbside site

FAR   The fraction of false alarms over the total forecast positive events

FN   Number of episode days incorrectly identified as non-episode days

FP   Number of non-episode days incorrectly identified as an episode

HAM   Tower Hamlets Roadside site

KEN   North Kensington site

MAR   London Marylebone Road site

POD   The fraction of correctly forecast episode days

TN   Number of non-episode days correctly identified

TP   Number of episode days correctly identified

WST   London Westminster site

Subscripts

$-i$ The index of a set without the point $(\mathbf{x}_i, \mathbf{y}_i)$

$_{bg}$ Predictors collected from a background air quality monitoring site

ji Information path from predictor value i to the jth node in the succeeding node of a hidden layer

Acronyms and Nomenclature

kj	Information path from hidden node value j to the kth node in the succeeding node of an output layer
max	Maximum value of a given predictor
min	Minimum value of a given predictor
$norm$	Normalised value of a given predictor
ALL	MLP model utilising all set of predictors
MET	MLP model utilising a set of meteorological predictors
NCA	MLP model utilising a set of predictors selected by the NCA technique
POL	MLP model utilising a set of emission predictors
RT	MLP model utilising a set of predictors selected by the RT technique
SFS	MLP model utilising a set of predictors selected by SFS technique
SR	MLP model utilising a set of predictors selected by the SR technique
TIME	MLP model utilising a set of temporal-type predictors

Contents

Contents

Contents

xix

Contents

Contents

List of Figures

List of Figures

xxv

List of Figures

List of Figures

List of Figures

List of Tables

Chapter 1

Introduction

Research on the development of air quality models continues to rapidly progress as outdoor air pollution (AP) remains the biggest environmental risk to public health (Oliveri Conti et al., 2017). For instance, 4.2 million premature mortalities worldwide per year were linked to outdoor AP the majority of which were due to ischaemic heart diseases and strokes (see Figure 1.1a). Furthermore, the number of respiratory diseases among children and adults and premature deaths due to outdoor AP is projected to skyrocket at substantial levels from 2010 to 2060 (see Figure 1.1b). Although outdoor air quality has improved in many areas especially in high-income countries over the past decades (Fann et al., 2009; Crippa et al., 2016), peak AP levels continue to escalate at an alarming rate affecting economies and public health due to rapid industrialisation.

(a)   (b)
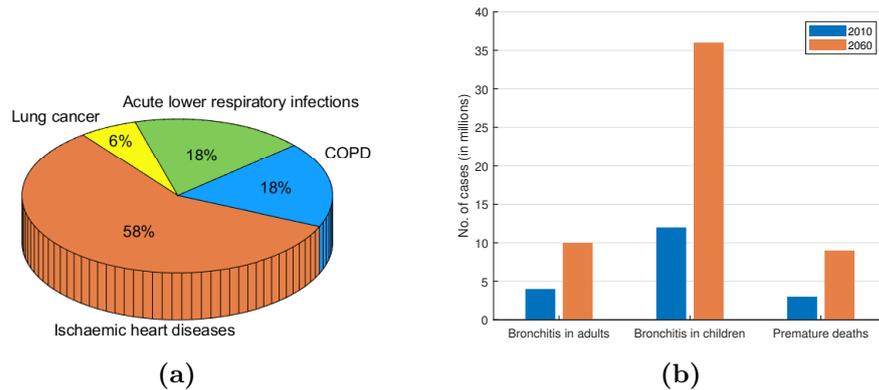
Fig. 1.1. (a) Breakdown of outdoor AP-related premature death causes per year (WHO, 2016), and (b) Projected number of cases (in millions) of AP-related respiratory diseases and premature deaths from 2010 to 2060 (OECD, 2016).

Chapter 1. Introduction

As such, various approaches have been implemented by policy-makers and the scientific community to manage and mitigate the adverse effects of outdoor AP levels especially on public health. Firstly, abatement measures have been put in place to ensure that concentration levels of pollutants such as ozone (O_3), fine particles (PM_{10} and $PM_{2.5}$), and nitrogen dioxide (NO_2), fall within acceptable legislative standards (Baklanov et al., 2007; Gers et al., 2001). For instance, European Union (EU) member states are mandated to assess their AP levels especially in densely-populated areas via measurements, modelling and other empirical techniques and report the collected data to the European Commission for compliance (Nagl et al., 2016). On the other hand, the modelling of outdoor AP levels has been employed to assist urban planners in providing well-informed decisions that reduce the impacts of outdoor AP (Baklanov et al., 2007; Chen et al., 2018; McLaren and Williams, 2015). For instance, the restriction on driving and certain industries has been implemented during the predicted periods of peak pollution episodes in Santiago, Chile (Gers et al., 2001). Such an action was reported to minimise $PM_{2.5}$ concentration levels by 20% and reduced around 8 deaths per day. Rapid alerts derived from AP forecasting models also act as early-warning systems which can reduce the health and economic burdens of peak AP episodes in urbanised locations (McLaren and Williams, 2015; Kelly et al., 2012). In detail, the public especially those with underlying respiratory conditions can enforce personal mitigation measures against outdoor AP. For instance, daily pollution forecasts in Canada have been found to help decrease the number of asthma-related cases by 25% (Chen et al., 2018). This has lead to the increasing demand for intensive research efforts to help improve the performance of existing AP modelling systems.

However, AP modelling is a non-trivial task, and the difficulty of such task primarily depends on the specific purpose of the AP model itself. Colls (2001) summarises the different purposes of AP models as follows:

- to determine which sources are responsible for what proportion of AP concentration at any location where a measuring device is placed;

- to estimate population exposure on a higher spatial or temporal resolution than is practicable by measurement;

- to target emission reductions on the highest contributors; and

- to forecast and predict AP concentration variations over time.

In the context of building AP forecasting models, modellers have faced limitations due to the complex, non-linear, and poorly understood dynamics that exist between outdoor air pollutants and many parameters in the environment, e.g. meteorology, traffic, etc. (Colls, 2001; Gardner and Dorling, 1998; Sportisse, 2010). Colls (2001) describes a perfect AP model as "one that is able to predict the spatial and temporal variations of AP concentrations to sufficient accuracy for all practical reasons, deeming other measurements unnecessary." Since then, various outdoor AP forecasting approaches attempting to satisfy the conditions of the above-mentioned definition have already been developed. They are fundamentally categorised as either *physics-based* or *statistical* approaches.

1.1 Physics-based Approaches

Physics-based approaches model air quality through a detailed understanding of the physical, chemical and fluid dynamical processes in the atmosphere (Colls, 2001; Gardner, 1999). They require parameters that describe the formation, generation and dispersion of air pollutants in the ambient environment In order to operate (Jacobson, 1997). The main advantage of implementing physics-based models is their ability provide insights describing the mechanisms between AP and meteorological variables in the atmosphere.

Popular examples of physics-based models include the Urban Airshed Model (UAM) (Chang and Cardelino, 2000), Weather Research and Forecasting Model with Chemistry (WRF/Chem) (Chuang et al., 2011), Community Multiscale Air Quality (CMAQ) model (Mueller and Mallard, 2011) and ensemble systems (Ganev et al., 2008).

However, physics-based models are limited in various aspects. Firstly, they rely on highly-simplified assumptions concerning air pollutant behaviours, e.g. constant source emissions and homogeneous atmospheric conditions over space and time (Gardner, 1999; Lagzi et al., 2013). This can restrict the usefulness and accuracy of model out-

puts especially in locations with highly dynamic conditions of variables associated with outdoor AP levels (Lu and Xue, 2015; Luecken et al., 2006).

Physics-based models also yield results that do not share the spatial resolution of real-world observations from monitoring sites (Colls, 2001; Chave and Levin, 2003; National Research Council, 2007). In more detail, the dynamics describing the dispersion of air pollutants is usually complex that modellers resort to producing forecasts on coarse grids. As such, the resulting forecasts can be useless in most cases when utilised to explain the outdoor AP conditions in local situations (Valput et al., 2019). Furthermore, such results cannot be automatically compared to conventional observations of real-world pollutant levels.

Physics-based models also require several types parameters which are difficult to obtain and usually estimated by modellers based on experience leading to reduced accuracy of results (Xu et al., 2017; De Ridder et al., 2012).

Finally, the implementation of physics-based models can be computationally expensive and intractable (Beelen et al., 2009; Dutot et al., 2007; Fernando et al., 2012; Gocheva-Ilieva et al., 2014).

1.2 Statistical Approaches

In contrast, statistical models estimate the underlying dynamics between a pollutant behaviour directly from measured data. That is, data on AP concentration levels at a given site alongside explanatory data that might affect them are used to estimate their future values. Hence, statistical models do not share the difficulties exhibited by physics-based models. Many of the traditional statistical AP models include regression methods, time series analysis, classification and cluster analysis and principal component analysis (Gardner, 1999; Shahraiyni and Sodoudi, 2016). Given the complex non-linear and ill-defined dynamics between outdoor air pollutants and environmental parameters, the performance of traditional linear statistical models is often limited (Gardner and Dorling, 1998; Shahraiyni and Sodoudi, 2016). In detail, most statistical models assume linearity between AP concentration levels and explanatory variables making them unable to capture the complex non-linear relationship among

different variables in the ambient environment.

Alternatively, statistical models that incorporate AI-based techniques have gained massive popularity in recent years as they are able to circumvent the limitations of traditional statistical models (Gardner and Dorling, 1998; Shahraiyni and Sodoudi, 2016; Abderrahim et al., 2016; Chen et al., 2008). Such a shift of interest can also be due to the availability of more powerful computing systems that are capable of performing more sophisticated algorithms (IEEE Spectrum, 2018). Popular examples include Artificial Neural Networks (ANNs) (McCulloch and Pitts, 1943), Support Vector Machines (SVMs) (Corani, 2005), Fuzzy Logic (FL) (Zadeh, 1988) and hybrid AI-based models. More specifically, research activity involving the use of ANNs in AP forecasting applications has skyrocketed over past the 19 years alone (Cabaneros et al., 2019). ANNs have successfully been employed to address the issues encountered by its linear statistical and physics-based counterparts (Gardner and Dorling, 1998; Shahraiyni and Sodoudi, 2016). Hence, this research solely focuses on the development of effective AP forecasting models based on ANNs.

1.3   Artificial Neural Network Models

ANNs are essentially simple abstractions of biological neural networks. The basic principle of an artificial neuron was first conceptualised by Warren McCulloch and Walter Pitts in their paper *"A Logical Calculus of the Ideas Immanent in Nervous Activity"* (McCulloch and Pitts, 1943).

Figure 1.2 illustrates the similarities between a biological and an artificial neuron. The tree-like nerve receptors called *dendrites* collect information from neighbouring neurons and carry electrical signals into the *cell body*. The cell body then processes the incoming signals by summing and mapping them. The single long fiber called an *axon* then carries the resulting signal from the cell body to the dendrites of other neurons. The point of contact between an axon of one neuron and a dendrite of another neuron is called the synapse. Synaptic weights determine the strength of the signals being transferred from one neuron to another. The adjustable network parameters are formally called the *weights* and *biases*, while the non-linear mapping is commonly called

as the *transfer function.*

(a)   **(b)**
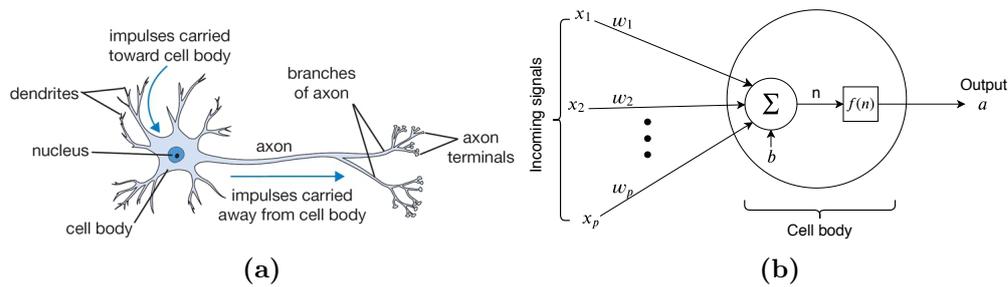
Fig. 1.2. (a) A cartoon drawing of a biological neuron (Durak, 2020), and (b) its mathematical model.

In comparison, an artificial neuron resembles its biological counterpart in two respects: 1) their building blocks are both highly interconnected, and 2) the connections between neurons dictate the function of the network (Hagan et al., 1995).

ANNs are popularly employed in a wide range of applications due to the following advantages:

- ANNs can approximate any mapping between a set of input variables or *predictors* to a set of target variable(s) or *predictands*. They also operate without any assumption regarding the statistical distribution of the data being used and instead rely on the data being used to "train" them.

- ANNs can model non-linear relationships, setting them apart from many traditional statistical models which assume linearity between predictors and predictands (Hornik et al., 1989). For instance, ANNs have been found to be suitable tools in the atmospheric sciences as the dynamics between numerous variables in the ambient environment is usually non-linear (Colls, 2001; Gardner and Dorling, 1998).

- ANNs can adaptively learn through network training, e.g. calibration of its weights and bias parameters, just as a biological neuron understands patterns over time through the modification of their synaptic strengths (see Chapter 3 Section 3.2.5).

Just like any other modelling scheme, ANN models also encounter limitations. Firstly, there is no clean-cut approach in developing them. One can only provide a set of

guidelines that help modellers build their own ANN models. Developing ANN models generally comprises several steps, e.g. division of data, data preprocessing, determination of model inputs and network architecture, training, validation (Maier and Dandy, 2000). The implementation of each step requires the selection of several settings that usually involve a combination of *a priori* knowledge, an ad-hoc approach and optimisation schemes. Consequently, this lack of one-size-fits-all approach creates more free parameters than any conventional physics-based or statistical models, thus creating more uncertainty among modellers. Maier and Dandy (2000) once argued that the lack of the comprehensive guide makes it difficult for future modellers to draw meaningful comparisons between existing ANN models. Gardner (1999) even pointed out that the unguided use of ANN models will not automatically guarantee superior model results.

Furthermore, ANN models are case-specific. In the context of AP modelling, ANN models, are site-specific, e.g. they can be only employed in areas where the data used to train them were collected. Thus, the application of AP forecasting models based on ANNs is limited in areas with the lack or absence of monitoring stations due to defective instruments, budget cuts, etc. In such cases, modellers usually resort to spatial interpolation techniques in which available measurements from other monitoring stations are utilised to model AP concentration levels at unmonitored stations. While ANN models have been employed to estimate spatiotemporal variations of AP levels more recently, this is currently a growing sub-field within the ANN modelling literature (Cabaneros et al., 2019). Most existing works also employ plain ANN models which can limit the overall performance of their developed models (Alimissis et al., 2018; Pfeiffer et al., 2009; Tzanis et al., 2019).

Finally, ANN models have been found to have difficulties when dealing with extreme AP concentration levels (Catalano et al., 2016; Gong and Ordieres-Meré, 2016). This issue is attributed to the limited continuous observations of extreme pollutant levels leading to fewer representative training data, e.g. *the imbalance data problem*, and the highly variable concentration levels at a local scale. To address the said issue, a number of studies dealing with hybrid or ensemble forms of ANN models has grown rapidly in the past five years alone (Cabaneros et al., 2019). One popular class of

hybrid models is known as data-intensive in which raw AP time series are initially decomposed into several sub-series in order to extract important features, and make the forecasting scheme more accurate. For instance, decomposition techniques based on wavelets have been shown to improve the performance of plain ANN models (Cabaneros et al., 2019; Siwek and Osowski, 2012; Osowski and Garanty, 2007). The use of more sophisticated forms of ANNs, including the Non-linear Autoregressive ANN model with Exogenous Inputs (NARX-ANN) and the deep learning techniques such as the Long Short-term Memory (LSTM) models, has been also been shown to outperform standard ANN models (Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019). On the other hand, the development of ensemble systems wherein the results of multiple models are integrated into one final output has been recently proposed in the literature (Siwek and Osowski, 2012; Di et al., 2019; Araujo et al., 2020). Ensemble modelling has been argued to take advantage of the strengths of the individual models. Simultaneously, it compensates for the weaknesses of the individual models. However, results pertaining to the effectiveness of both hybrid and ensemble approaches are still limited (Cabaneros et al., 2019).

1.4   Research Aim and Objectives

Therefore, the primary aim of this research is to develop temporal and spatiotemporal models based on state-of-the-art ANN models in forecasting outdoor AP levels. In detail, this work attempts to achieve the following objectives:

1. conduct a critical review of existing works and state-of-the-art ANN models to identify gaps in the literature;

2. develop AP forecasting models based on ANNs according to the identified ANN model-building protocol from the literature review;

3. build and evaluate wavelet-based data-intensive hybrid ANN models for the temporal and spatiotemporal forecasting of outdoor AP levels, e.g. extrapolation in the temporal and spatiotemporal domain, respectively, across several monitoring sites in Central London; and

4. build and evaluate data-intensive ensemble forms of ANN models in the spatiotemporal forecasting of outdoor AP levels across the Greater London area.

This work investigates the use of popular feedforward ANN model, e.g. the Multilayer Perceptron (MLP), and a deep recurrent ANN models, e.g. LSTM model, and their hybrid forms to model outdoor AP. The selection of these models was based upon the perceived benefits that the techniques offered when compared to traditional statistical and other forms of ANN models.

1.5   Project Methodology

This research will perform a thorough review of related literature from published peer-reviewed articles dealing with ANN models for outdoor AP forecasting. Research gaps will then be identified to justify the objectives of this thesis. Three general experiments will then be conducted to achieve the said objectives. The said experiments are as follows:

Experiment 1: Temporal modelling of hourly NO_2 levels in London Marylebone Road using ANN models and various feature selection techniques;

Experiment 2: Temporal and spatiotemporal modelling of hourly NO_2 levels in Central London with wavelet-based ANN models;

Experiment 3: Spatiotemporal modelling of hourly NO_2 levels in Greater London using wavelet-based fully-connected LSTM models;

The model development and parameter configurations of each experiment are described in detail in Chapters 3 and 4.

1.6   Thesis Structure

To encompass the overall aim and objectives, this thesis is structured as follows:

Chapter 1 provides a background on the use of ANN models in ambient AP forecasting. This chapter also provides the main aims and objectives of the research.

Chapter 1. Introduction

Chapter 2 presents the results of the undertaken critical review of the application of ANN models in forecasting outdoor AP levels. This chapter also provides a brief discussion of the stages in building ANN models and the taxonomy of options available for each step as guides in assessing existing works in the literature.

Chapter 3 discusses the underlying theories regarding the methods applied in each of the experiments carried out. Most importantly, the underlying mechanisms of feed-forward ANN models are described in more detail in this chapter.

Chapter 4 outlines the experimental settings adapted to performing the experiments. The said settings include details regarding the collected data and case study locations, and model and software parameter values utilised.

Chapter 5 compares and analyses the results obtained from the experiments. This chapter particularly makes critical comparisons between plain, hybrid and ensemble forms of ANN-based models developed in this research.

Chapter 6 draws conclusions based on the analysis of the experimental results presented in the preceding chapters and identifies the components of this research and current state of AP modelling using ANN models that need to be addressed by future researchers.

Chapter 2

Literature Review

2.1 Introduction

The objectives of this chapter are twofold: (1) to provide a critical review of the application of ANN models in forecasting outdoor AP levels, and (2) to identify current research gaps in the literature. A total of 139 papers have been selected and reviewed in this chapter (see Appendix A for a full list of the identified papers for review). Details regarding the methods and scopes applied in the selection process can be found in Cabaneros et al. (2019).

The papers are reviewed in terms of the following components: (a) basic model settings, (b) model development protocol, and (c) model performance. Basic settings such as the distribution of papers according to the year of publication, model time-steps, forecasting horizon, and utilised air pollutant parameters are discussed in Chapter 2 Section 2.2.1. The general protocol in building ANN models are described in Chapter 2 Section 2.2.2. The identified papers are then assessed in terms of model performance in Chapter 2 Section 2.2.3. A summary of the key findings of this review is found in Chapter 2 Section 2.3. Finally, the identified research gaps according to the aforementioned results are presented in Chapter 2 Section 2.4.

2.2 AP Forecasting using ANN models

2.2.1 Overview of identified modeling settings

The distribution of articles by year of publication is given in Figure 2.1. There is a growing number of published articles since year 2000 that deal with AP forecasting using ANN models. Almost half of the identified papers were published since 2015 alone. This can be well explained by the availability of more computing environments tailored for running algorithms supporting ANN models (IEEE Spectrum, 2018). Furthermore, large datasets of good quality are frequently collected due to the installation of more monitoring sites in many cities, especially those regulated by air quality directives (Date, 2018; Jinran, 2017; Upadhay, 2019).
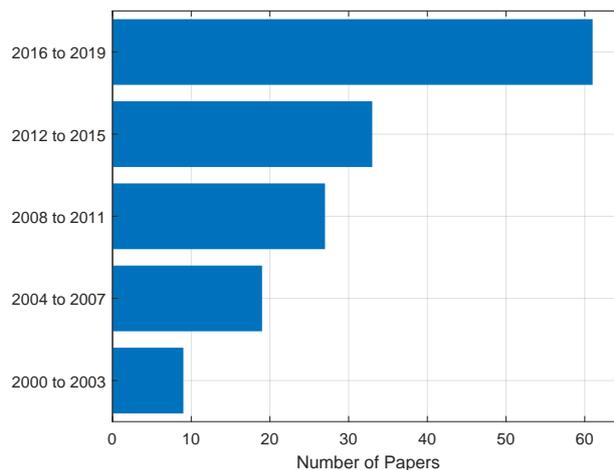
Fig. 2.1. Distribution of papers by year of publication.

The number of occurrences in which various air pollutant variables were examined is shown in Figure 2.2. Airborne particulate matter with a diameter smaller than 10 μm (PM$_{10}$) and 2.5 μm (PM$_{2.5}$), oxides of nitrogen (NO, NO$_x$, and NO$_2$) and ozone (O$_3$) are the most selected variables for estimation by the identified papers. In more detail, particulate matter was examined 89 times, almost 50% of which dealt with PM$_{10}$ prediction, while almost 45% of which for PM$_{2.5}$, and the remainder for fine dust and total suspended particles (TSP). The results above highlight the growing number

of studies that investigate the reduction of the disease burden resulting from PM_{10} and $PM_{2.5}$ exposure using early warning mechanisms such as ANN models (RCP, 2016). On the other hand, oxides of nitrogen and O_3 were examined by 53 and 45 papers, respectively. The forecasting of carbon monoxide (CO) and sulphur dioxide (SO_2) were carried out by 18 and 23 papers, respectively. It is also worth noting that a third of the identified papers examined more than one air pollutant species.
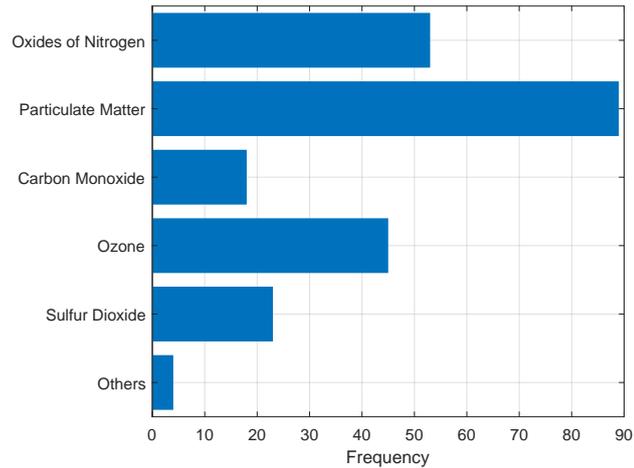
Fig. 2.2. Number of occurrences various pollutants were predicted.

The number of occurrences various time-steps have been utilised is shown in Figure 2.3. The hourly time step was most commonly utilised (71 times), followed by daily (51 times), yearly (5 times) and 5-minute steps (4 times). Several time steps including 2-hourly, 4-hourly, 1-minute and 30-minute were also used. The time steps of the selected variables are primarily determined by the sampling periods of the instruments used to measure pollutant species and meteorological data at monitoring stations. For instance, the predominant use of hourly time steps is influenced by the guidelines set by most legislative directives, and the availability of reliable measured data (DEFRA, 2004). In some cases, however, authors pre-process their data via averaging and linear interpolation techniques to create model datasets with consistent time steps.
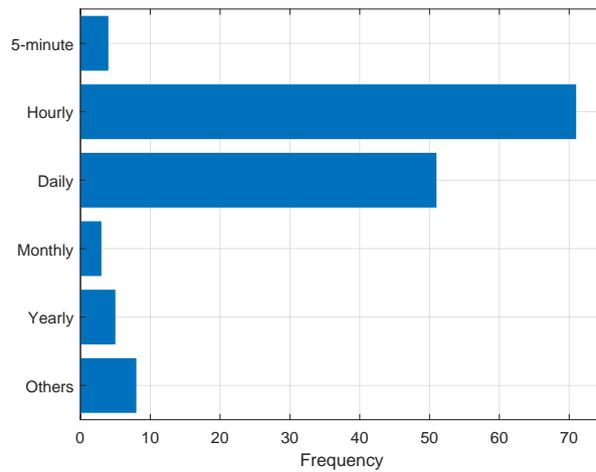
Fig. 2.3. Number of occurrences various time steps have been utilised.

Figure 2.4 shows the number of times various forecasting time steps have been used by the identified papers. Only 105 (64%) of the reviewed papers explicitly described the forecast length they used. Of these papers, short-term forecasting (forecast length $= 1$) was carried out 73 times. Long-term forecasting (forecast length > 1) was done 94 times. In more detail, $+ 24$ h, $+ 48$ h and $+ 72$ h forecasts were carried out 26, 7, and 4 times, respectively. This result highlights the need for early-warning mechanisms that can provide longer range forecasts. Finally, prediction (forecast length of 0) was carried out by 5 papers.
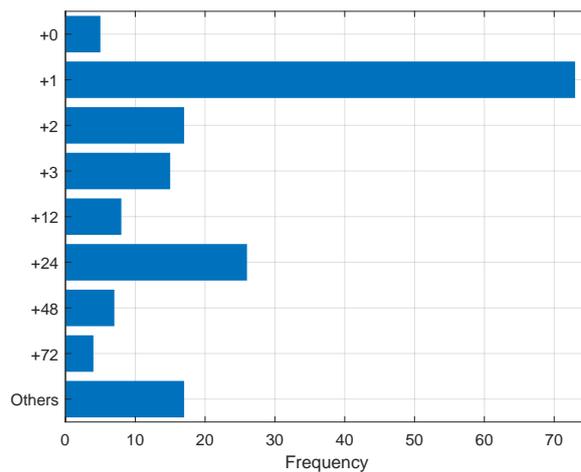
Fig. 2.4. Number of occurrences various forecasting horizons have been utilised.

2.2.2   Model development protocol

Building an ANN model generally requires eight main steps, namely: (1) data collection, (2) data preprocessing, (3) selection of predictors, (4) data splitting, (5) selection of model architecture, (6) determination of model structure, (7) model training, and (8) model performance validation (Cabaneros et al., 2019; Maier and Dandy, 2000). The said steps, alongside the manner the collected data move through them, and their corresponding outputs are presented in Figure 2.5. Each step, as well as their corresponding taxonomy of options, is presented in the following sections. The identified papers are then assessed according to the options presented above.

2.2.2.1   Data Collection

ANN models are data-dependent which means their performance primarily relies on the following factors: (a) size and (b) type of data utilised to train them.

In terms of data size, one important requirement is for the dataset to span the full range of input space for which the network will be trained to approximate (Hagan et al., 1995). In AP forecasting applications, the use of predictors covering a period of a year or more has been highly recommended to ensure that seasonal factors that strongly influence AP levels are taken into account (Colls, 2001; Arhami et al., 2013; Kumar et al., 2017).

The number of occurrences various lengths of input data have been collected is shown in Figure 2.6. All but six studies provided details regarding input data lengths. The majority of the papers utilised input data with lengths covering more than a period of one year. Datasets with a duration from one to three years were used 60 times, while those with lengths longer than three years 43 times. The use of datasets covering a period of less than 6 months occurred 22 times, while only 8 studies used data with lengths between six months to one year.

**Outcomes/
Decisions**

**Steps in Model
Development
Process**

Flow of Data

Available Data

Number of
Outputs

1. Data Collection
(inputs and outputs)

Selected Raw Data

2. Data Pre-processing

Selected Processed
Data

Number of
Predictors

3. Predictor Selection

Model Development
Data

4. Data Splitting

Model
Architecture

5. Model Architecture
Selection

Training
Data

Testing
Data

Validation
Data

6-7. Model Structure and
Learning Parameters
Calibration

Model
Structure
Selection

Performance
evaluation

Trained
Model

Model
Calibration

Validated
Model

8. Model Performance
Evaluation

Fig. 2.5. Basic stages in building an ANN model (Maier et al., 2010).
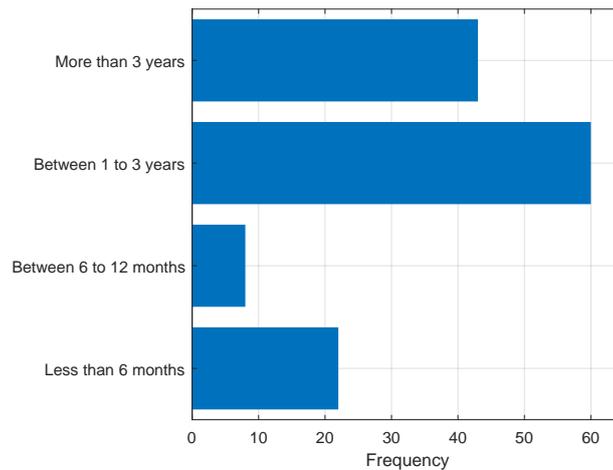
Fig. 2.6. Number of occurrences various data lengths have been collected.

The selection of various predictor types also affects the performance of ANN models as air quality is a complex function of meteorology, emissions and other parameters (Colls, 2001). In this research, predictors are categorised as either *meteorological*, *emissions*, *traffic*, or *others* for simplicity. Meteorological predictors refer to parameters influenced by the turbulence within the atmosphere. Variables such as wind speed, wind direction, relative humidity and atmospheric turbulence have been found to have a massive influence on the dispersion and concentration of several air pollutants such as O_3, NO_2, PM_{10} and $PM_{2.5}$ (Colls, 2001; Kumar et al., 2017; Dominick et al., 2012; Peng et al., 2017). Emissions predictors refer to primary and secondary air pollutant species in the ambient environment. They are also considered important as they are highly correlated to other air pollutants (Colls, 2001; WHO, 2018). Traffic predictors refer to parameters that characterise traffic behaviour. These include traffic flow density, speed, occupancy degree, queues length, and travel time which are typically monitored on roads in close proximity to air quality stations.

As shown in Figure 2.7, the use of both meteorological and pollutant emissions variables was observed 92 times. The removal of both meteorological and emissions predictors has been found to cause negative effects on model performance (Al–Dabbous et al., 2017; Juhos et al., 2009; Rahimi, 2017). On the other hand, the utilisation of only meteorological predictors was carried out 9 times, while the sole use of emissions

data 14 times. Additionally, the use of traffic data alongside other variables occured in 18 papers. The use of data based on satellite-derived imagery also appeared in 6 papers. For instance, the utilisation of satellite-derived Aerosol Optical Depth (AOD) variables was explored in the forecasting of $PM_{2.5}$ on several occasions (Mao et al., 2017; Wen et al., 2019; Yeganeh et al., 2017). Other predictors such as land-use, economic and stability predictors were used by only a few identified papers, limited mainly to spatial or spatiotemporal forecasting of AP levels (Chen et al., 2008; Yeganeh et al., 2017; Antanasijević et al., 2018; Alam and McNabola, 2015; Yeganeh et al., 2018).
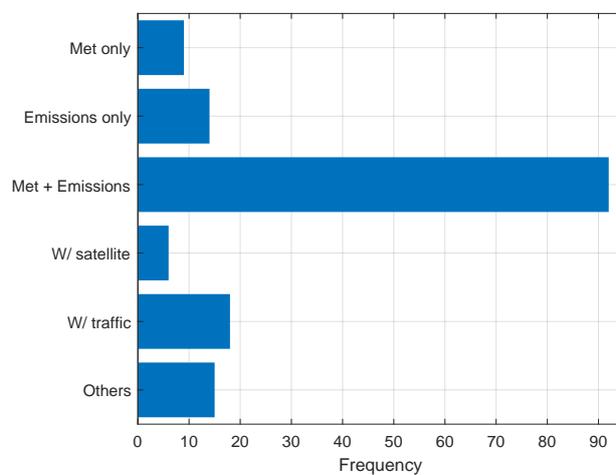
Fig. 2.7. Number of occurrences various sets of predictors have been utilised.

2.2.2.2 Data Preprocessing

Data preprocessing refers to the preliminary techniques that improve the representation of predictors. Two popular data preprocessing techniques in the field of AP modelling are *data normalisation* or *missing data imputation*.

Normalisation is applied to make sure that predictor values fall within a similar range. This process is important as inputs with large values disproportionately mask the impact of those with smaller ones (Hagan et al., 1995). Normalisation techniques are either categorised as *range scaling* or *standardisation* techniques (see Figure 2.8). Under range scaling, predictor values are being mapped to a range of values between 0 and 1 or -1 to 1, inclusive. In contrast, standardisation converts an old variable into a

new variable with zero mean and unit standard deviation.
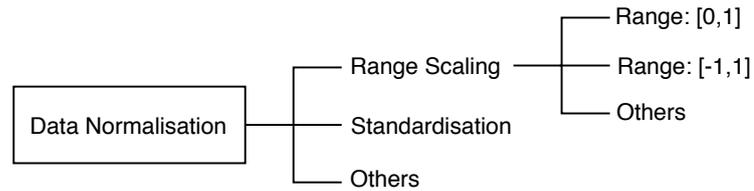
Fig. 2.8. Taxonomy of data normalisation techniques.

The majority of the studies identified (76 papers) did not provide sufficient details describing the methods utilised for data normalisation. Of those that did, 43 papers employed the standard normalisation scheme. There are 10 occasions where the input data were adjusted to have zero mean and unity variance. However, there are only two occasions where other normalisation methods were used. For instance, Osowski and Garanty (2007) transformed all their data using reverse Mallat filtering, while Zhu et al. (2018) divided each sample by the maximum value of the entire dataset for each predictor.
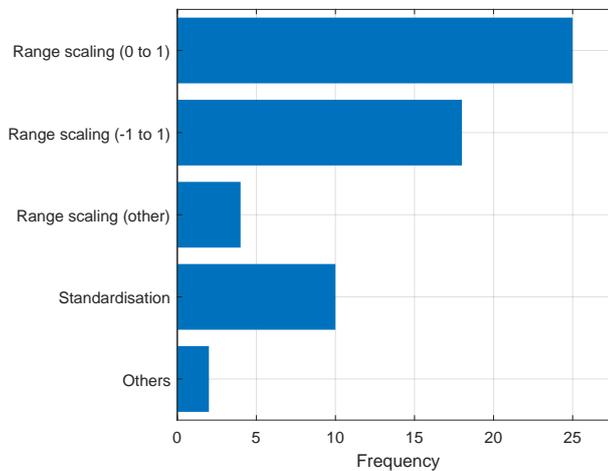
Fig. 2.9. Number of occurrences various missing data normalisation techniques have been implemented.

Missing data imputation addresses the issue of missing data, which is a common problem in air quality forecasting (Junninen et al., 2004). Missing data is caused by many factors such as insufficient sampling, errors in measurements or faults in data

acquisition (Junninen et al., 2004). Over the years, several imputation techniques have been developed (Junninen et al., 2004; Pigott, 2001; Plaia and Bondı, 2006). A taxonomy of missing data imputation approaches is given in Figure 2.10.
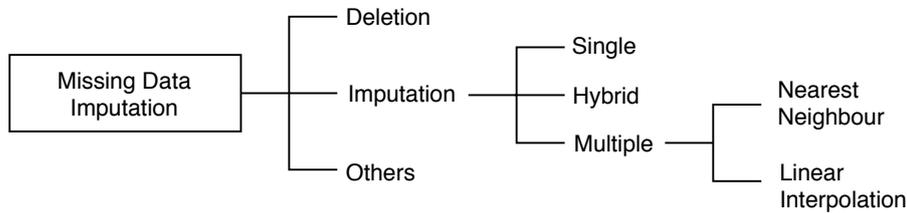
Fig. 2.10. Taxonomy of missing data imputation procedures.

One popular missing data imputation approach is the list-wise or pair-wise deletion of predictors with missing data. Another one is the substitution of missing values with the mean of the entire dataset. However, the said practices are highly discouraged as they can disrupt the inherent structure of the original dataset, thus potentially degrading the performance of a model (Junninen et al., 2004). Other missing data imputation techniques such as univariate, multivariate, nearest neighbour, and the hybrid of the previous approaches are also available (Junninen et al., 2004; Plaia and Bondı, 2006). Univariate methods include linear (LIN), spline and nearest neighbour (NN) interpolation. Multivariate methods include regression-based imputation, nearest neighbour interpolation, Self-organizing Map (SOM) and MLP models. A more detailed discussion of the said techniques can be found in Junninen et al. (2004) and Plaia and Bondı (2006).

Figure 2.11 shows the number of occurrences various missing data imputation approaches has been undertaken. Only 34 papers provided details regarding missing data. The deletion of predictors with missing data was the most predominant step undertaken (16 times). Under univariate methods, the NN interpolation was carried out 5 times, whereas linear interpolation was only used 4 times. There are 6 instances in which the combination of multiple methods was implemented to address specific gap lengths.
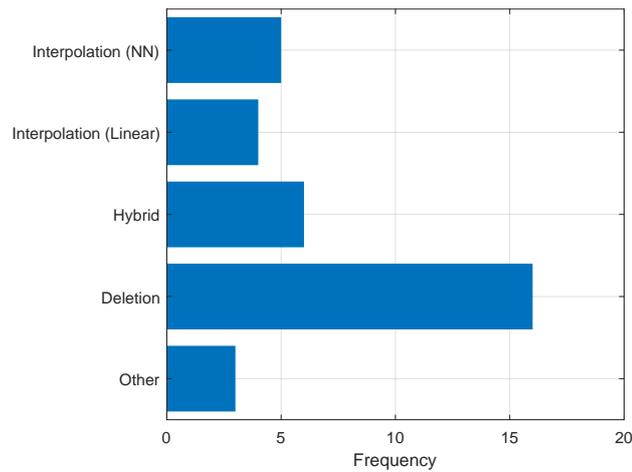
Fig. 2.11. Number of occurrences various missing data imputation techniques have been implemented.

2.2.2.3 Predictor Selection

The performance of ANN models is highly dependent on model predictors (Hagan et al., 1995; Bishop, 1995). However, the inclusion of too many predictors results in more network connections leading to overfitting issues (Hagan et al., 1995). Conversely, the absence of relevant predictors inhibits the model from correctly approximating the underlying dynamics between predictors and predictands (Maier and Dandy, 2000). As such, several approaches that determine the most significant predictors have been proposed and tested. They are categorised as either *model-free* or *model-based* (Maier et al., 2010), see Figure 2.12.
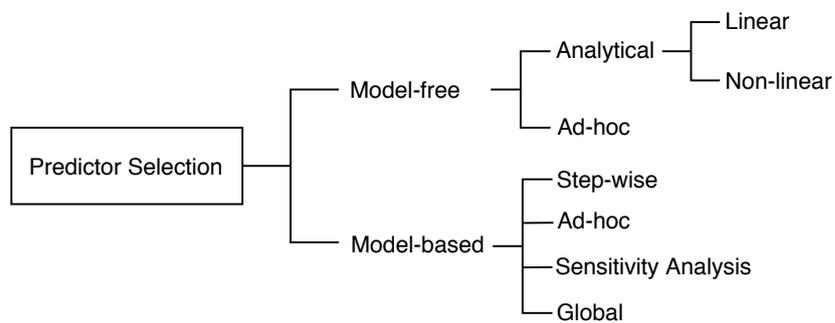
Fig. 2.12. Taxonomy of approaches to selection of optimal model predictors.

Model-free approaches select predictors without relying on the performance of the developed ANN models. In detail, such approaches carry out the selection before an ANN model is trained. Model-free approaches can further be classified as either *ad-hoc* or *analytical*. The selection of predictors in an arbitrary manner or based on domain knowledge falls under the ad-hoc approach. In contrast, the analytic approach employs a statistical measure of dependence between predictors and predictands. This is mostly carried out through Correlation Analysis (CA). On the other hand, model-based approaches operate by determining the effect of a candidate predictor or set of predictors on the overall model performance. One popular example is the stepwise selection, where a network iteratively selects, e.g. *forward selection*, or removes, e.g. *backward elimination*, a candidate predictor. An ad-hoc approach can also be carried out in which arbitrary combinations of model predictors are tested. Global approaches employ global optimisation algorithms to select the combination of predictors that maximises model performance. Finally, an approach based on sensitivity analysis examines the plots of the influence of each predictor to a predictand or set of predictands.

The number of occurrences various predictor selection approaches have been applied is shown in Figure 2.13.
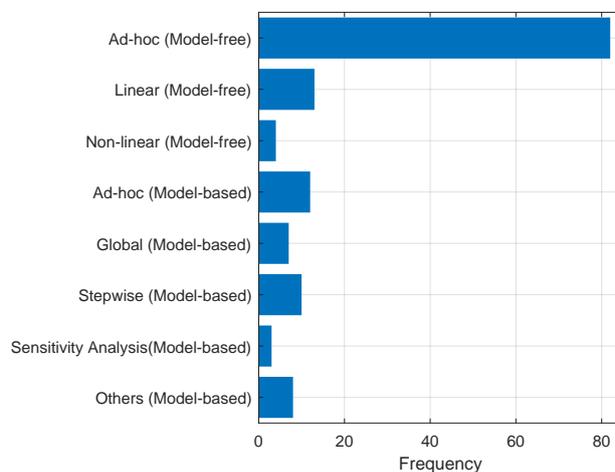
Fig. 2.13. Number of times various predictor selection techniques have been used.

Model-free approaches were implemented 95 times, while model-based approaches were employed 40 times. There are several factors that might explain the predominant

use of the former over the latter. The implementation of the model-based approach is time-consuming as ANN models also need to be trained in the process. The approach might also not clearly identify the effects of sets of predictors on model performance, as the latter is also a function of other model components such as model structure and calibration settings. Of the model-free approaches, ad-hoc methods were most widely implemented occurring in 82 papers. Linear approaches, especially correlation analysis, were utilised in 13 papers, while non-linear methods only 4 times. Stepwise techniques were applied 10 times, while the use of global search approaches 7 times.

2.2.2.4 Data Splitting

Data division is carried out by splitting the available data into three subsets, namely, the *training*, *validation* and *test* sets. The training subset is used for computing the gradient and adjusting the network weights and biases (see Chapter 2 Section 2.2.2.7). On the other hand, the validation subset is utilised to stop the network training before *model overfitting* occurs. Model overfitting is a scenario in which the network memorises the data in the training subset making it unable to generalise new situations or unforeseen data (Hagan et al., 1995). The error in the validation subset monitors the network performance during training. When this error begins to increase for several iterations, the training is stopped. The weights and bias values that yielded the minimum error are then used as the final trained network weights and biases. Lastly, the testing subset is used to determine the generalisation ability of the developed model.

Data splitting approaches can be categorised as either *supervised* or *unsupervised* approaches (see Figure 2.14). Supervised approaches split the collected dataset into three subsets while accounting for the statistical properties of each subset. On the other hand, unsupervised approaches do not take the statistical properties of the subsets into account explicitly. Only stratified unsupervised approaches ensure that the statistical properties of the subsets are similar (Maier et al., 2010). Random unsupervised approaches split data in a random manner. The v-fold cross-validation randomly divides the dataset into v independent subsets, of which $v - 1$ subsets are chosen for training while remaining data is used for testing. The process is repeated several times

until a criterion is met. Physics-based approaches split data into different classes according to knowledge about the underlying physical processes. Ad-hoc approaches split data in an ad-hoc manner. One popular example is known as block splitting which allocates the first N observations for the training set, and the next group of observations for the validation and testing sets.
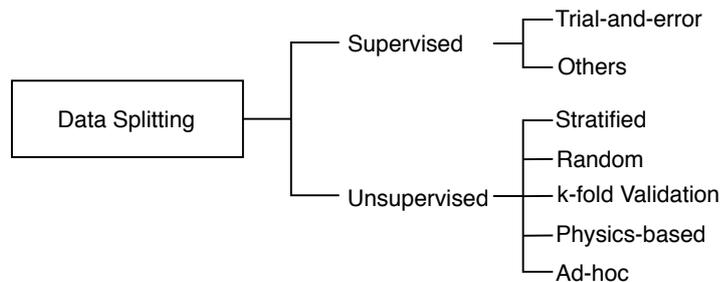
Fig. 2.14. Taxonomy of data splitting techniques.

Only a few identified papers did not discuss the process of data division explicitly. Of those that did, only unsupervised methods were implemented (see Figure 2.15). In more detail, the ad-hoc method was implemented 79 times, while random data division methods were carried out 40 times. The v-fold cross-validation method was implemented only 8 times. Finally, only a small number of papers employed unsupervised stratified techniques.
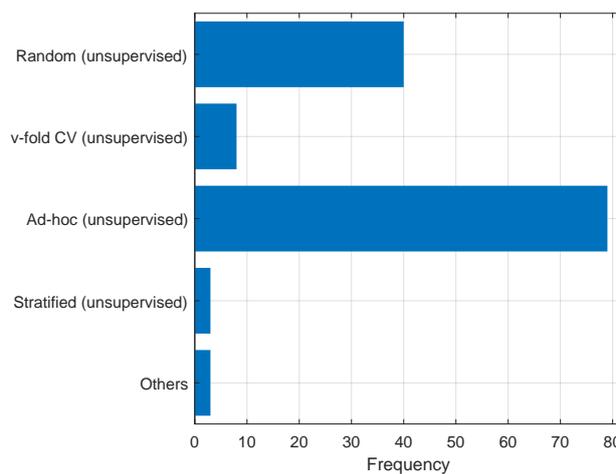
Fig. 2.15. Number of occurrences various unsupervised data splitting methods have been implemented.

2.2.2.5 Selection of Model Architecture

Model architecture refers to how information flows from one layer to the subsequent layers in a given network. The taxonomy of model architectures is shown in Figure 2.16.
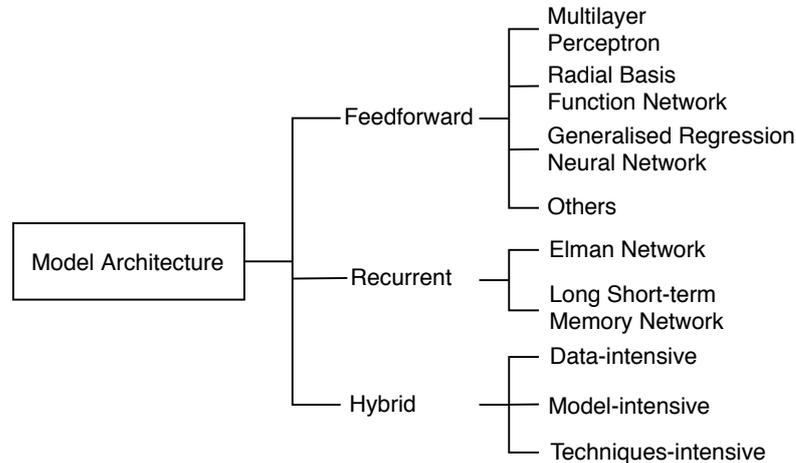
Fig. 2.16. Taxonomy of model architectures.

Two of the most popular network architectures for prediction and function approximation are the feedforward and recurrent networks (Hagan et al., 1995). In a feedforward network, information moves from the input layer through the succeeding layers in a single direction. The Multilayer Perceptron (MLP) is a very popular feedforward ANN model for the non-linear approximation of AP variables (Gardner and Dorling, 1998; Cabaneros et al., 2019). Other examples of feedforward ANNs include Radial Basis Function Networks (RBFNs), General Regression Neural Networks (GRNNs), Ward Neural Networks (WNNs) and Extreme Learning Machine (ELM). In contrast, Recurrent Neural Networks (RNNs) allow feedback in which some output nodes are connected to the nodes of the preceding layers. One popular example of RNNs is the Elman network, while recent sophisticated examples of RNNs include LSTM networks. RNNs have been argued to be potentially more effective than feedforward ANNs because of their feedback mechanism which improves their capacity to learn (Hagan et al., 1995; Samarasinghe, 2006).

The application of hybrid ANN models has also been highlighted in recent years (Makridakis et al., 2018). The hybrid modelling approach has been argued to capitalise

on the strengths and overcomes the weaknesses of the individual models involved (Chen et al., 2008; Shahraiyni et al., 2015; Sharma et al., 2005). In this work, hybrid models are categorised as either *data-intensive*, *model-intensive* or *technique-intensive* (Maier et al., 2010). Data-intensive approaches initially classify input data with respect to various dynamics dependent on the problem specifications or the criteria set by the modeller. Separate models are then developed for the identified separate classes. Popular examples include the use of ANNs and techniques such as Principal Component Analysis (PCA), k-means clustering, Ensemble Empirical Mode Decomposition (EEMD) and Wavelet Transformation (WT). The model-intensive approach employs different models for different sub-components of the overall physical system and then aggregates the various responses calculated from different models. Popular examples include Fuzzy-neuro networks, e.g. the hybrid of a feedforward ANN and fuzzy systems, multiple Restricted Boltzmann Machine (RBM) layers and a Back-propagation (BP) layer, and LSTM and Convolutional Neural Networks (CNNs). Lastly, a technique-intensive approach is one that combines an ANN model with a different technique to develop an ensemble approach exploiting the advantages offered by different techniques. Examples include the use of MLP with SVM, and Stacked Auto-encoders (SAE) with a Learning Regression (LR) layer.

Figure 2.17 shows the number of times various model architectures have been used by the identified papers. MLP models were found to be the most commonly used model architecture, implemented in 73 papers. The number of studies in which alternative network architectures were employed was reasonably uniform, ranging from 4 to 13. Other feedforward ANN models were also applied, including the GRNN (Antanasijević et al., 2018, 2013; Singh et al., 2012; Stamenković et al., 2017) and RBFN (Ibarra-Berastegi et al., 2008; Ordieres et al., 2005; Zhou et al., 2014; Zito et al., 2008). Plain recurrent models such as the Elman network (Brunelli et al., 2007; Biancofiore et al., 2015, 2017; Chelani et al., 2002) and LSTM models (Li et al., 2017; Freeman et al., 2018; Wang et al., 2018) were also chosen.
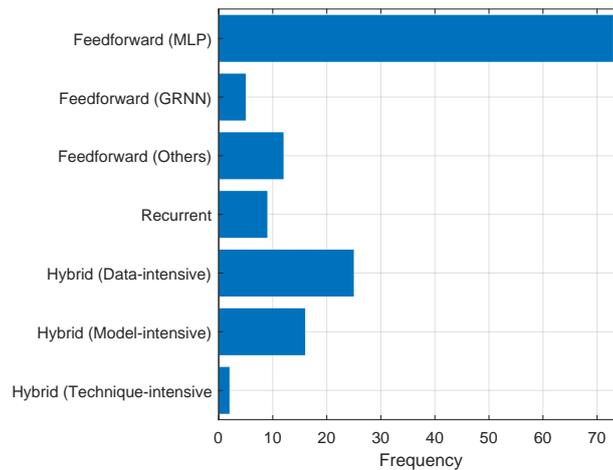
Fig. 2.17. Number of occurrences various model architectures have been used.

The majority of the implemented hybrid models are data-intensive (25 papers). One of the most popular data-intensive approaches involves the use of wavelets. The prediction of several subseries with lower variability has been shown to yield more accurate estimation results than predicting the original time series of higher variability (Siwek and Osowski, 2012; Osowski and Garanty, 2007; Dunea et al., 2015; Prakash et al., 2011). Other examples include the use of EMMD and GRNN model (Zhou et al., 2014), Imperialist Competitive Algorithm (ICA), EMMD and BPNN model (Jiang et al., 2018), and ICA, EMMD and ELM model (Li and Zhu, 2018). Several model-intensive approaches were also applied in 18 occasions. Examples include the use of CNN and LSTM models (Wen et al., 2019; Huang and Kuo, 2018; Pak et al., 2018; Qin et al., 2019), Fully-connected LSTM models (Zhao et al., 2019), and Neuro-Fuzzy (NF) models (Yeganeh et al., 2017, 2018; Dursun and Taylan, 2015; Jain and Khare, 2010; Taylan, 2017).

Network architecture also refers to the way information is mapped from the input nodes through the nodes of the succeeding layers. The said mapping is influenced by the choice of transfer function. Furthermore, the superimposition of different transfer functions determines the ability of an ANN to approximate different input-output dynamics (Gardner and Dorling, 1998; Hagan et al., 1995; Bishop, 1995). In fact, feedforward ANNs with linear transfer functions and without hidden nodes are equivalent

to linear statistical models (Bishop, 1995). Transfer functions are selected based on the nature of the task and the layer in which they are to be utilised. In AP forecasting, *sigmoidal* functions are commonly used in the hidden layer as they are non-linear and easily differentiable (Gardner and Dorling, 1998). Popular sigmoid functions include the *logistic* and *hyperbolic tangent* functions. *Linear* or *identity function* is commonly used in the output layer (see Chapter 3 Section 1.3).

Almost 30% of the identified papers did not provide details concerning their use of transfer functions. Among those that did provide information, the logistic sigmoid function was predominantly used in the hidden layer nodes, to be followed by hyperbolic tangent function. In the output nodes, identity function was widely utilised, followed by logistic sigmoid.

2.2.2.6 Determination of Model Structure

Model structure refers to the number of layers and the number of nodes in each layer. Determining the number of nodes in both input and output layers is fairly easy as they correspond to the number of predictors and predictands, respectively. However, finding the optimal number of hidden layers and nodes is a non-trivial task as the relationship between predictors and predictands is captured in the hidden layer (Hagan et al., 1995). Furthermore, the hidden layer structure directly corresponds to the number of network weights which directly influences model complexity.

Hidden layer parameters are case-specific as they are dependent on the data complexity of a specific application (Gardner and Dorling, 1998). Unfortunately, a general method for model structure determination still remains unknown (Cabaneros et al., 2019). As a result, different approaches have been employed to address the said uncertainty.

The methods for determining the optimal ANN model structure can be classified either as *global*, *stepwise trial-and-error* or *ad-hoc* (see Figure 2.18). Global approaches employ global methods based on competitive evolution found in nature, e.g. Genetic Algorithm (GA), Particle Swarm Optimisation (PSO), simulated annealing, etc. Using this approach, it is possible to simultaneously optimize the network weights and biases,

and the number of hidden layers and nodes. If implemented properly, global methods are likely to yield the best ANN structure and parameters. However, they are found to be computationally expensive (Maier and Dandy, 2000).
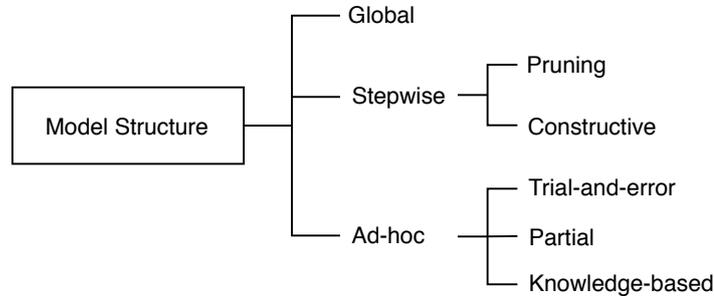
Fig. 2.18. Taxonomy of model structure determination procedures.

Stepwise trial-and-error approaches can also be used, in which a basic ANN structure is first assumed, which is modified with each trial with the objective of achieving a structure that is neither too complex nor too simple. They can further be categorised into two types, one based on pruning algorithms and the other on constructive approaches. Lastly, ad-hoc approaches determine the optimal model structure without adhering to strict pruning and constructive techniques. One ad-hoc approach is based entirely on trial-and-error approach. The partial ad-hoc approach is based on the use of both a trial-and-error approach and an empirical formula that provides upper and/or lower bounds of the number of hidden nodes. Finally, the knowledge-based ad-hoc approach is based on the experience or intuition of the modeller.

The number of occurrences various model structure optimisation techniques has been used is shown in Figure 2.19. The ad-hoc approach was by far the most popular, with 132 instances. It is worth noting that 15 identified papers implemented both a trial-and-error and empirical rules. Of the structured approaches, constructive stepwise approaches were implemented 4 times, while global approaches only three times.
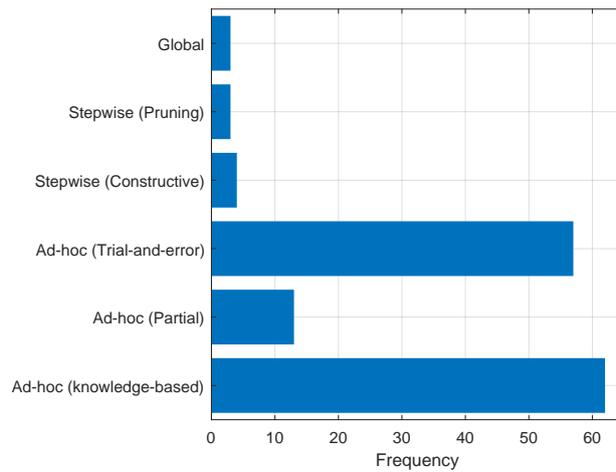
Fig. 2.19. Number of occurrences various model structure optimisation techniques have been used.

2.2.2.7 Calibration of Model Parameters

Training an ANN is the process of calibrating the connection network weights to improve the performance of the resulting model. Weights and biases are usually initialised randomly (Bishop, 1995). The network is repeatedly presented with the desired response for each input pattern as the network weights and biases are calibrated until the target outcome, e.g. the acceptable difference between the desired and actual output, is met. Network calibration methods are classified as either *local* or *global* approaches (see Figure 2.20).
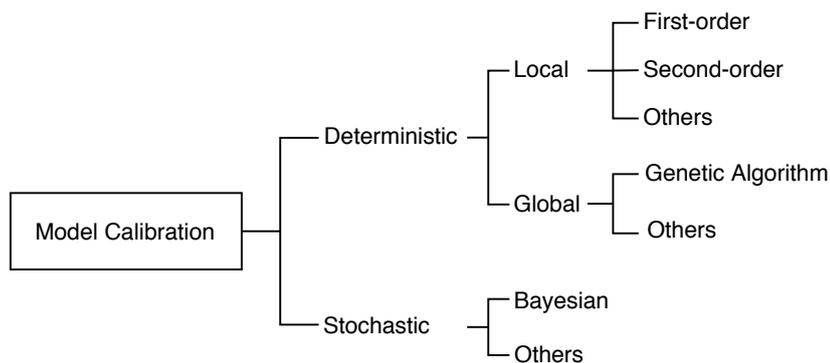
Fig. 2.20. Taxonomy of model calibration techniques.

Local methods usually work on gradient information and are therefore prone to becoming trapped in local optima if the error surface is reasonably rugged. However, these methods are generally computationally efficient. Gradient methods can be further subdivided into *first-order*, e.g. back-propagation, or *second-order methods*, e.g. Newton's method and conjugate gradient method (Hagan et al., 1995; Bishop, 1995). With back-propagation, suitable values of the network training parameters, e.g. *learning rate* and *momentum term*, also need to be initialised. The suitable values of the said parameters are also case-specific. A few empirical formulas for calculating them are available in the literature (Gardner and Dorling, 1998; Hagan et al., 1995; Samarasinghe, 2006). There are several established criteria for initialising the training parameters to attain optimum model performance. Global optimization methods, such as genetic algorithms, have an increased ability to find global optima in the error surface, although this is generally at the expense of computational efficiency. Alternatively, stochastic calibration methods can be applied to account for parameter uncertainty. These approaches can be used to obtain distributions of the model parameters, rather than finding a single parameter vector. This advantage lies in their ability to estimate confidence intervals. In order to achieve this, Bayesian methods are commonly used (Maier et al., 2010; Bishop, 1995).

The number of occurrences various training methods have been used is shown in Figure 2.21.
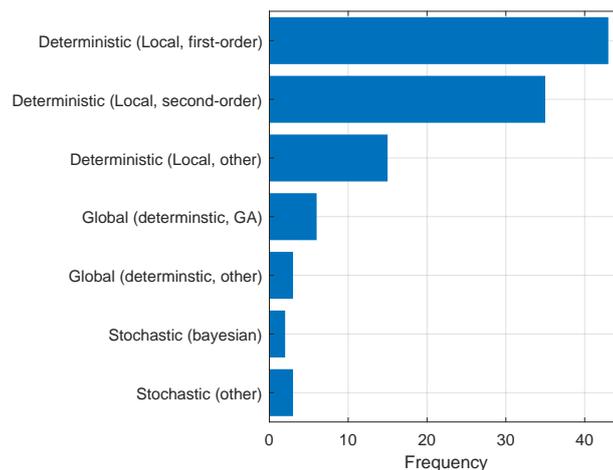
Fig. 2.21. Number of occurrences various training methods have been used.

Deterministic local calibration techniques were predominantly used, e.g. 93 times out of the 107 cited techniques by the identified papers. Of those deterministic techniques, first-order and second-order approaches were employed in 43 and 35 instances, respectively. Additionally, there were 9 studies that utilised global techniques and 3 that employed stochastic techniques.

2.2.2.8 Validation of Model Performance

A wide range of statistical indices must be employed to quantify model performance which is usually assessed using quantitative error metrics. There are three aspects of model validity, namely, *replicative validity*, *predictive validity*, and *structural validity*. In this review, the studies were assessed only in terms of predictive validity, e.g. the ability of models to approximate unforeseen or independent data. Other works such as Humphrey et al. (2017), Gass (1983) and Dawson et al. (2007) provide a more detailed discussion regarding model validation. The commonly used predictive validity metrics are categorised as squared, relative, absolute or correlation metrics.

Squared errors are based on the squares of the differences between actual and modelled output values. Common examples include the Mean Squared Error (MSE), Sum of Squared Errors (SSE) and Root Mean Squared Errors (RMSE). Absolute errors are based on the absolute differences between actual and modelled outputs. Metrics under this type include the Mean Sum of Absolute Deviations (MSAD) and Total Sum of Absolute Deviations (TSAD). Relative errors measure the performance of models with outputs. Common examples under this type include the Average Absolute Relative Error (AARE), Normalised Root Mean Squared Error (NRMSE) and Normalised Mean Bias Error (NMBE). Correlation errors measure the empirical error between actual and modelled outputs. One common example is the Pearson correlation coefficient (R). Other metrics include the information criteria, such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), which consider model complexity in addition to model error. Visual inspection tools qualitatively describe the performance of developed models by either comparing the plots between real and estimated values or providing several charts of the model residual errors, e.g. bar charts, error

scatter plots, and error distribution.

The number of times various model predictive performance metrics were used to validate the models is shown in Figure 2.22. While correlation and squared error metrics were also widely used (163 and 118 times, respectively), measures based on absolute and relative errors, were also employed extensively.
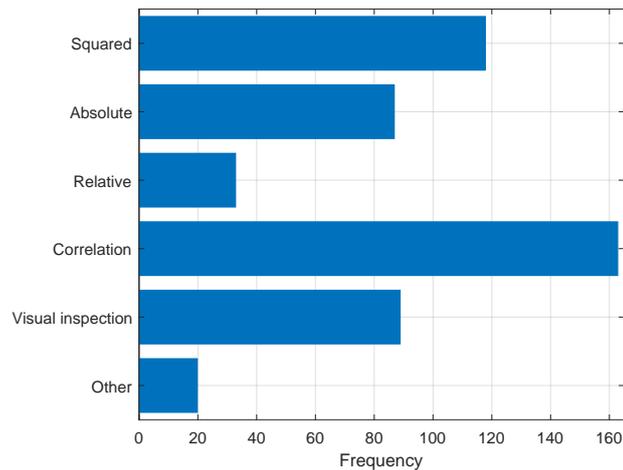
Fig. 2.22. Number of occurrences various predictive validity metrics have been considered.

Visual inspection was carried out by the majority of the papers identified. The comparison between plots of actual and predicted values was predominantly shown (44 times), to be followed by scatter plots (36 times) and error histogram and surfaces (9 times).

2.2.3 Comparison of model performance

2.2.3.1 MLP vs. linear statistical models

As one of the most popular ANN architectures applied for outdoor AP modelling, MLP models have been demonstrated by many studies to outperform traditional linear statistical models such as Multiple Linear Regression (MLR) models (Gardner and Dorling, 1998; Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019).

The scope and results of the said studies are summarised in Table 2.1. Note that those with the superscript * denotes results derived from a testing set collected from warm season, while ** from cold season. Although the studies included in the table

33

have different forecast targets and locations and a direct comparison of the results is not possible, they were included nonetheless for completeness.

Table 2.1. Identified studies comparing the performance of MLP and linear statistical models.

		Performance	
Publication	Location	MLP	MLR
a) Forecasting of PM$_{10}$:			
Kukkonen et al. (2003)	Helsinki	R^2 = 0.30-0.43; R^2 = 0.20	R^2 = 0.03-0.38; R^2 = 0.26
Vlachogianni et al. (2011)	Athens	R = 0.64-0.72*	R = 0.59-0.73*
Paschalidou et al. (2011)	4 sites in Cyprus	R^2 = 0.65-0.76	R^2 = 0.33-0.35
Russo et al. (2015)	Lisbon	R = 0.81	R = 0.75
b) Forecasting of NO$_2$:			
Kukkonen et al. (2003)	Helsinki	R^2 = 0.44-0.73; R^2 = 0.28-0.59	R^2 = 0.47-0.48; R^2 = 0.01
Agirre-Basurko et al. (2006)	4 sites in Bilbao	R = 0.89-0.91; R = 0.85-0.89; R = 0.89-0.90; R = 0.88	R = 0.88; R = 0.84; R = 0.88; R = 0.88
Elangasinghe et al. (2014)	Auckland	R^2 = 0.77	R^2 = 0.17-0.60
Rahimi (2017)	Tabriz	R^2 = 0.92	R^2 = 0.41
c) Forecasting of NO$_x$:			
Vlachogianni et al. (2011)	Athens	R = 0.85-0.76*; R = 0.62-0.74**	R = 0.64-0.73*; R = 0.60-0.69**
Rahimi (2017)	Tabriz	R^2 = 0.94	R^2 = 0.44
d) Forecasting of O$_3$:			
Agirre-Basurko et al. (2006)	4 sites in Bilbao	R = 0.91-0.92; R = 0.92-0.93; R = 0.91-0.93; R = 0.83-0.90	R = 0.88; R = 0.90; R = 0.91; R = 0.89

The majority of the studies indicate a significant improvement in terms of R or R^2 score whenever MLP models are compared to MLR models. In the study of Rahimi (2017) to provide short-term forecasts of hourly NO$_2$ and NO$_x$ concentrations at two locations in Tabriz, the results report the superiority of MLP models (R^2 of 0.92 and 0.94 for NO$_2$ and NO$_x$, respectively) over linear statistical models (R^2 of 0.41 and 0.44 for NO$_2$ and NO$_x$, respectively). Elangasinghe et al. (2014) also demonstrated the superiority of MLP models over linear regression models in their study to forecast NO$_2$ levels in a major highway in Auckland. Their results reported an improvement in terms of

R^2 scores from 0.17 to 0.60 of the benchmark model results when MLP models were employed. Kukkonen et al. (2003) developed several MLP models with a linear model and a deterministic modelling system in estimating NO_2 and PM_{10} levels at two urban locations in Helsinki. Despite the poor performance of the said developed, their findings revealed a better agreement between the actual and predicted NO_2 and PM_{10} values from the MLP models than those by the benchmark models. Agirre-Basurko et al. (2006) implemented MLP and MLR models to forecast NO_2 and O_3 levels at four locations in Bilbao. The work developed a total of 32 models, accounting for the four stations at each forecasting time step k, for $k = 1, 2, \ldots, 8$. Their results showed that the MLP models performed better than the MLR models in 75% of the cases (in six forecasts up to eight) in the study area.

However, a few studies report a marginal improvement of forecasting results achieved by the MLP models. Vlachogianni et al. (2011) suggests the use of simpler statistical techniques as alternatives to the use of ANN models especially for regulatory purposes, while Russo et al. (2015) reports no clear superiority of non-linear models over linear models in general.

2.2.3.2 Hybrid vs. plain ANN models

As shown in Table 2.2, hybrid forms of ANN models have been shown to be superior to plain ANN models by numerous identified studies.

Table 2.2. Identified studies comparing the performance of MLP and linear statistical models.

	Model Performance		Model Details*	
Publication	Hybrid	Plain	Hybrid	Plain
a) Forecasting of PM_{10}:				
Díaz-Robles et al. (2008)	$R^2 = 0.98$	$R^2 = 0.77$	ARIMAX; MLP	MLP
Siwek and Osowski (2012)	$R = 0.92 \pm 0.03$	$R = 0.61 \pm 0.09$	WT; MLP	MLP
Ul-Saufie et al. (2013)	$R^2 = 0.78$	$R^2 = 0.64$	PCA; MLP	MLP
Dunea et al. (2015)	$R = 0.97$	$R = 0.76$	WT; MLP	MLP
Bai et al. (2016)	RMSE = 15.39	RMSE = 23.62	WT; MLP	MLP
b) Forecasting of $PM_{2.5}$:				

Continued on next page

Table 2.2 – *Continued from previous page*

Publication	Model Performance		Model Details*	
	Hybrid	Plain	Hybrid	Plain
Zhou et al. (2014)	RMSE = 29.41	RMSE = 32.31	EEMD-GRNN	GRNN
Feng et al. (2015)	R = 0.86	R = 0.71	WT; MLP	MLP
Yeganeh et al. (2017)	$R^2 = 0.84$	$R^2 = 0.61$	Neuro-Fuzzy	MLP
Li and Zhu (2018)	$R^2 = 0.90$	$R^2 = 0.21$	ICEEMDAN -ICA-ELM	MLP
Bai et al. (2019)	RMSE = 12.08	RMSE = 24.30	EEMD-LSTM	MLP
Zhao et al. (2019)	RMSE = 35.82	RMSE = 37.71	LSTM-FC	MLP
Qi et al. (2019)	AAR = 0.81**	AAR = 0.74**	GC; LSTM	MLP
Wen et al. (2019)	RMSE = 12.08	RMSE = 39.92	C-LSTME	SVR
Qin et al. (2019)	R = 0.97	R = 0.92	CNN; LSTM	MLP
Huang and Kuo (2018)	RMSE = 23.83	RMSE = 29.99	CNN; LSTM	MLP
Ma et al. (2019)	RMSE = 8.24	RMSE = 11.20	BLSTM; IDW	MLP
	RMSE = 8.24	RMSE = 8.98	BLSTM; IDW	LSTM
c) Forecasting of NO_2:				
Dunea et al. (2015)	R = 0.79	R = 0.76	WT; MLP	MLP
Bai et al. (2016)	RMSE = 2.62	RMSE = 5.41	WT; MLP	MLP
Catalano et al. (2016)	R = 0.92	R = 0.91	Ensemble: SARIMAX; MLP	MLP
Li and Zhu (2018)	$R^2 = 0.90$	$R^2 = 0.50$	ICEEMDAN -ICA-ELM	MLP
d) Forecasting of O_3:				
Feng et al. (2011)	RMSE = 18.01	RMSE = 67.89	SVM-ANN	ANN
Mishra and Goyal (2016)	R = 0.85	R = 0.83	Neuro-Fuzzy	ANN
	R = 0.85	R = 0.84		
	R = 0.84	R = 0.67		
	R = 0.92	R = 0.79		
Pak et al. (2018)	RMSE = 3.20	RMSE = 21.37	CNN; LSTM	MLP
	RMSE = 3.20	RMSE = 4.98	CNN; LSTM	LSTM
Li and Zhu (2018)	$R^2 = 0.91$	$R^2 = 0.55$	ICEEMDAN -ICA-ELM	MLP
e) Forecasting of SO_2:				
Li and Zhu (2018)	$R^2 = 0.97$	$R^2 = 0.72$	ICEEMDAN -ICA-ELM	MLP

To cite a few, Mishra and Goyal (2016) demonstrated that hybrid NF models provided approximately 14% more accurate predictions than the benchmark plain ANN model. Yeganeh et al. (2018) showed that their NF models improved the performance of plain ANN models. Bai et al. (2016) reported that the estimates of wavelet-based

hybrid models are closer to the actual observations than those yielded by the plain ones. The same conclusions were reached in a study by Dunea et al. (2015) where hybrid ANN models based on the Daubechies Db3 wavelets were utilised to estimate the hourly AP levels at four various urban locations in Romania.

Data-intensive hybrid models have also been shown to significantly improve the performance of benchmark models in almost all cases (Siwek and Osowski, 2012; Dunea et al., 2015; Li and Zhu, 2018; Ul-Saufie et al., 2013; Bai et al., 2016; Feng et al., 2015). On the other hand, model-intensive hybrid models based on deep learning have been shown to outperform MLP and LSTM models, especially in the forecasting of $PM_{2.5}$ levels (Huang and Kuo, 2018; Pak et al., 2018; Qin et al., 2019; Qi et al., 2019).

Finally, Table 2.2 reveals that MLP models have been widely used as benchmarks for hybrid models. It appears that it is standard practice to compare the performance of the proposed hybrid approaches to MLP models.

2.2.4   Spatial or spatiotemporal modelling using ANN models

Several studies have also attempted to incorporate ANN models in estimating the spatial and spatiotemporal variations of AP concentration levels in ambient environments.

One popular approach is by utilising predictors from neighbouring sites to estimate the concentration levels of a given target site. For instance, Li et al. (2017) estimated $PM_{2.5}$ levels in China using multiple RBM (Restricted Boltzmann machine) layers and a BP layer. The proposed model was trained using $PM_{2.5}$ observations from both neighbouring sites and the same station along with other predictors. Alimissis et al. (2018) applied MLP models to spatially forecast NO_2, NO, O_3, CO and SO_2 levels at several locations in Athens. The models were trained using only the information from the neighbouring sites to estimate the concentration levels at the target site. A similar methodology was employed by Tzanis et al. (2019) to spatially estimate PM_{10} and $PM_{2.5}$ levels at several sites in Athens. The results also demonstrated the superior performance of MLP models when compared to other spatial interpolation schemes, e.g. MLR, Inverse Distance Weighting (IDW) and Linear Mixed Effects (LME) models.

Papaleonidas and Iliadis (2013) estimated O_3 levels at an urban location in Athens utilising various predictors from neighbouring sites using feedforward ANN models. The selection of the optimal neighbouring sites was carried out using correlation analysis. Kurt and Oktay (2010) built three MLP models based on geographic modelling to predict the concentration levels of SO_2, CO, and PM_{10} at ten different locations in Istanbul. The first geographic model utilised additional observed AP values from a selected neighbouring site, the second used two neighbouring sites instead of one, and the third considered the distance between the triangulating districts.

The utilisation of predictors revealing spatial information of a given geographical area of interest is also a popular approach. For instance, Yeganeh et al. (2018) applied an NF model to estimate the spatiotemporal variability of the monthly mean NO_2 levels at 1km spatial resolutions at a location in Australia using land-use, meteorological, satellite, and traffic predictors from 12 monitoring sites. Yeganeh et al. (2017) applied NF, SVR and MLP models to estimate the spatiotemporal variations of the monthly $PM_{2.5}$ levels in using land-use, satellite, and meteorological predictors from eight monitoring stations in Queensland. Di et al. (2019) proposed a generalised additive model based on ANN, random forest and gradient boosting to estimate daily $PM_{2.5}$ levels at a 1 km x 1 km resolution across the contiguous USA. Each model of the ensemble approach was trained using several predictors including satellite, meteorological, land-use, emissions and others. Mao et al. (2017) developed MLP models using meteorological and satellite-based predictors. Backward air mass trajectory predictors were also used to account for the contribution of regional transport of $PM_{2.5}$. Each grid cell on the AOD retrievals was assigned to a prediction model trained by information from the nearest neighbouring station.

The use of sophisticated deep learning ANN models, especially LSTM models, have been recently applied to extract both temporal and spatial features from collected datasets. Such studies usually entail a vast amount of training data from numerous monitoring sites. For instance, Li et al. (2016) developed a model based on a Stacked Autoencoder (SAE) and a logistic regression layer to forecast $PM_{2.5}$ levels at 12 stations in Beijing. Tong et al. (2019) applied a bidirectional LSTM model to perform

38

spatiotemporal interpolation of $PM_{2.5}$ levels in Florida using several predictors from neighbouring sites. The authors argued the use of a bidirectional variant of an LSTM model to account for both past and future values of $PM_{2.5}$ levels which can provide valuable information for the estimation of the current $PM_{2.5}$ levels. Qi et al. (2019) proposed a hybrid approach based on graph CNN and LSTM models to estimate the spatiotemporal variation of $PM_{2.5}$ levels in Jing-Jin-Ji. Li et al. (2017) applied extended LSTM models to forecast the spatiotemporal variations of $PM_{2.5}$ levels at 12 monitoring stations in Beijing. Wang et al. (2018) proposed an ensemble approach consisting of several LSTM models with regression layers to forecast multiple pollutant species at 35 stations in Beijing. The relevant neighbouring sites were selected based on the Granger causality analysis of the spatial correlations among the stations. Zhao et al. (2019) proposed an approach combining LSTM models and a feedforward ANN model to respectively extract temporal and spatial features of the input data. In more detail, the hybrid model was applied to estimate the $PM_{2.5}$ concentration of a given central site Beijing using the previous measurements of the same site and neighbouring sites.

2.3 Summary

A critical review of 139 papers dealing with the forecasting of outdoor AP levels using ANN models was conducted for this research. The results of the comprehensive review are as follows:

1. Pollutants such as PM_{10}, $PM_{2.5}$, and oxides of nitrogen have been the most examined variables. Most identified studies also utilised data covering more than one year for ANN model development which is considered a good practice (Hagan et al., 1995; Arhami et al., 2013). However, the practice does not fully address other prevailing issues regarding input data including the imbalanced data problem, and high levels of data variability (Fernando et al., 2012; Gong and Ordieres-Meré, 2016; Bai et al., 2019);

2. Meteorological and pollutant emissions data have been most commonly utilised predictors. However, the choice of predictor types and size should be dealt with extra

care as the use of too many predictors tend to reduce non-linear models such as ANN models to linear models (Ibarra-Berastegi et al., 2008);

3. Data-processing has been handled really well by most practitioners, except for missing data imputation in which data deletion was commonly practised;

4. Ad-hoc approaches have been employed predominantly in predictors selection, data splitting and model architecture and structure selection. That is, there has been a little adoption on the utilisation of systematic approaches limiting future modellers from repeating the results of the identified papers (Wu et al., 2014);

5. Much effort has been directed towards the application of existing popular ANN architectures, especially MLP models. However, there has been a growing interest in the use of novel and more sophisticated network types recently. In more detail, hybrid models incorporating deep learning ANN models have been built in the spatial and spatiotemporal estimation of pollution levels;

6. First-order and second-order local search procedures, such as the backpropagation algorithm, have been primarily used in training ANN models. However, studies investigating the potential benefits of using global optimization techniques in terms of improving the predictive ability of ANN models are rather limited; and

7. A wide range of performance criteria, mainly those assessing the predictive performance of models, was used predominantly by the identified papers. While this suggests a good modelling practice, the replicative and structure validity of the models developed were generally ignored.

2.4   Research Gaps

Based on the results of the undertaken review, the following gaps were identified:

1. *Systematic approaches that implement most of the steps in developing ANN models.* Analytic techniques in developing ANN models should be further examined. This is critical given the lack of a clean-cut set of ANN model settings for all applications.

This can be circumvented by adhering to a set of general model building guidelines to minimise uncertainty and increase the reproducibility of the results of a given application: data preprocessing (Bowden et al., 2012), predictor selection (Galelli et al., 2014), data splitting (Wu et al., 2013), model architecture selection (Hunter et al., 2018), model structure selection (Kingston et al., 2008), and model validation (Humphrey et al., 2017; Dawson et al., 2007).

2. *The development of data-intensive ANN models.* Coupling ANN models with techniques that reveal additional information regarding the predictors can address existing issues such as the imbalanced data problem and measurements exhibiting high variability (Gong and Ordieres-Meré, 2016; Siwek and Osowski, 2012).

3. *The use of sophisticated models, especially those that employ deep learning and ensemble modelling techniques.* Data-intensive hybrid model approaches based on deep learning models and ensemble modelling techniques should be tested and validated given the rapidly increasing availability of data from continuous sensors in recent years.

4. *The demand for robust ANN models capable of revealing inherent features from a small number of predictors.* Although the utilisation of several significant predictors is encouraged, parsimonious ANN models developed especially in monitoring sites where data is limited should be developed and tested.

The aim of this research is to fill the said gaps as mentioned in Chapter 1 Section 1.4.

Chapter 3

Methodology

3.1 Introduction

This chapter presents the methods involved in implementing the experiments briefly outlined in Chapter 1 Section 1.5. A diagram of the framework for each experiment is provided to guide the readers.

3.2 Modelling Hourly NO_2 using ANN Models with Predictor Selection Techniques

This section discusses the various predictor selection techniques implemented in Experiment 1.

As shown in Figure 3.1, the experiment has two main components: (1) predictor selection, and (2) the training of the feedforward ANN model. The experiment attempts to determine the optimal set of p out of a candidate of set P predictors ($p \leq P$) using various predictor selection techniques, e.g. model-free and model-based techniques. The said predictor selection techniques are discussed in Section 3.2.1 through Section 3.2.4. Finally, the MLP model, as well the basic theories behind ANN computations to be trained using the selected predictors, is thoroughly described in Section 3.2.5.
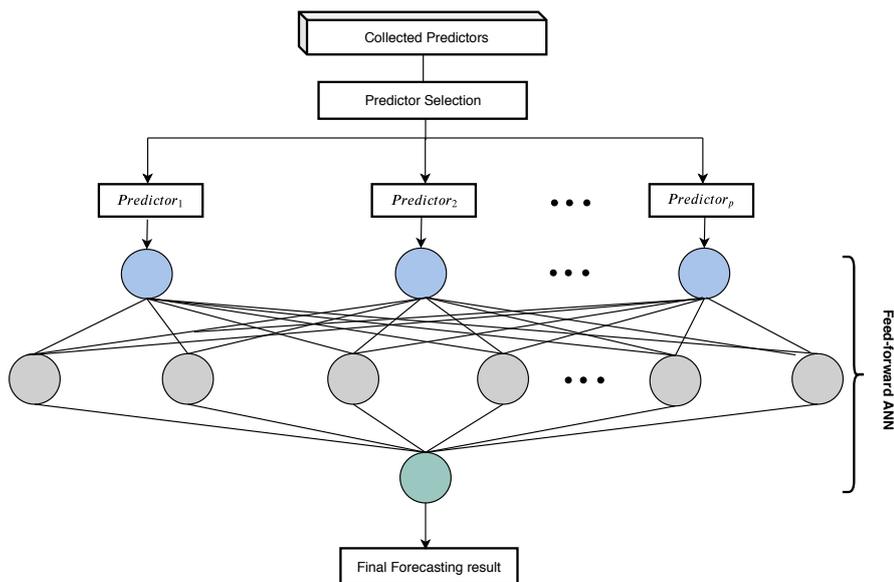
Fig. 3.1. The framework of the proposed forecasting model in Experiment 1.

3.2.1 Stepwise Regression

Stepwise regression is a linear search strategy procedure based on two known predictor selection techniques, namely, *forward selection* and *backward elimination*. It is a filter type class (model-free) selection technique as the results of the linear search strategy do not rely on the results of an ANN model. (The following notations are consistent with those used by Wilks (2006).)

Forward selection begins without any predictor, yielding an uninformative prediction model:

$$\hat{y} = b_0, \tag{3.1}$$

where $\hat{y}$ is the estimate of the predictand, Y, and b_0 is the intercept which is the sample mean of the predictand. The process proceeds by subsequently adding a predictor, $\mathbf{x}_1$, from a set of P predictors that is most linearly correlated with the predictand. The resulting model is then given by

$$\hat{y} = b_0 + b_1 \mathbf{x}_1, \tag{3.2}$$

44

where b_i is the coefficient of $\mathbf{x}_i$. Note that b_0 is no longer the average of the Y values. Together with the previously selected predictor, $\mathbf{x}_1$, an additional predictor from the remaining $(P-1)$ predictors is then selected. The predictor that provides the best regression performance of the model given by

$$\hat{y} = b_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 \tag{3.3}$$

is selected as $\mathbf{x}_2$. The performance of the model is usually assessed in terms of the R^2, MSE and F values. The same selection process is carried out in the subsequent steps until a stopping criterion is met. Note that as the regression equations are updated, the coefficients for the intercept and for the previously chosen predictors will change.

Algorithm 1 provides the pseudocode for the basic steps of Forward Selection technique.

Algorithm 1 Forward Selection

---

1. Define L that measures the correlation between a predictor X^+ and the predictand Y

2. Create an empty candidate set of predictors: $C^{(k)} = \{\emptyset\}$, $k = 0$

3. Select the best remaining predictor:

$$X^+ \leftarrow \arg \max_{X^+ \in C^{(k)}} \left[L\left(C^{(k)} \cup X^+\right)\right]$$

4. If $L\left(C^{(k)} \cup X^{(k)}\right) > L\left(C^{(k)}\right)$

    a) Update: $C^{(k+1)} \leftarrow C^{(k)} \cup X^+$

    b) $k \leftarrow k+1$

    c) Repeat step 2

On the other hand, backward elimination is the opposite of Forward Selection. Backward elimination begins with the regression model utilising all P predictors:

$$\hat{y} = b_0 + b_1\mathbf{x}_1 + b_2\mathbf{x}_2 + \ldots, b_P\mathbf{x}_P. \tag{3.4}$$

At each step of the process, the predictor whose coefficient is smallest in absolute value

relative to its estimated standard error is removed from the equation.

3.2.2 Neighbourhood Component Analysis (NCA) with Regularisation

Neighbourhood component analysis (NCA) with regularisation is a nearest neighbour-based method predictor weighting technique developed by Yang et al. (2012). NCA is a filter type (model-free) predictor selection technique which learns the predictor weights that minimise an objective function evaluating the prediction loss over the training data. (The following notations are consistent with those used by Yang et al. (2012).)

Let $T = \{(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$, be a set of training samples, where $\mathbf{x}_i$ is P-dimensional vector which corresponds to the i-th sample of the all predictors, y_i is the i-th sample of the predictand Y, and N is the number of samples. The objective of NCA is to find a weighting vector $\mathbf{w}$ that lends itself to select the predictor subset optimising nearest neighbour prediction.

Now consider a randomised regression model that randomly picks a reference point for x, $\text{Ref}(x)$, and sets the response value at x equal to the response value of the reference point $\text{Ref}(x)$. The probability $P(\text{Ref}(x) = x_j \mid T)$ that point x_j is picked from T as the reference point for x is given by:

$$P(\text{Ref}(\mathbf{x}) = \mathbf{x}_j \mid T) = \frac{k(d_w(x, x_j))}{\sum_{j=1}^{n} k(d_w(x, x_j))}, \tag{3.5}$$

where $d_w(\mathbf{x}_i, \mathbf{x}_j)$ measures the weighted distance between two samples $\mathbf{x}_i$ and $\mathbf{x}_j$,

$$d_w(\mathbf{x}_i, \mathbf{x}_j) = \sum_{p=1}^{P} w_p^2 |x_i p - x_j p|, \tag{3.6}$$

where w_p is the weight associated with the p-th predictor. Afterwards, a leave-one-out approach is applied in which the response for $\mathbf{x}_i$ using the data in T^{-i}, e.g. the set T without the point $(\mathbf{x}_i, \mathbf{y}_i)$, is estimated. Hence, the probability $P(\text{Ref}(x) = x_j \mid T)$

that point x_j is picked from T as the reference point for x is

$$P\left(\mathrm{Ref}\left(\mathbf{x}\right) = \mathbf{x}_j \mid T^{-i}\right) = \frac{k\left(d_w\left(x, x_j\right)\right)}{\sum_{j=1, j\neq i}^{n} k\left(d_w\left(x, x_j\right)\right)}. \tag{3.7}$$

Now, let $L : \mathbb{R}^2 \mapsto \mathbb{R}$ be a loss function that measures the disagreement between the predicted value of the randomised regression model, $\hat{y}_i$, and response, y_i. Then, the average value of $L\left(y_i, \hat{y}_i\right)$ is given by

$$L_i = E\left(L\left(y_i, \hat{y}_i\right) \mid T^{-i}\right) = \sum_{j=1, j\neq i}^{N} p_{ij} L\left(y_i, y_j\right). \tag{3.8}$$

The objective function to be minimised is then given by:

$$f\left(w\right) = \frac{1}{N} \sum_{i=1}^{N} L_i + \lambda \sum_{r=1}^{P} w_r^2, \tag{3.9}$$

where λ is a regularisation parameter for all weights.

Algorithm 2 provides the pseudocode for the basic steps of NCA predictor selection technique.

Algorithm 2 Neighbourhood Component Analysis

---

1. Initialise: $\mathbf{w}^{(k)} = (1, 1, \ldots, 1)$, $k = 0$

2. Select $\mathrm{Ref}(x)$

3. Compute: $d_w\left(x_i, x_j\right)$ and $P\left(\mathrm{Ref}\left(x\right) = x_j \mid T\right)$ using $\mathbf{w}^{(k)}$

4. $\mathbf{w} \leftarrow \arg\min_{\mathrm{Ref}(x)^{(k)}} f\left(w\right)$

5. Update: $\mathbf{w}^{(k)} \leftarrow \mathbf{w}$

6. $k \leftarrow k + 1$

7. Repeat step 2

3.2.3 Sequential Forward Selection (SFS)

Sequential forward selection (SFS) is a technique that identifies a subset of optimal predictors by sequentially selecting predictors until there is no improvement in the

prediction or a stopping criterion is met.

SFS adapts the same forward selection approach of stepwise regression in Section 3.2.1 by starting from an empty set of candidate predictors. However, unlike stepwise regression, SFS carries out a 10-fold cross-validation procedure by utilising the candidate set of predictors to train a given regression model. This makes SFS a wrapper (model-based) predictor selection technique. SFS then assesses the importance of the predictors based on a predefined loss function. SSE is a commonly used loss function. A candidate predictor that minimises the mean loss value is then selected, and SFS continues until the addition of more predictors does not reduce the loss value.

Algorithm 3 provides the pseudocode for the basic steps of SFS technique.

Algorithm 3 Sequential Forward Selection
---

1. Initialise predictor set: $C^{(k)} = \{\emptyset\}$, $k = 0$

2. Define estimator F such that $\hat{y}_i = F\left(C^{(k)}\right)$

3. Define criterion E measuring the error between estimated $\hat{y}_i$ and $y_i \in Y$

4. Select an optimal predictor:

$$X^+ \leftarrow \arg \min_{X^+ \in C^{(k)}} \left[E\left(C^{(k)} \cup X^+\right)\right]$$

5. If $E\left(C^{(k)} \cup X^{(k)}\right) < E\left(C^{(k)}\right)$

   a) Update: $C^{(k+1)} \leftarrow C^{(k)} \cup X^+$

   b) $k \leftarrow k + 1$

   c) Repeat step 4

3.2.4 Classification and Regression Trees (CART)

Classification and Regression Trees (CART) is a set of statistical procedures introduced by Breiman et al. (1984). CART employs the methodology known as the *Binary Recursive Partitioning* (BRP) algorithm.

As described by Merkle and Shaffer (2011), the BRP algorithm has three main characteristics:

(1) it estimates the value of the predictand by partitioning the training data into subgroups based on the predictors;

(2) at any step, it partitions the training data into two subgroups; and

(3) within the subgroups produced from a predictor, it repeats the partitioning process based on other predictors or other splits of the same predictor.

Classification trees are designed for predictands that take a finite number of unordered values, while *regression trees* are for predictands that take continuous or ordered discrete values. Since this work only utilises predictors with continuous values, this paper shall be focusing on regression trees henceforth.

In a regression problem, there is a training dataset of n observations on a predictand Y and P predictors, $X_1, X_2, \ldots, X_P$. A regression tree is formed by finding the split in the predictor space X that best discriminates between classes of the predictand with continuous values. The result of the said process is a binary decision tree, whose terminal nodes represent distinct classes or categories of data.

The binary decision tree starts with a root node t derived from a predictor in the dataset that minimises the measure of impurity or entropy of the two sibling nodes. The node impurity is usually evaluated by the sum of squared deviations about the mean and the node predicting the sample mean of Y. The regression model is then fitted to each node to give the predicted values of Y. The binary splitting of the original dataset is repeatedly performed to determine the predictors that minimise Eq. (3.10) and until some pre-specified criteria are met. The importance of the predictors are then assessed by summing the changes in the mean squared error incurred due to splits on every predictor and dividing the sum by the total number of branch nodes formed: The said node impurity is given by

$$i\left(j,s\right) = \sum_{i:x_i \in C_1(j,s)} \left(y_i - \hat{y}_{C_1}\right)^2 + \sum_{i:x_i \in C_2(j,s)} \left(y_i - \hat{y}_{C_2}\right)^2, \qquad (3.10)$$

where C_1 and C_2 denote the two distinct non-overlapping regions (child node 1 and child node 2) produced by dividing the predictor space set using the splitting point $s \in S$

for any predictor j, and y and $\hat{y}$ are the actual and predicted values using regression according to the points in C_1 and C_2.

Algorithm 4 gives the pseudocode for the basic steps of the Regression Tree predictor selection technique.

Algorithm 4 Pseudocode for Regression Tree construction

---

1. Start at the root node

2. For each predictor X_j, select the best split s

$$s^* \leftarrow \arg \min_{s^* \in S} \left[i(j, s) \right]$$

3. If a stopping criterion is reached, exit

4. Otherwise, repeat step 2 to each child node

3.2.5 The Multilayer Perceptron (MLP) model

As mentioned in Chapter 2 Section 2.2.2.5, the MLP is one of the most commonly used feedforward ANN forms for prediction and function approximation (Gardner and Dorling, 1998). Although the inner mechanisms of feedforward ANNs were mentioned several times in the preceding chapters, the following provides a mathematical description of their structure and training. The MLP consists of the following components:

- an input layer with p inputs, corresponding to the number of utilised predictors,

- one or more hidden layer, where each layer has an arbitrary number of nodes, and

- an output layer with o number of nodes, corresponding to the number of outputs of the ANN.

Each layer in an MLP network is fully-connected to the succeeding layer, forming a forward connection between each node in layer l to each node in layer $l + 1$. Figure 3.2 shows a graphical representation of an MLP with P input nodes, one hidden layer with M nodes, and an output layer with K nodes. The MLP shown in the figure has $(P + M + K)$ nodes.
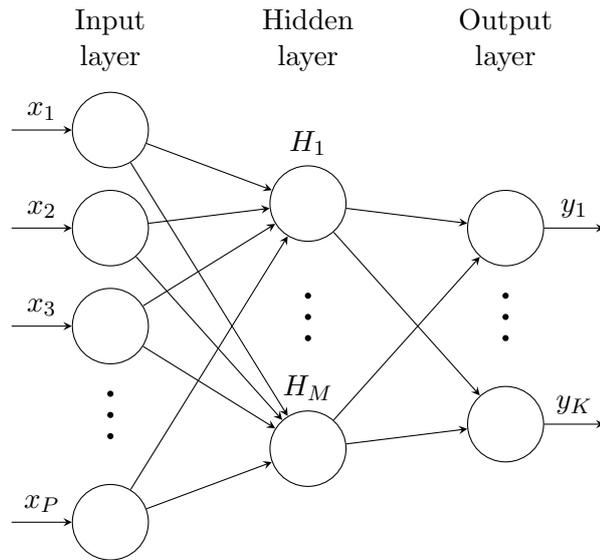
Fig. 3.2. A feedforward ANN with N input, M hidden, and K output nodes.

The structure of a node is shown in Figure 3.3. Each node in the hidden layer is initially fed with the values of the input nodes, $x_1, \ldots, x_P$, each of which is scaled by the network connection weights **w** and biases **b**. The resulting value n is then mapped by f to each node in the succeeding layers.
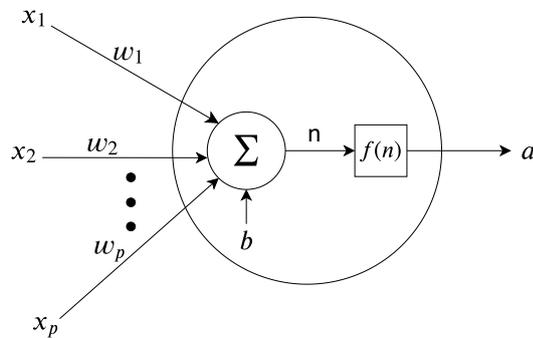
Fig. 3.3. Structure of an ANN hidden node.

Adapting the notations used in Bishop (1995), the forward process described above is generally given by Eq. (3.11):

$$a_j = f\left(\sum_{i=1}^{p} x_i w_{ji}^{(1)} + b_j^{(1)}\right), \tag{3.11}$$

where $j \in [1, M]$, and M is the number of hidden nodes, $w_{ji}^{(l)}$ and $b_j^{(l)}$ are the weights and bias parameters, respectively, $f(\cdot)$ is the mapping commonly referred to as the transfer function, and the superscript (l) denotes that the corresponding parameters are those in the l-th layer of the network.

Transfer functions can be any mapping and are normally chosen based on the nature of the modelling task (Hagan et al., 1995). One of the most popular transfer functions in environmental modelling applications is the *sigmoid transfer function*. As shown in Figure 3.4a, a sigmoid function has a graph that looks like a stretched 'S' and has a range between either 0 to 1 or -1 to 1. These characteristics enable an ANN to approximate any non-linear and complex relationships between predictors and predictands (Hornik et al., 1989; Bishop, 1995). As such, the sigmoid function is commonly used in the nodes of the hidden layers (Cabaneros et al., 2019).

(a) $f(x) = \frac{1}{1+e^{-x}}$ **(b)** $f(x) = x$
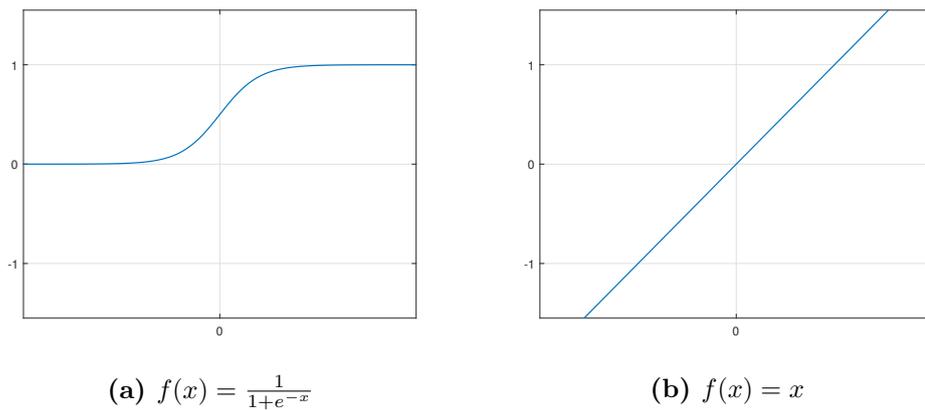
Fig. 3.4. The graphs of the (a) sigmoid and (b) linear transfer functions.

Another popular transfer function is the *linear transfer function* which is commonly used in the output nodes of ANNs employed to estimate continuous predictand values (see Figure 3.4b). For a comprehensive list of transfer functions, the reader is referred to Hagan et al. (1995).

Following Eq. (3.11), the final result of the output layer of an ANN with only one

hidden layer can be computed as follows:

$$\hat{y}_K = \sum_{j=1}^{M} a_j w_{kj}^{(l_o)} + b_k^{(l_o)}, \tag{3.12}$$

where $k \in [1, k]$, and l_o refers to the output layer.

The training of feedforward ANNs is carried out in a supervised manner in which a series of input and target output values are presented to the model. That is, an ANN model is fed with the following sequence of proper network behaviour:

$$\{p_1, y_1\}, \{p_2, y_2\}, \ldots, \{p_Q, y_Q\}, \tag{3.13}$$

where $\mathbf{p}_q$ is an input to the ANN model and $\mathbf{y}_q$ is the corresponding target output. The values of the weights and bias parameters are initially set at random. During the training process, the weights and bias values are adjusted based on the network error, e.g. the difference between the input and target values (see Figure 3.5). The process is repeated until a predefined threshold is met.
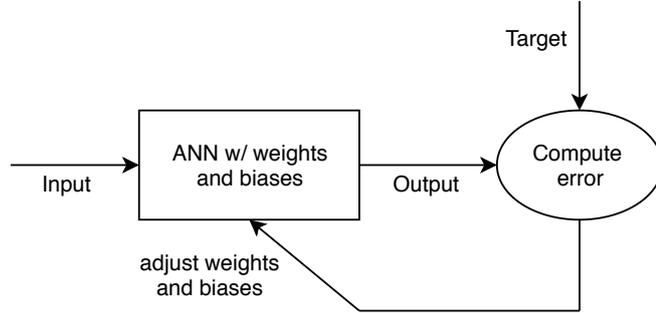
Fig. 3.5. ANN training algorithm (Cabaneros et al., 2017).

The network error is given by:

$$E = \frac{1}{2Q} \sum_{i=1}^{Q} (y_i - \hat{y}_i)^2, \tag{3.14}$$

where Q is the total number of training data. ANNs are trained using the *backpropagation* algorithm which adjusts the weights and biases of the network according to the

error gradient to descend the error surface. The algorithm attempts to improve the performance of an ANN by reducing the network error along its gradient.

Note that E in Eq. (3.14) can be expressed as the function of the network weights and biases. The MSE of the network outputs can then be given by:

$$F\left(\mathbf{w}\right) = \left(y\left(k\right) - \hat{y}\left(k\right)\right)^{T}\left(y\left(k\right) - \hat{y}\left(k\right)\right) = e^{T}\left(k\right)e\left(k\right), \qquad (3.15)$$

As mentioned in Chapter 2 Section 2.2.2.7, there are several variants of the gradient descent methods. The Levenberg-Marquardt (LM) algorithm is utilised in the experiments of this thesis. The LM algorithm is a variant of the backpropagation algorithm for minimising functions that are sums of squares of other non-linear functions. Since the performance or error index for MLP models is the MSE, the use of the LM algorithm is considered appropriate (Hagan et al., 1995; Bishop, 1995). A full description of the algorithm is provided in Appendix B.

Based on Figure 3.2 and Eq. (3.12), $\hat{y}$ is dependent on the transfer function of the output node and the collective outputs of the hidden nodes. As such, the final output of an ANN in general depends on the inputs and hidden-node transfer functions.

3.2.6   Multiple Linear Regression (MLR) model

The MLR model is one of the most popular statistical models in AP forecasting applications (Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019).

The general form of the MLR model is given by:

$$\hat{Y}_i = \beta_0 + \beta_1\mathbf{x}_{i,1} + \beta_2\mathbf{x}_{i,2} + \ldots + \beta_P\mathbf{x}_{i,P} + \epsilon_i, \qquad (3.16)$$

where, for the i-th observation, $\hat{Y}_i$ is the estimated value of the predictand Y_i, β_0 is the regression coefficient, and β_p are the coefficients of the P independent predictors $\mathbf{x}_{i,p}$, $p = 1, 2, \ldots, P$, and ϵ_i is the residual error.

The MLR model assumes linearity between the predictors and a predictand, independence between predictors, and normality and independence between residual errors ϵ_i, with 0 mean and σ^2 constant variance. As such, a linear predictor selection technique

is normally applied to identify the set of independent predictors of the model.

To employ the MLR model, β_i coefficients are initially estimated using the observations during the calibration stage. The estimated coefficients are then used to apply the model using the test set for assessing model performance.

3.3 Temporal and spatiotemporal modelling of hourly NO_2 levels in Central London using wavelet-based ANN models

This section describes the wavelet-based modelling approach adapted in Experiment 2 which has been split into two sub-experiments, namely the temporal and the spatiotemporal analyses. With regards to the temporal modelling of NO_2 levels, data from the target site, e.g. MAR site, are utilised in training the proposed and benchmark models. On the other hand, the spatiotemporal modelling involves the estimation of pollutant NO_2 levels at a given site using only the data from the neighbouring sites. In detail, six sites in Central London are examined in the sub-experiment (see Chapter 4 Section 4.4).

As shown in Figure 3.6, the experiment has three main stages, namely: (1) wavelet preprocessing or decomposition, (2) ANN model development and training, and (3) reconstruction of results. In more detail, a wavelet decomposition technique is initially applied on the collected AP concentration time series. Discrete wavelet transformation is described in Section 3.3.1, while the Daubechies wavelets are defined in Section 3.3.2. The resulting subseries are then considered as the new predictands. Afterwards, separate ANN models are developed and trained to estimate the said predictands. Finally, the results of the ANN models are combined to yield the final wavelet-based ANN model result. Finally, other forms of ANN models are considered in the current experiment instead of the standard feedforward approach which was applied in Experiment 1. The said model are discussed in Section 3.3.3.
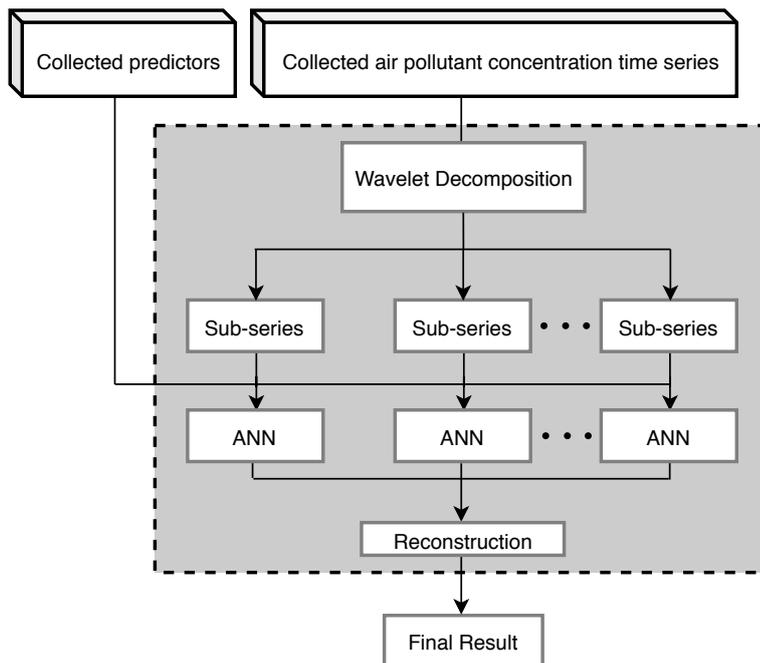
Fig. 3.6. The framework of the proposed spatio-temporal forecasting model in Experiment 2.

3.3.1 Discrete Wavelet Transformation (DWT)

Discrete Wavelet Transformation (DWT) is a technique that decomposes a given time series into several subseries of various scales. In general, wavelet transformation can be thought of as a prism that exhibits the properties such as points of abrupt changes, seasonality or periodicity of a given time series (Parmar and Bhardwaj, 2013). The following describes the way the original time series is decomposed using DWT.

A sequence of wavelet and binary scale functions, denoted by $\psi_{j,k}(t)$ and $\varphi_{j,k}(t)$, respectively, can be calculated as

$$\psi_{j,k}(t) = 2^{j/2}\psi\left(2^j t - k\right), \tag{3.17}$$

$$\varphi_{j,k}(t) = 2^{j/2}\varphi\left(2^j t - k\right), \tag{3.18}$$

where t, j, and k denote the time, scaling, and translation parameters, respectively,

$\psi(t)$ is a given mother wavelet function, and $\varphi(t)$ is its corresponding scaling function. The initial step of DWT is to map the elements of a given time series S to its wavelet coefficients, and from these coefficients, two components are formed, namely a smooth version called *approximation* and a component corresponding to the deviations called *details* of the signal:

$$S(t) = \sum_{k=1}^{n} c_{j,k} \psi_{j,k}(t) + \sum_{j=1}^{J} \sum_{k=1}^{n} d_{j,k} \psi_{j,k}(t), \qquad (3.19)$$

where $c_{j,k}$ and $d_{j,k}$ denote the approximation and detailed coefficients, respectively, at scale j and location k, n is the size of the original time series, and J is the decomposition level.

For instance, a decomposition of $S(t)$ into a low frequency part $A_1(t)$ and a high frequency part $D_1(t)$ is given by $S(t) = D_1(t) + A_1(t)$. The same process is carried out on $A_1(t)$ in order to obtain decomposition in finer scales: $A_1(t) = D_2(t) + A_2(t)$ (see Figure 3.7).
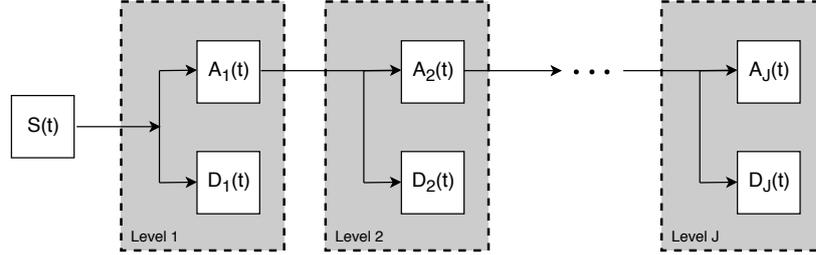
Fig. 3.7. Decomposition of time series S into approximation A_J and detailed components D_i $i \in [1, J]$.

Hence, Eq. (3.19) can be simplified into

$$S(t) = \sum_{i=1}^{J} D_i(t) + A_J(t), \qquad (3.20)$$

where $D_i(t) = \sum_{k=1}^{n} d_{j,k} \psi_{j,k}(t)$ and $A_J(t) = \sum_{k=1}^{n} c_{j,k}$. In other words, DWT represents S in terms of the sum of subseries consisting of high frequency detail signals $D_1, D_2, \ldots, D_J$ and a low frequency approximation signal A_J.

The proposed approach in this work is to use the separately-built ANN models in

estimating the detailed and approximation signals, instead of a single original time series S_t. The form of the final estimated value of $S(t)$ is then given by

$$\hat{S}(t) = \sum_{i=1}^{J} \hat{D}_i(t) + \hat{A}_J(t), \tag{3.21}$$

where $\hat{D}_i(t)$ and $\hat{A}_j(t)$ are respectively the detailed and approximation coefficients separately estimated by ANN models.

3.3.2 Daubechies wavelets

The Daubechies mother wavelets (Daubechies, 1988) are chosen in this work because they provide a smoother overall approximation of any given signal (Nievergelt, 2013). Daubechies wavelets are based on a basic "building block" or "scaling function", denoted by φ. The Daubechies basic building block φ is shown in Figure 3.8.
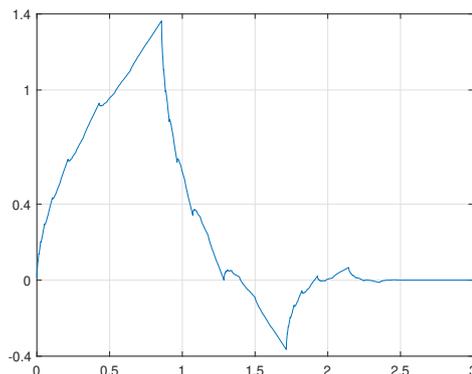
Fig. 3.8. Daubechies' scaling function φ.

The said building block function has no closed form, although it satisfies several algebraic relations that prove as useful as a formula for the purpose of calculations (Nievergelt, 2013). For instance, φ equals zero outside $[0,3]$:

$$\varphi(r) = 0 \text{ if } r \leq 0 \text{ or } 3 \leq r. \tag{3.22}$$

Furthermore, starting with the initial values $\varphi(0) = 0$, $\varphi(1) = \frac{1+\sqrt{3}}{2}$, $\varphi(2) = \frac{1-\sqrt{3}}{2}$,

and $\varphi\left(3\right)=0$, the building block φ satisfies the following recurrence relation:

$$\begin{aligned}\varphi\left(r\right)=&\frac{1+\sqrt{3}}{4}\varphi\left(2r\right)+\frac{3+\sqrt{3}}{4}\varphi\left(2r-1\right)\\&+\frac{3-\sqrt{3}}{4}\varphi\left(2r-2\right)+\frac{1-\sqrt{3}}{4}\varphi\left(2r-3\right).\end{aligned} \tag{3.23}$$

The function φ serves as the basic building block for the associated Daubechies mother wavelet, denoted by ψ, defined by the following recursion:

$$\begin{aligned}\psi\left(r\right)=&-\frac{1+\sqrt{3}}{4}\varphi\left(2r-1\right)+\frac{3+\sqrt{3}}{4}\varphi\left(2r\right)\\&-\frac{3-\sqrt{3}}{4}\varphi\left(2r+1\right)+\frac{1-\sqrt{3}}{4}\varphi\left(2r+2\right).\end{aligned} \tag{3.24}$$

The Daubechies wavelet function ψ is shown in Figure 3.9.
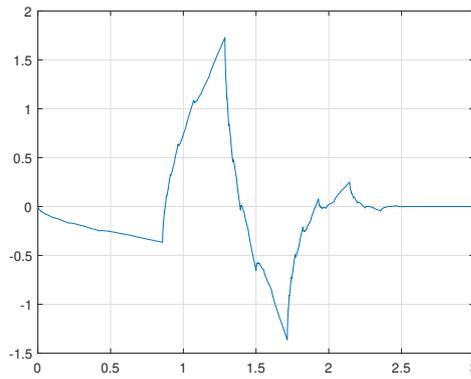
Fig. 3.9. Daubechies' wavelet function ψ.

Figure 3.10 illustrates a sample decomposition in four subsequent components of the NO_2 time series collected at Westminster-Marylebone Road site from 01/01/2007 at 00:00 to 11/02/2007 at 13:00 (King's College London, 2019). It is evident that the fluctuation patterns of each subseries differ, and that higher wavelet levels exhibit lower signal variability. Consequently, the task of estimating the said wavelets using a general approximator such as ANNs becomes significantly easier as demonstrated by previous studies (Siwek and Osowski, 2012; Dunea et al., 2015).
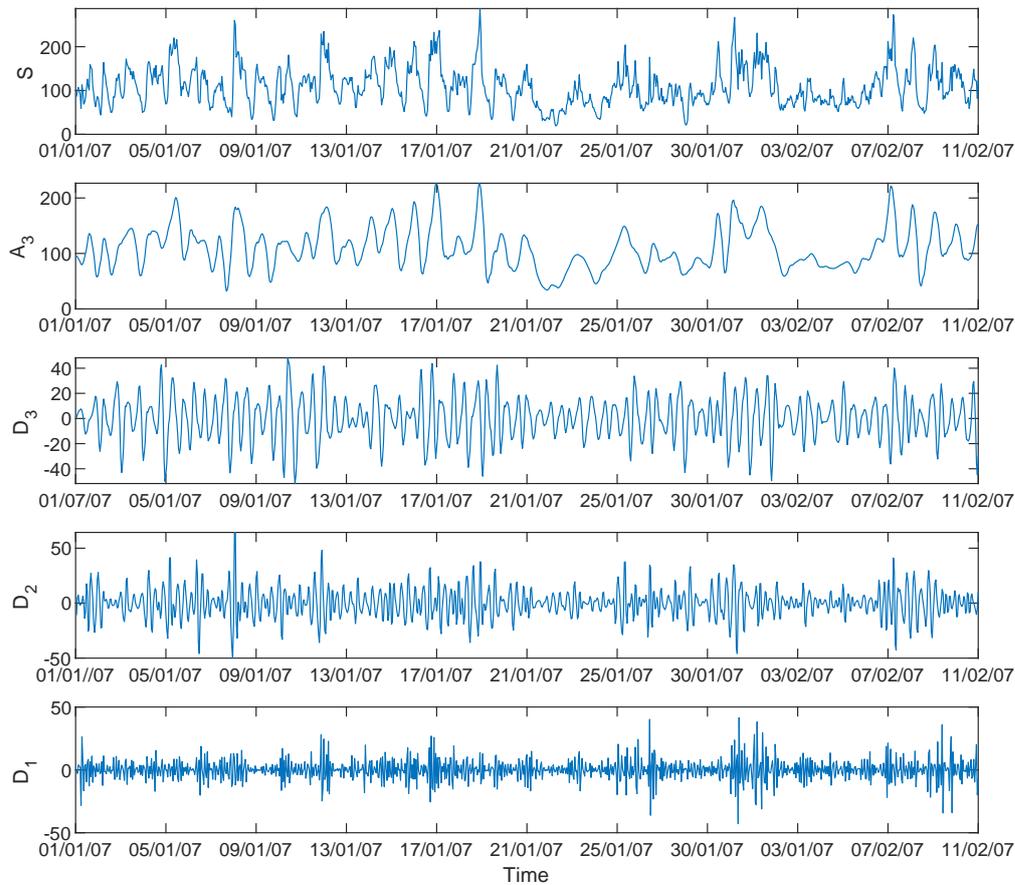
Fig. 3.10. Example of time series decomposition for NO_2 concentrations in $\mu g/m^3$ recorded at MY1 site using Daubechies db4 level-3 mother wavelet, where D_1 to D_3 are the detail signals and A_3 is the level 3 approximation signal.

3.3.3 Long Short-term Memory Neural Network (LSTM) Model

LSTM neural networks are special types of RNNs in which self-connected neurons are used to allow a cyclic flow of information. The LSTM neural network was proposed by Hochreiter and Schmidhuber (1997).

LSTM neural networks consist of one input layer, one output layer, and a series of memory blocks. Each LSTM block is composed of one or more self-recurrent memory cells and three multiplicative units, i.e. input, output and forget gates. Each gate provides continuous analogs of read, write and reset operations for the blocks, respectively. The blocks enable the network to accumulate enough information to update their train-

ing parameters. These mechanisms allow LSTM models to extract long-range temporal dependencies of inputs in the form of sequential series. An LSTM memory block with a single cell is illustrated in Figure 3.11.
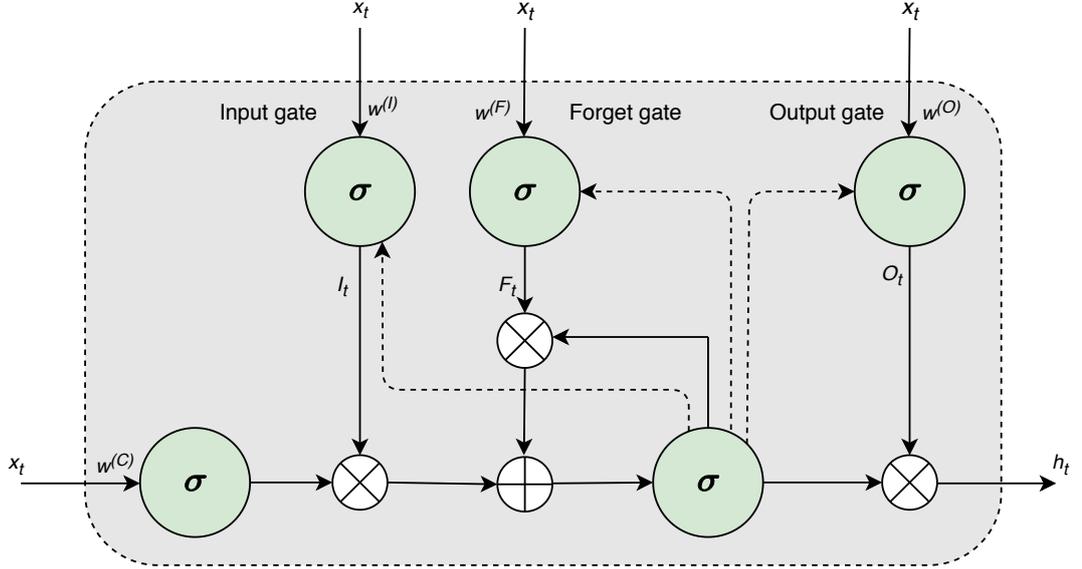
Fig. 3.11. A diagram of a single LSTM memory block

The input gate allows incoming information to modify the state of the nodes. The output gate permits or impedes the cell state from affecting other neurons. The forget gates were designed to learn and reset memory cells once their status is out of date, thereby preventing the cell status from growing without bounds and causing saturation of the transfer functions.

The forward training process of an LSTM unit with x_t, c_t and h_t as the input, memory cell status and output values, respectively, can be formulated as described in Eq. (3.25) to Eq. (3.29):

$$F_t = \sigma \left(w^{(F)} \cdot x_t + w^{(H)} \cdot h_{t-1} + w^{(C)} \cdot c_{t-1} + b^{(F)} \right) \tag{3.25}$$

$$I_t = \sigma \left(w^{(I)} \cdot x_t + w^{(H)} \cdot h_{t-1} + w^{(C)} \cdot c_{t-1} + b^{(I)} \right) \tag{3.26}$$

$$c_t = F_t \cdot c_{t-1} + I_t \cdot \sigma \left(w^{(C)} \cdot x_t + w^{(H)} \cdot h_{t-1} + b^{(C)} \right) \tag{3.27}$$

$$O_t = \sigma \left(w^{(O)} \cdot x_t + w^{(H)} \cdot h_{t-1} + w^{(C)} \cdot c_t + b^{(O)} \right) \tag{3.28}$$

$$h_t = O_t \cdot \sigma\left(c_t\right) \tag{3.29}$$

where I_t, O_t, and F_t are the function values of the input, output and forget gates at time t, respectively, c_t and h_t represent the activation vector for each cell and memory block, respectively, σ denotes the transfer function, and w and b are the weighting and bias constants. Commonly used transfer functions for the LSTM model include the *sigmoid*, *tanh* and *relu* functions.

3.4 Spatiotemporal modelling of hourly NO$_2$ levels in Greater London using fully-connected wavelet-based LSTM models

3.4.1 Framework of the fully-connected Hybrid Deep Learning Approach

The proposed modelling approach is presented in this section and is depicted in Figure 3.12. The proposed modelling approach consists of four parts, each of which is briefly summarised as follows:

(1) The k-optimal neighbouring sites of each target site are selected based on the Pearson correlation coefficient, $r\left(\cdot\right)$, of the concentration values between sites s_i and s_j:

$$r\left(s_i, s_j\right) = \frac{Cov\left(s_i, s_j\right)}{\sigma\left(s_i\right)\sigma\left(s_j\right)}, \tag{3.30}$$

where $Cov\left(\cdot\right)$ denotes the covariance function, and $\sigma\left(\cdot\right)$ is the standard deviation. A correlation threshold value is initially set, and the k-optimal sites with correlation values higher than the threshold value are chosen. Hence, the number of neighbouring sites for each target site may vary. It is also worth noting that the most relevant neighbouring sites may not necessarily be the nearest ones due to the presence of other meteorological factors and human activities (Wang and Song, 2018). Hence, the distance constraint is not considered in the selection of the neighbouring sites.

The selection process is carried out to reduce the tendency of model overfitting due
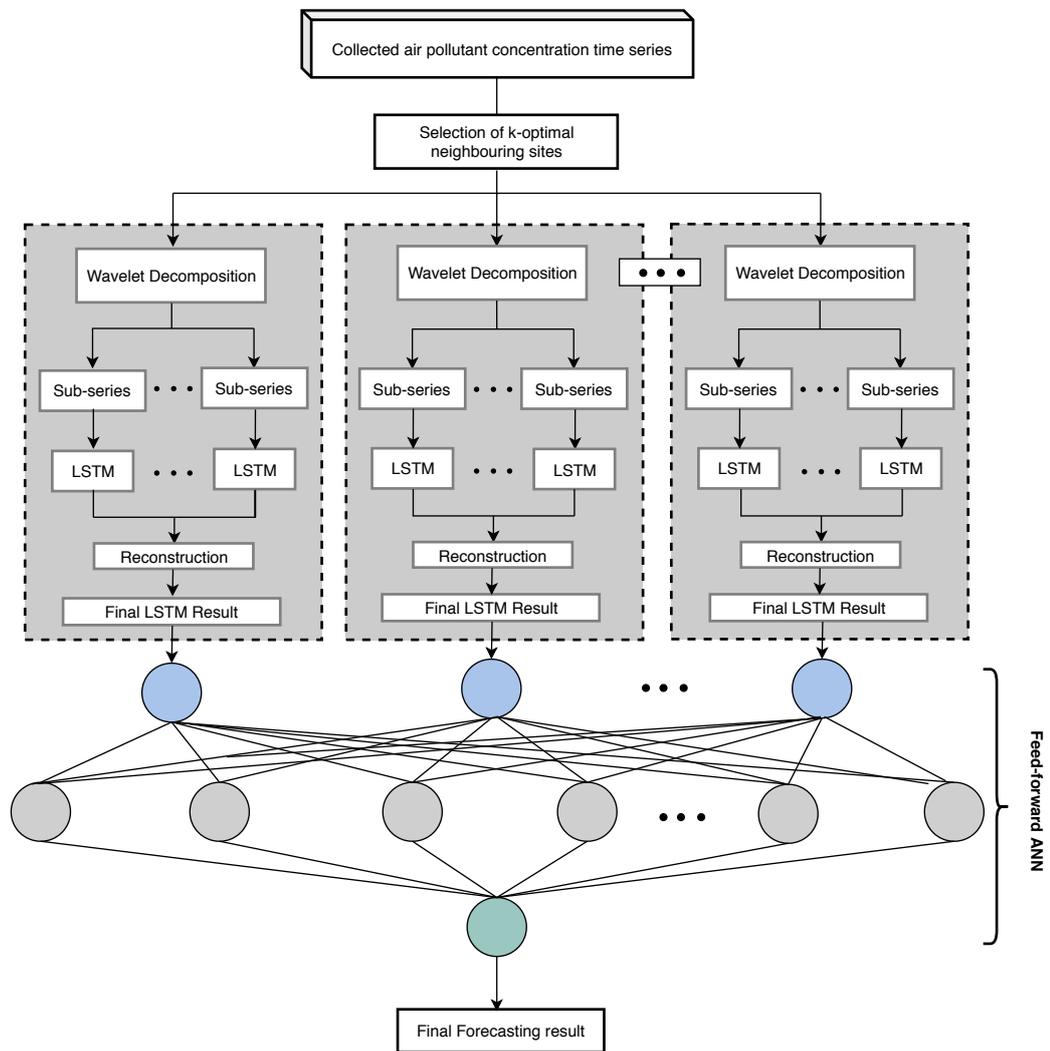
Fig. 3.12. The framework of the proposed fully-connected hybrid LSTM modelling system in Experiment 3.

to the utilisation of too many irrelevant or redundant predictors (Bishop, 1995). It is worth noting that as much as previous works have reported the advantages of utilising other auxiliary predictors such as meteorological and traffic variables (Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019), this work only considered the past concentration values from the selected neighbouring sites. This is to ensure that the results are influenced not by the relationship between the utilised predictors but only by the past NO_2 data and complexity of the proposed model architectures. Furthermore, the said

scheme attempts to assess the spatiotemporal performance of hybrid ANN architectures using only minimum number of predictors.

(2) The hourly AP concentration data from the selected k neighbouring sites are then decomposed into several sub-series through DWT (see Section 3.3.1).

(3) Each decomposed data, i.e. the subseries, from each selected sites is then fed to a single temporal approximator represented by the LSTM models (see Section 3.3.3). LSTM models are selected to extract temporal information from the input time series. The individual LSTM model estimates are then reconstructed using Eq. (3.21) to form the final LSTM result. Note that the said results represent the initial estimates (or pre-estimates) of the pollutant concentration at the target station by the LSTM models using only the information from the k neighbouring sites.

(4) The LSTM pre-estimates are then fed to a fully-connected ANN represented by the MLP model (see Section 3.2.5). The MLP is chosen to extract information regarding the spatial variation of the individual forecasts from the neighbouring sites. The output of the feed-forward model is considered as the final forecasting result of the proposed hybrid spatiotemporal wavelet-based model.

3.5   Summary

The Chapter has described the methods that are applied to implement the various experiments needed to achieve the objectives of this research. The schematics for each experiment were discussed in detail, while the individual techniques needed to carry out each component of the said schematics were fully described.

The basic components such as the network architecture, hidden node configurations, and calibration schemes of a standard feedforward ANN model, e.g. the MLP model, were fully covered. Sophisticated recurrent variant of the MLP model, e.g. the LSTM model, was also covered.

The basic theory behind the proposed predictor selection techniques, e.g. stepwise regression, NCA with regularisation, SFS and CART, were also discussed in this chapter. Furthermore, distinctions between filter-, wrapper-, and ensemble-type techniques were described according the mechanisms of the selected predictor selection techniques.

Chapter 3. Methodology

Finally, the underlying principles into developing data-intensive hybrid and ensemble ANN models were discussed. For instance, the development of hybrid ANN models that underwent wavelet preprocessing and various ensemble modelling approaches were fully covered in this chapter.

Chapter 4

Experimental Design and Model Development

4.1 Introduction

In this chapter, the settings of the experiments in this thesis are presented. Details including the model-building protocol, data ratification and monitoring site types, the multi-step forecasting scheme, and performance evaluation indices are first covered in Section 4.2, while the settings of the experiments mentioned in Chapter 1 Section 1.6 are described in Sections 4.3 through 4.5.

4.2 An overview of the experimental settings

4.2.1 Model-building protocol

As discussed in Chapter 2 Section 2.2.2, the development of ANN models generally consists of the following steps: (1) data collection, (2) data preprocessing, (3) predictor selection, (4) data splitting, (5) selection of model architecture selection, (6) model structure determination, (7) model training, and (8) model validation. The settings of each experiments in the following sections are therefore described according to the said steps. The choice of certain parameters and techniques that were incorporated in the experiments are also justified in the succeeding sections.

67

4.2.2 Data collection and site location details

A special emphasis on the details regarding the collected data and case study locations are provided due to the data-driven nature of ANN models. There are a couple of technical terms relating to the locations where the data have been obtained that need to be defined. Such details include site environment types and data ratification practices.

First, several site environment types were considered in the following experiments. A detailed description of various air quality monitoring site types is provided in Table 4.1. Common instruments used for the continuous monitoring of O_3 concentrations include the Dobson spectrophotometer and the Brewer spectrophotometer, while NO_2 concentrations are monitored using the chemiluminescent analyser (DEFRA, 2004). Fine particles such as PM_{10} and $PM_{2.5}$ are measured using the Tapered Element Oscillating Microbalance (TEOM) analysers Model 1400AB with different sampling heads design, Filter Dynamics Measurement System (FDMS) and β-attenuation analysers (Charron and Harrison, 2005).

Finally, the data presented online by DEFRA are either marked as *provisional* or *fully ratified*. Hourly mean monitoring results from the Automatic Urban and Rural Monitoring Network (AURN) are uploaded as provisional data. Data validation is then undertaken in which the provisional data are screened and "cleaned". Additional reviews and screenings on a quarterly basis are then carried out to exclude faulty values as far as possible (DEFRA, 2017). When provisional data have been quality-assured, they are considered fully ratified or verified. All data utilised in all the experiments in this research have been fully ratified, and are therefore fit for computational analyses and interpretation (DEFRA, 2017). A screenshot of DEFRA's online data selector interface is shown in Appendix C.

Table 4.1. Description of various site environment types (DEFRA, 2004).

Type	Description
Urban area (U)	Area with complete or at least highly predominant building-up of the street front side by buildings with at least two floors or large detached building with at least two floors. With the exception of city parks, the built-up area is not mixed with non-urbanised areas. For the measurement of O_3, the urban area is described as residential or commercial areas of cities, parks (away from trees), big streets or squares with very little or no traffic, open areas characteristic of educational, sports or recreation facilities. Urban sites should measure air quality which is representative of a few km^2.
Suburban area (S)	Area with contiguous settlement of detached buildings of any size with a lesser building density than those in urban areas. The built-up is mixed with non-urbanised areas such as agricultural, lakes and woods. For the measurement of O_3, suburban areas include where population, sensitive crops or natural ecosystems located in the outer fringe of an agglomeration are exposed to high ozone levels. Suburban sites should measure air quality which is representative of some tens of km^2.
Rural area (R)	Area situated more than 20 km away from agglomerations and more than 5 km away from other built-up areas, industrial installations or motorways major roads. Rural sites should measure air quality which is representative of at least 1000 km^2.
Traffic station (T)	Area located such that its pollution level is determined predominantly by the emissions from nearby traffic, such as roads, motorways, and highways. Traffic sites should measure air quality representative of a street segment no less than 100m length. Furthermore, sampling probes shall be at least 25m from the edge of major junctions and more than 10m from the kerbside.
Industrial station (I)	Area located such that its pollution level is influenced by emissions from nearby single industrial sources or industrial areas with many sources. Industrial stations should measure air quality representative of at least 250m $\times$ 250m.
Background station (B)	Area located such that its pollution level is not influenced significantly by any single source or street, but rather by the integrated contribution from all sources upwind of the station, e.g. by all traffic, combustion sources, etc. upwind of the station in the city. Background sites should measure air quality representative for several km^2.

4.2.3   Forecasting horizon

The structure of a multi-step ahead forecasting scheme which was adapted in the experiments is given in Figure 4.1. It can be seen that some related measurements in the previous hours is utilised by the model. The parameter h refers to the forecasting

horizon. It can be seen that the farther out the forecasting horizon, i.e. the greater the value of h, the lesser samples become available due to the time shifting required. As such, the total amount of samples available for training and training could be calculated as total samples $= (N - h)$, where N is the total number of samples. On the other hand, the parameter d refers to the delay or lag, which determines how much data in the past are used for model training.
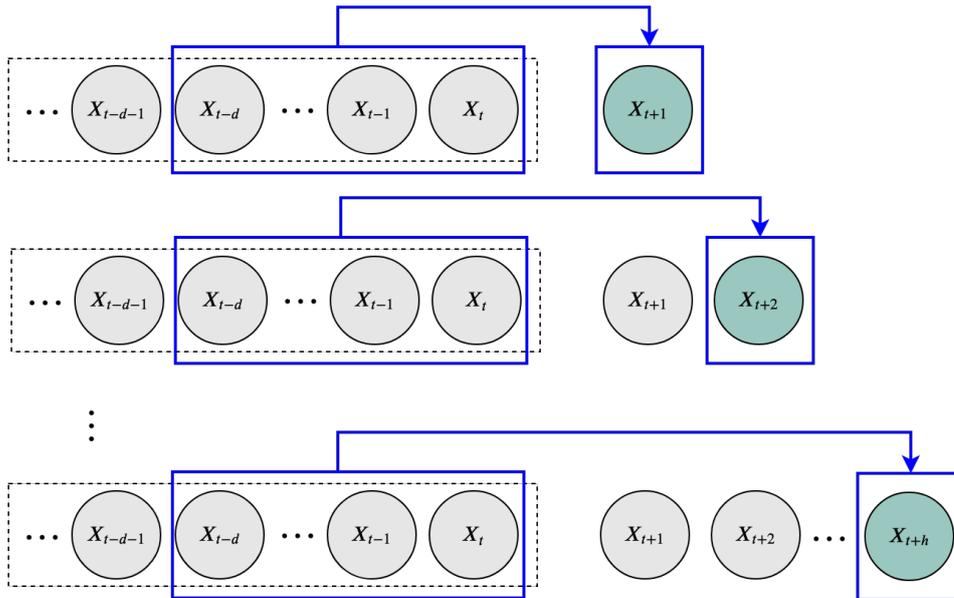
Fig. 4.1. The structure of the multi-step ahead forecasting scheme.

4.2.4 Performance Indices

The models that were developed in all the experiments are validated using various comprehensive performance metrics. In detail, the replicative and predictive validity of the model results are assessed to ensure that the models can be utilised for their intended purpose with confidence.

As thoroughly discussed by Humphrey et al. (2017), replicative validation metrics ensure that the model has captured the underlying relationship in the training data, while predictive validation metrics assess the generalisation ability of the model.

The following statistical metrics were employed: the RMSE, MAE, and Index of Agreement (IA) or Willmott's index (Willmott et al., 2012). The said metrics are

described in Table 4.2. The RMSE is one of the most popular indices based on the squares of the differences between actual and modelled output values, while MAE is based on the absolute differences between actual and modelled output values. Lower RMSE and MAE values translate to better model performances. IA is a dimensionless index that is more rationally accounts the accuracy of model results than any other existing indices. In fact, the use of the more traditional linear correlation coefficient, R, and its square, R^2, are not recommended as a measure of model performance as their magnitudes are not consistently related to the accuracy of model results (Willmott, 1982). The metric values are bounded by -1.0 and 1.0, and should be closer to 1.0 for best-performing models.

Table 4.2. Model performance metrics.

Abbreviation	Definition	Formula
RMSE	Root Mean Squared Error	$RMSE = \sqrt{\dfrac{1}{N}\sum_{i=1}^{N}(P_i - O_i)^2}$
MAE	Mean Absolute Error	$MAE = \dfrac{1}{N}\sum_{i=1}^{N}|P_i - O_i|$
IA	Index of Agreement	$IA = 1 - \dfrac{\sum_{i=1}^{N}(P_i - O_i)^2}{\sum_{i=1}^{N}\left[|P_i - \bar{O}| + |O_i - \bar{O}|\right]^2}$

Note that P_i and O_i are the predicted and observed values of NO_2 concentration, respectively, and $\bar{P}$ and $\bar{O}$ are the mean value of the predicted and observed values of the predictand, respectively.

The models developed in Experiment 2.2 are further assessed by determining their ability to accurately forecast hourly AP limit exceedances or hourly episodes, see the contingency table shown in Table 4.3 The columns are the forecast values and the rows are the actual values. In the matrix,

- TN is the number of hourly episodes correctly identified as an episode;

- FP is the number of hourly episodes incorrectly identified as an episode, e.g. false alarm;

- FN is the number of hourly episodes incorrectly identified as a non-episode, and

- TP is the number of hourly episodes correctly identified as an episode.

The NO$_2$ level of 100 μg/m^3 was selected as it is found that levels beyond it are considered unhealthy for sensitive groups, including those with lung disease, children and older adults (USEPA, 2016).

Table 4.3. Contingency table for a two-category forecast.

		Forecast	
		< 100 μg/m^3	≥ 100 μ g/m^3
Actual	< 100 μg/m^3	TN	FP
	≥ 100 μg/m^3	FN	TP

Based on Table 4.3, several metrics can be calculated:

- probability of detection (POD) which represents the fraction of correctly forecast NO$_2$ limit exceedance, ranging between 0 to 1 and with the best value of 1; and

- false alarm rate (FAR) which is the fraction of false alarms over the total forecast positive events, ranging between 0 to 1 and with the best value of 0.

Ideally, the POD score should be reasonably high while the FAR score should be reasonably low to maintain public confidence in the NO$_2$ level early warnings. The said ratios are given by Eq. (4.1), and Eq. (4.2), respectively:

$$POD = \frac{TP}{FN + TP}, \tag{4.1}$$

$$FAR = \frac{FP}{FP + TP}. \tag{4.2}$$

Lastly, a visual inspections was undertaken to reveal the performance of the models. For instance, the following figures were shown: scatter plots of actual versus predicted data, plots of actual and predicted data, and histogram of residuals.

4.2.5 Software and Hardware Specifications

All computations such as the development of the ANN models and implementation of all preprocessing techniques are carried out in MATLAB R2019b environment (The

MathWorks, 2019). All spatial analyses and mapping were performed by ArcGIS software version 10.6 (Environmental Systems Research Institute, 2020). Furthermore, the computations were performed in an Intel(R) Xeon(R) E5-2620 CPU with 2.10GHz processors, 8 physical cores, 16 logical processors, NVIDIA Quadro K620 2 GB DDR3 GPU memory, and 16 GB memory.

4.3 Settings for Experiment 1

4.3.1 Collected data and site locations

The selected target site for Experiment 1 is the London Marylebone Road (MAR) which is a busy road comprising of three lanes of traffic in each direction and carrying approximately 90,000 vehicles per weekday (DEFRA, 2020). As such, MAR site is considered as one of the most polluted streets in the United Kingdom, exceeding annual legal limit of 40 μg/m^3 for NO$_2$ levels several times in the past four years (The Guardian, 2017). The dataset collected from MAR site thus exhibits a wide range of NO$_2$ levels, particularly numerous peak levels. As such, the dataset was as selected to assess the performance of the proposed models of all experiments in this thesis. MAR site is also expected to represent many other urban locations exhibiting a similar range of NO$_2$ concentration values. To account for the influence of the AP levels from a neighbouring site, measurements from the London Bloomsbury (BLM) site, which is approximately 2 km away from MAR site, were also collected. As shown in Figure 4.2, MAR and BLM sites fall under traffic and background types, respectively.

(a) MAR site   (b) BLM site

Fig. 4.2. Maps describing a) MAR and b) BLM sites (via Google Maps).

Continuous measurements of emissions and meteorological predictors were used to train the MLP models in Experiment 1. In more detail, hourly values of NO, NO_2, CO, PM_{10}, $PM_{2.5}$, O_3, barometric pressure (BP), temperature (T), wind direction (WD) and wind speed (WS) levels measured from January 2014 to December 2016 were collected. The set of predictors were initially selected based on availability and completeness. Furthermore, a combination of primary and secondary gaseous and particulate air pollutants were chosen as relationships between them are known to occur. The meteorological parameters were selected as they influence the dispersion of air pollutants in the ambient environment. Note that the wind-related variables were transformed into two components, namely, $W_x = WS \cos(WD)$, and $W_y = -WS \sin(WD)$, to account for their cyclic characteristic and avoid sudden jumps of values (Arhami et al., 2013).

Table 4.4 lists the selected predictors and their statistical properties for the specified collection period. Note that those ending with '_bg' denote the predictors collected from a background site, e.g. BLM site.

Table 4.4. Descriptive statistics of hourly air pollutant and meteorological data for the period January 2014 to December 2016 (DEFRA, 2004).

Predictor	Unit	Mean	Median	Std. Dev.	Min	Max	Missing data (%)
NO_2	$\mu g/m^3$	90.35	84.63	40.91	8.06	321.91	1.34
O_3	$\mu g/m^3$	14.81	10.28	13.18	0	98.69	3.06
NO	$\mu g/m^3$	142.19	108.19	115.37	1.53	872.83	1.34
CO	$\mu g/m^3$	0.52	0.47	0.27	0	2.51	6.56
PM_{10}	$\mu g/m^3$	25.44	22.70	14.14	0	118.50	6.19
$PM_{2.5}$	$\mu g/m^3$	16.63	14.40	10.88	0	97.10	3.37
WD	°	197.40	225.00	98.14	0	360.00	3.10
WS	m/s	3.42	3.10	1.68	0	12.60	3.10
T	°C	10.16	10.10	5.70	-6.70	29.60	3.10
BP	$\bar{m}$	1009.99	1011.00	9.90	976.00	1038.00	3.10
O_3_bg	$\mu g/m^3$	28.12	26.14	19.88	0	138.09	1.03
NO_bg	$\mu g/m^3$	18.91	8.66	29.71	0.06	432.42	1.49
NO_2_bg	$\mu g/m^3$	46.92	45.23	21.21	0.48	192.43	1.54
PM_{10}_bg	$\mu g/m^3$	19.40	16.10	12.61	0	147	12.28
$PM_{2.5}$_bg	$\mu g/m^3$	12.51	9.40	10.71	0	96.50	4.62
WD_bg	°	202.33	229.80	92.36	0	360	3.01
WS_bg	m/s	3.34	3.00	1.72	0	13.10	3.01
T_bg	°C	10.34	10.30	5.75	-6.80	29.40	3.01

The collected NO_2 data from MAR site has the mean and median concentrations of

90.35 μg/m^3 and 84.63 μg/m^3, respectively. The statistics reveal the relatively higher pollution levels in MAR site when compared to those collected from BLM site. Furthermore, the emissions data in MAR site exhibit higher variability than those from BLM site according to the computed standard deviation values of the NO$_2$ data from both sites. Finally, it is evident that the collected dataset suffer from varying amounts of missing values, ranging from 1.03% to 12.28%.

Figure 4.3 displays the complete (and imputed) time series of the hourly NO$_2$ data, the parameter to be estimated in the current experiment. Note that the Data Augmentation (DA) was used to address the issue of missing data (see Section 4.3.2 for more details). The collected hourly NO$_2$ levels appear to vary rapidly and cover a wide range of values between short periods of time. Due to the complexity of the dynamics between NO$_2$ pollution and other parameters, a more comprehensive visual inspection needs to be done. The time series plots of the predictors utilised in Experiment 1 can be found in Appendix E.
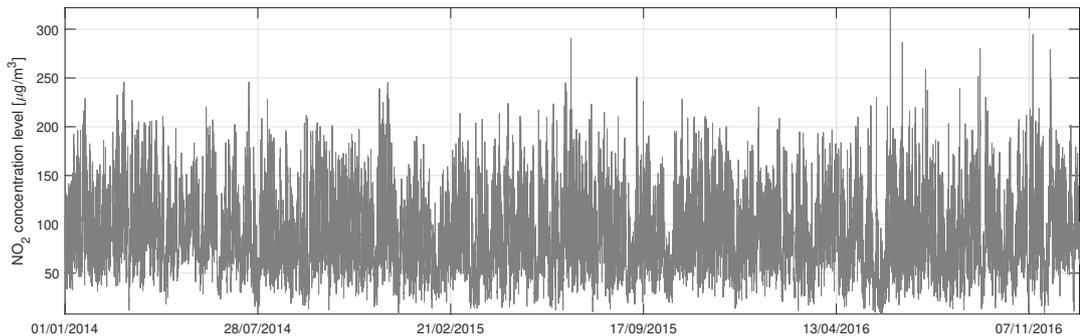
Fig. 4.3. Hourly NO$_2$ measured in MAR site during from January 2014 to December 2016.

Figure 4.4 shows the plot of the same hourly NO$_2$ values over hourly (at the top) and daily (at the bottom) cycles. Higher mean values and amplitudes take place from 7am to 10am and from 3pm to 6pm in comparison to the other times of the day. The said observed peaks represent peak travel hours demonstrating the strong influence of road transportation especially from diesel vehicles on roadside NO$_2$ levels (DEFRA, 2004; Colls, 2001; WHO, 2003). The said differences in the mean concentration level

and the amplitude of the hourly time series indicate the presence of hourly seasonality. Furthermore, the box plot of the collected data over daily cycles exhibits seasonality of NO_2 levels. For instance, higher levels are observed on weekdays while lower ones are seen during on weekends.
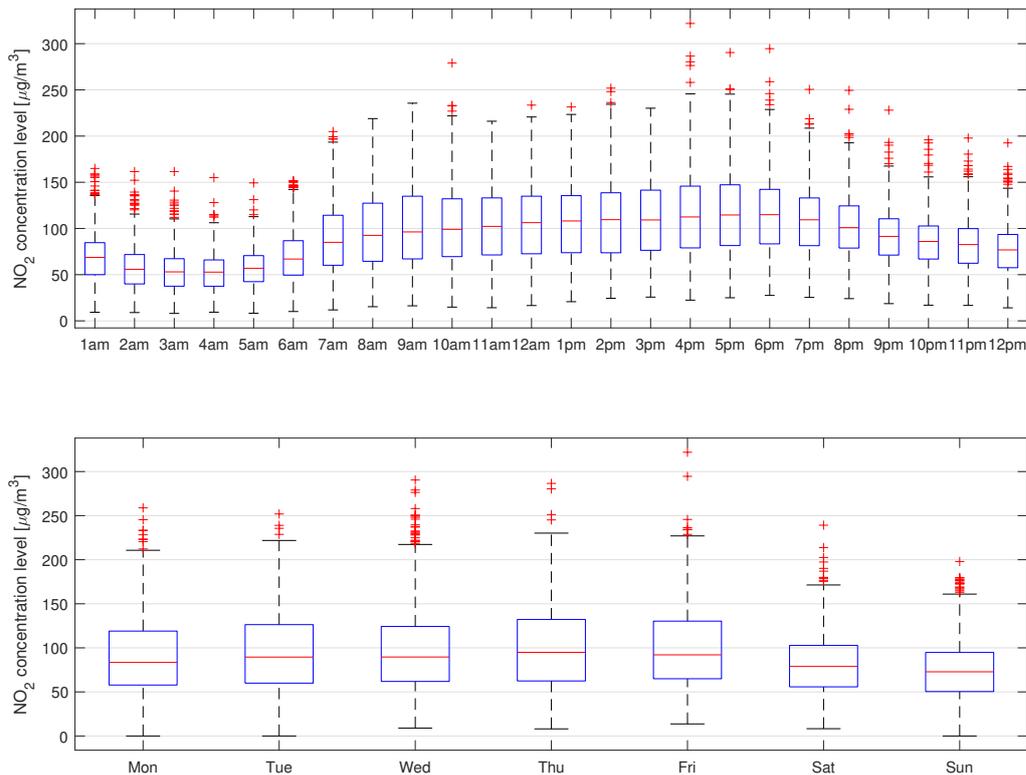
Fig. 4.4. Hourly box plots of the collected NO_2 data for each hour from January 2014 to December 2016.

Given the said findings, temporal predictors such as hour of the day, e.g. $1, 2, \ldots, 24$, and day of the week, e.g. $1, 2, \ldots, 7$, were also included in the candidate set of predictors. However, such values introduce abrupt changes especially when fed to sigmoidal transfer function. To avoid discontinuity and account for their cyclic trend, the temporal predictors were transformed using the following expressions: $\sin(2\pi h/24)$, $\cos(2\pi h/24)$, $\sin(2\pi d/7)$, and $\cos(2\pi d/7)$, where h is the hour of the day, and d the day of the week (Chelani et al., 2002).

76

4.3.2   Model Development

The Data Augmentation (DA) algorithm (Tanner and Wong, 1987) was applied to address the missing data. It consists of an iterative process that alternatively fills in missing data and makes inferences about the unknown parameters in a random manner. Details regarding the algorithm can be found in Appendix D. Furthermore, the max-min data normalisation was performed to ensure that all values fall between zero to one (Hagan et al., 1995). Given a predictor set X, the technique can be described as follows:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}, \tag{4.3}$$

where x_{norm} is the normalised value of $x \in X$, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of X, respectively.

Several MLP models based on different sets predictors selected by stepwise regression, NCA with regularization, SFS, and regression trees techniques were built in this experiment. In addition, a MLP model that use all set of collected predictors was also developed to serve as a benchmark. Table 4.5 shows the settings used to implement the said predictor selection techniques.

Table 4.5. List of parameter settings utilised by the predictor selection techniques.

Technique	Parameter	Value
Stepwise Regression	Intercept term	Constant
	Distribution	Normal (default)
	Criterion	SSE
NCA with regularisation	Regularisation (λ)	Optimised via 5-fold CV partition
	Fitting method	Fitting using all data
	Weight solver	Stochastic Gradient Descent
	Loss function	MSE
	Gradient tolerance	1e-4
	Epsilon value	(Interquartile range of Y)/13.49
	Iteration limit	30
SFS	No. of CV partition	10-fold
	Objective function	MSE of MLP model results
Regression Tree	Leaf Merge Flag	On
	Minimum parent size	10
	No. of bins	Empty
	Predictor Selection	Curvature Test

Furthermore, a plain MLR model was also developed to serve as linear benchmark to the employed non-linear models in this experiment. Note that the only stepwise regression was applied to identify the best predictors for the MLR model. Table 4.6 lists all the models built in Experiment 1.

Table 4.6. List of proposed and benchmark models developed in Experiment 1.

Model Code	Model Details
MLP_{SR}	MLP model using the predictors selected by Stepwise Regression
MLP_{NCA}	MLP model using the predictors selected by NCA with regularisation
MLP_{SFS}	MLP model using the predictors selected by SFS
MLP_{RT}	MLP model using the predictors selected by Regression Tree
MLP_{ALL}	MLP model using all predictors
MLP_{MET}	MLP model using only meteorological predictors
MLP_{POL}	MLP model using only emission-type predictors
MLP_{TIME}	MLP model using only time-scale predictors
MLR_{SR}	MLR model using the predictors selected by Stepwise Regression

Both 10-fold cross-validation and block splitting approaches were applied to train and evaluate the performance of the developed models. In detail, all points during the collection period, e.g. from January 2014 to December 2016, were utilised for the training and testing of the model under 10-fold cross-validation. On the other hand, the first 90% of the collected hourly data were allocated to train the models while the remaining points were used for validation purposes under block splitting.

The MLP-based models built in this experiment have the structure P-H-1, where P denotes the number of predictors used, H is the number of nodes in the hidden layer, and 1 output node. Only one hidden layer was applied as it was found to be sufficient in approximating any smooth measurable mapping between predictors and predictands (Hornik et al., 1989). The value of H is determined through a trial-and-error approach in which several models with various hidden node configurations were run 100 times to account for the sensitivity of the initial random values of the weight and bias parameters. The value of H that yields the least average residual scores of the model using the training set was then chosen for the model testing stage. Finally, the log sigmoid transfer function was chosen for the hidden layers, and the linear function for the output layer.

The MLP models in this experiment were trained using the LM algorithm. In

addition, the forecasting horizon values used in this experiment are $h = 0, 1, 2, 3$. Lastly, the models were evaluated using the statistical indices described in Table 4.2.

4.4 Settings for Experiment 2

4.4.1 Collected data and site locations

The models employed in Experiment 2 were developed based on the air quality data collected in Central London. For the Experiment 2.1 which involves the temporal modelling of NO_2 levels, predictors from MAR site were utilised. The selection of the optimal predictors is based on the results of Experiment 1. On the other hand, several monitoring sites surrounding MAR site were chosen in Experiment 2.2 dealing with the spatiotemporal modelling of NO_2 levels. The number of sites considered for investigation was limited to five, representing those closest to MAR site. As such, the following sites were selected: (1) London Marylebone Road, (2) London Bloomsbury, (3) Camden Kerbside, (4) North Kensington, (5) Tower Hamlets Roadside, and (6) London Westminster (see Figure 4.5).
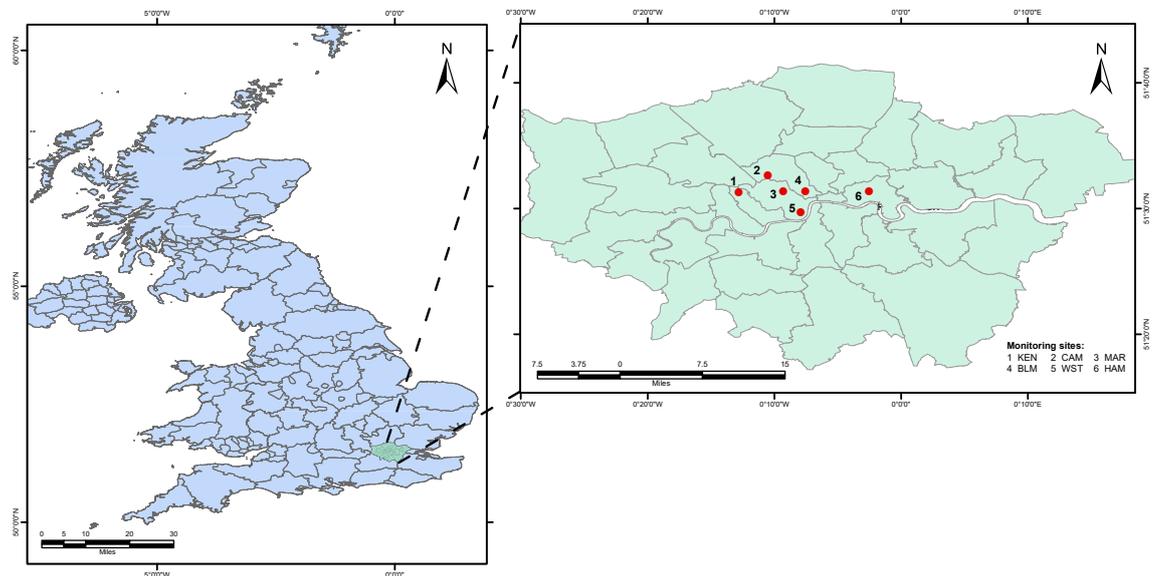
Fig. 4.5. Case study area.

The number of sites considered for investigation was limited to five, representing those closest to MAR site. The neighbouring sites were also selected based on data availability and site environment type. In more detail, only urban type monitoring sites and those with missing values of less than or equal to five percent of the total number of data from January 2013 to December 2014 were selected. A total of 105,120 data points were collected in total, e.g. 17,520 data points from each chosen site.

Table 4.7 provides the location and description of the selected monitoring sites, while the maps describing the selected sites and their vicinity are given in Figure 4.6.

Table 4.7. Air quality monitoring sites, coordinates, and site environment type.

Site name	Site Code	Latitude / Longitude	Environment type
London Marylebone Road	MAR	51° 31 ′ 21.10 ″ N, 0° 9 ′ 16.59 ″ W	Urban traffic
London Bloomsbury	BLM	51° 31 ′ 20.24 ″ N, 0° 7 ′ 33.20 ″ W	Urban background
Camden Kerbside	CAM	51° 32 ′ 39.15 ″ N, 0° 16 ′ 30.96 ″ W	Urban traffic
North Kensington	KEN	51° 31 ′ 15.78 ″ N, 0° 12 ′ 48.57 ″ W	Urban background
Tower Hamlets Roadside	HAM	51° 31 ′ 21.10 ″ N, 0° 2 ′ 31.75 ″ W	Urban traffic
London Westminster	WST	51° 29 ′ 40.81 ″ N, 0° 7 ′ 54.92 ″ W	Urban background

All selected monitoring stations are part of the AURN that has been monitoring AP levels since 1997 (DEFRA, 2004). The key characteristics of the said sites are described as follows:

1) MAR is a kerbside site located within one meter of the edge of a busy six-lane road, A501. Its surrounding area forms a canyon;

2) KEN is a background site situated at a mainly residential area and is 5 meters away from a quiet residential road, St. Charles Square. CAM is a kerbside site situated at the southern end of a broad street canyon where the road is often busy;

3) BLM is a background site situated within the north-east corner of a central London garden with all four sides surrounded by a two-lane one-way road system. The site is surrounded by small buildings;

4) WST is a background site situated in the car park of a building, Westminster Coroner's Court, 17 meters away from an intersection between two-lane roads, B323 Horseferry Road and Regency Street. The site is surrounded by a mix of commercial and residential areas; and

(a) MAR **(b)** BLM

(c) CAM **(d)** KEN

(e) HAM **(f)** WST

Fig. 4.6. Maps of the selected sites and their vicinity (via Google Maps).

5) HAM is a kerbside site situated within an existing building, part of Queen Mary and Westfield College, on a busy dual carriageway road, A11 Mile End Road. Its surrounding area consists of commercial and residential buildings.

Table 4.8 presents the measure of central tendency and dispersion of the collected NO_2 concentration data from all chosen sites. It is evident that the NO_2 concentration measured from MAR, CAM and HAM sites during the period 2013-2014 suffer from high NO_2 concentrations with recorded mean values of 89.18 $\mu g/m^3$, 68.10 $\mu g/m^3$ and 61.29 $\mu g/m^3$. In fact, breaches of the legal limit were observed 127 and 56 times at MAR and CAM sites, respectively, while a breach was recorded only twice at HAM site. On the other hand, the annual mean NO_2 concentration levels at KEN, BLM and WST sites

were significantly lower, e.g. 35.63 μg/m^3, 51.46 μg/m^3, and 45.40 μg/m^3, respectively, compared to those from the roadside sites. Finally, the missing NO$_2$ concentration data in this study ranged from approximately 0.5% to 2.0%.

Table 4.8. Descriptive statistics of the collected hourly NO$_2$ concentration data.

	Sites					
	MAR	KEN	CAM	BLM	WST	HAM
Mean [μg/m^3]	89.18	35.63	68.10	51.46	45.40	61.29
Median [μg/m^3]	82.99	31.03	62.46	50.27	43.62	59.36
Maximum [μg/m^3]	280.74	173.73	368.86	192.43	174.72	237.79
Minimum [μg/m^3]	7.25	0	5.56	1.14	0.29	0.92
Standard deviation [μg/m^3]	40.30	21.32	35.58	22.47	22.57	28.63
Missing data [%]	1.32	1.63	0.47	1.26	1.03	1.79

The observations above are in agreement with the box plots of the hourly variations of the collected NO$_2$ concentration data shown in Figure 4.7. The time series plots of the collected NO$_2$ data can be found in Appendix H. Higher mean NO$_2$ concentration values and range are exhibited by the data collected from MAR, CAM and HAM sites in comparison to those shown by the data from BLM, KEN and WST sites. Results from using the said collected datasets with a wide range of characteristics could provide valuable information revealing the influence of site location types on model performance.
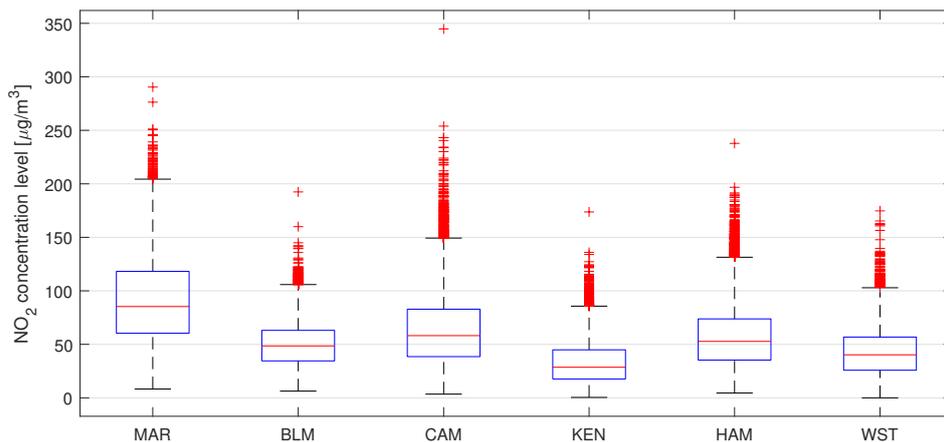
Fig. 4.7. Hourly box plots of the collected NO$_2$ data from the selected monitoring sites during the two-year study period.

4.4.2   Model development

In this experiment, the schemes employed in the data preprocessing in the previous experiment are also applied. Experiment 2 can be split into two main sub-experiments, namely, temporal modelling, and spatiotemporal modelling. As such, there are major distinctions regarding the predictors selected for the models of each sub-experiment.

For the temporal modelling tasks, the schemes applied for prediction selection in Experiment 1 adapted. In contrast, leave-one-out cross-validation methodology and correlation analysis were carried out to train the spatiotemporal models. In more detail, the information from the neighbouring sites will be used to estimate the pollutant level of a given target site. For instance, the NO_2 levels in MAR site will be modelled using only NO_2 levels from the neighbouring sites. The aim of this approach is to determine to ability of the proposed wavelet-based ANN models estimate the spatiotemporal variations of air pollutant levels.

A similar ANN architecture, e.g. a feedforward ANN (the MLP model), was also chosen as benchmark models for both sub-experiments. However, a deep RNN, e.g. LSTM model, was developed in Experiment 2.2 for the spatiotemporal estimation of NO_2 levels.

Data-intensive hybrid models that underwent wavelet preprocessing were proposed to test their superiority over the plain MLP models in both sub-experiments. As described in Chapter 3 Section 3.3.1, a wavelet decomposition technique is applied to split the original time series exhibiting high variability into several subseries with lower variability. Daubechies wavelets were chosen to approximate the raw time series signal for various reasons. Firstly, Daubechies wavelets are a family sophisticated wavelets capable of approximating continuous signals more accurately with fewer fixed scaling functions (Nievergelt, 2013). Secondly, Daubechies wavelets have been found to perform well in the past (Siwek and Osowski, 2012; Osowski and Garanty, 2007; Dunea et al., 2015). The names of the Daubechies family wavelets are usually written as DbN_v, where N_v is the number of vanishing moments which determine the ability a wavelet to approximate any given signal (Nievergelt, 2013).

Several types and levels of the Daubechies wavelet were tested and selected based on the ratio $\mathrm{std}\left(A_j\right)/\mathrm{std}\left(S\right)$. That is, the standard deviation of A_j must be substantially smaller than that of the original time series S. However, choosing a larger value of J increases the number of terms in Eq. (3.19), thus accumulating more approximation errors when Eq. (3.19) is computed (Osowski and Garanty, 2007). As such, the levels of the tested Db wavelets were limited from 3 to 5. On the other hand, the number of vanishing moments, N_v, was limited from 4 to 6. Considering the above-mentioned conditions, various configurations of Daubechies wavelets were selected.

The use of wavelets has implications on the resulting structure of the ANN models. Since several wavelets are applied to decompose each time series, the hidden layer configurations of resulting ANN models corresponding to the number of wavelets need to optimised. The same trial-and-error procedure for determining the hidden layers for the MLP models in Experiment 1 was applied here.

All models except the LSTM model are trained and tested using the same process described in Chapter 4 Section 4.3.2. The LSTM model trained using the 'adam' training algorithm. Finally, the forecasting horizon values for Experiment 2.1 are $h = 0, 1, 6, 12$. Only 1-h ahead forecasts are to be carried out in Experiment 2.2.

The list of models to be developed in Experiment 2.1 are as follows: W-MLP model, e.g. the wavelet-based MLP temporal model, and MLP model, e.g. the plain MLP temporal model.

On the other hand, the list of models to be developed in Experiment 2.2 is provided in Table 4.9. Note that CA denotes correlation analysis.

Table 4.9. List of proposed and benchmark models developed in Experiment 2.2.

Model code	Model description
W-LSTM-CA	Wavelet-based LSTM with predictors selected via CA
W-LSTM	Wavelet-based LSTM with all predictors
LSTM-CA	Plain LSTM with predictors selected via CA
LSTM	Plain LSTM with all predictors
W-MLP-CA	Wavelet-based MLP with predictors selected via CA
W-MLP	Wavelet-based MLP with all predictors
MLP-CA	Plain MLP with predictors selected via CA
MLP	Plain MLP with all predictors

4.5 Settings for Experiment 3

4.5.1 Collected data and site locations

The proposed models were developed using the hourly NO_2 concentration data from monitoring sites across the Greater London area. Figure 4.8 shows the distribution of the selected air quality monitoring sites across the Greater London area.
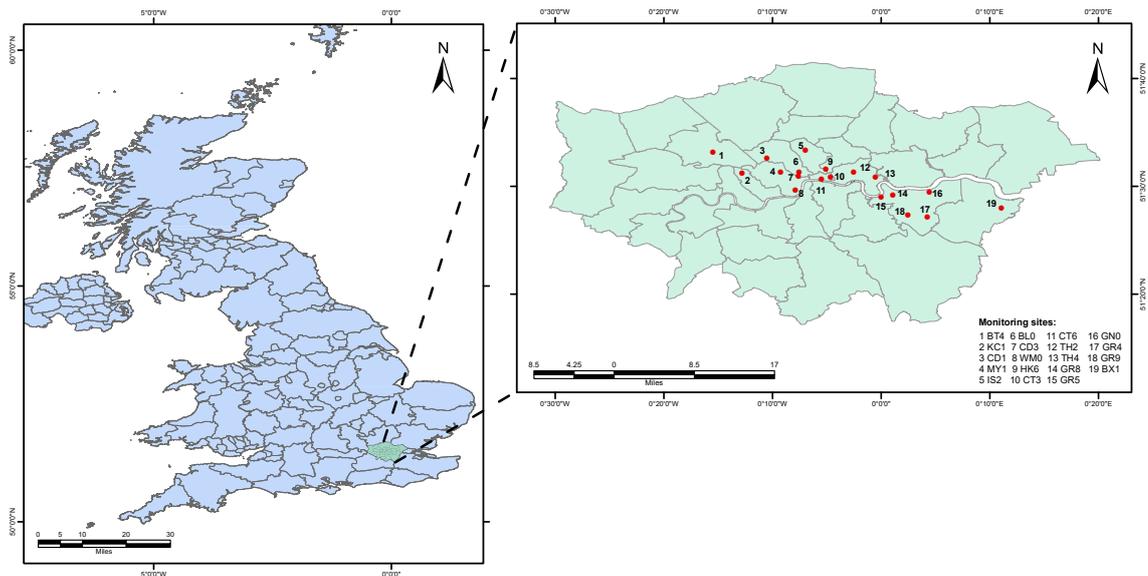
Fig. 4.8. The selected monitoring sites for Experiment 3.

The monitoring sites are part of the AURN of the United Kingdom, and maintained by the London boroughs, DEFRA, and Transport for London (TfL). Only urban or suburban monitoring sites with at least 80% of available NO_2 concentration data were considered. As such, 19 monitoring sites were selected, of which approximately 74% are considered as strategic for AP level assessments by Moorcroft and Marner (2011). As shown in Table 4.10, the selected sites consist of two kerbside, eleven roadside, two suburban, and four urban background sites.

Table 4.10. Air quality monitoring sites, coordinates, and site environment type (King's College London, 2019).

Site name	Site code	Site type	Latitude / Longitude
Bexley - Slade Green	BX1	Suburban	51° 27 ′ 57.71 ″ N, 4° 11 ′ 5.30 ″ E
Brent - Ikea	BT4	Roadside	51° 33 ′ 8.98 ″ N, 0° 15 ′ 29.86 ″ W
Camden - Bloomsbury	BL0	Background	51° 31 ′ 20.24 ″ N, 0° 7 ′ 33.20 ″ W
Camden - Shaftesbury Avenue	CD3	Roadside	51° 30 ′ 55.92 ″ N, 0° 7 ′37.30 ″ W
Camden - Swiss Cottage	CD1	Kerbside	51° 32 ′39.16 ″ N, 0° 10 ′30.97 ″ W
City of London - Sir John Cass	CT3	Background	51° 30 ′49.85 ″ N, 0° 4 ′39.96 ″ W
City of London - Walbrook Wharf	CT6	Roadside	51° 30 ′ 37.80 ″ N, 0° 5 ′ 29.8 ″ W
Greenwich - A206 Burrage Grove	GN0	Roadside	51° 29 ′ 25.92 ″ N, 0° 4 ′ 26.4 ″ E
Greenwich - Trafalgar Road	GR5	Roadside	51° 29 ′ 0.42 ″ N, 0° 0 ′0.52 ″ W
Greenwich - Westhorne Avenue	GR9	Roadside	51° 27 ′ 22.90 ″ N, 0° 2 ′ 26.59 ″ E
Greenwich - Woolwich Flyover	GR8	Roadside	51° 29 ′ 12.78 ″ N, 0° 1 ′ 4.44 ″ E
Greenwich - Eltham	GR4	Suburban	51°27 ′9.32 ″ N, 0°4 ′15.28 ″ E
Hackney - Old Street	HK6	Roadside	51° 31 ′ 35.23 ″ N, 0° 5 ′ 5.68 ″ W
Islington - Holloway Road	IS2	Roadside	51° 33 ′ 19.41 ″ N, 0° 6 ′ 58.06 ″ W
Kensington and Chelsea - North Ken	KC1	Background	51° 31 ′ 15.78 ″ N, 0° 12 ′ 48.57 ″ W
Tower Hamlets - Blackwall	TH4	Roadside	51° 30 ′ 54.17 ″ N, 0° 0 ′ 30.31 ″ W
Tower Hamlets - Mile End Road	TH2	Roadside	51° 31 ′ 21.11 ″ N, 0°2 ′ 31.76 ″ W
Westminster - Horseferry Road	WM0	Background	51° 29 ′ 40.81 ″ N, 0° 7 ′ 54.95 ″ W
Westminster - Marylebone Road	MY1	Kerbside	51° 31 ′ 21.11 ″ N, 0° 9 ′ 16.59 ″ W

The kerbside and roadside sites are located within 1 and 12 meters from major roads. Many of the sites, i.e. CD3, CD1, GR5, GR8, TH4, are situated near a junction while CT6, HK6, IS2 and MY1 sites are located in street canyons. Only BT4, GN0, GR9 and TH2 sites are placed in open areas. On the other hand, the urban background and suburban sites, i.e. CT3, BX1, BL0, GR4, KC1 and WM0, are mostly located on the outskirts of a city which are away from major AP sources.

Table 4.11 presents the descriptive statistics for the collected hourly NO_2 concentration data from the 19 selected monitoring sites within the period 2007 to 2012. The average hourly NO_2 concentrations at all sites are between 24.94 $\mu g/m^3$ at GR4 and 118.78 $\mu g/m^3$ at CT6. In all selected sites except BX1, GR4 and KC1, the mean NO_2 values fell above the EU limit value of 40 $\mu g/m^3$ annual average. The majority of the high average NO_2 levels are observed from roadside and kerbside sites while relatively low average NO_2 levels are exhibited by the urban background and suburban sites (Beckerman et al., 2008).

Table 4.11. Descriptive statistics for the collected NO_2 concentration data from the 19 selected monitoring sites.

Site code	Available Data (%)	Mean NO_2 (μg/m^3)	Median NO_2 (μg/m^3)	Min NO_2 (μg/m^3)	Max NO_2 (μg/m^3)	Standard deviation (μg/m^3)
BX1	95.48	33.22	29.00	-0.51	187.00	21.36
BT4	86.65	69.73	67.00	3.20	331.30	29.46
BL0	94.27	55.95	53.00	6.00	300.00	23.32
CD3	89.93	81.72	78.80	10.20	303.90	29.43
CD1	90.30	78.06	71.00	-8.88	390.00	43.84
CT3	96.01	55.32	53.00	6.10	247.65	23.49
CT6	85.81	124.27	117.70	9.60	367.00	49.50
GN0	96.60	52.65	49.00	-2.34	251.50	29.65
GR5	86.51	52.34	49.60	1.10	258.70	26.80
GR9	96.28	44.07	40.00	-1.05	209.50	27.58
GR8	97.29	74.32	69.50	0.10	310.40	39.00
GR4	93.11	26.56	21.00	-3.56	151.00	19.49
HK6	90.02	66.88	64.40	3.50	294.20	26.53
IS2	98.99	63.58	61.10	0.10	281.50	29.11
KC1	94.93	35.59	31.00	-1.95	401.00	23.57
TH2	81.49	63.56	61.00	2.00	300.00	30.66
TH4	95.08	63.35	60.35	4.80	257.10	27.05
WM0	93.85	41.33	38.00	-1.06	225.00	23.51
MY1	97.49	108.31	99.00	4.00	378.00	55.13

Figure 4.9 shows the box plots of the collected NO_2 concentration data from the selected monitoring sites across the Greater London area.
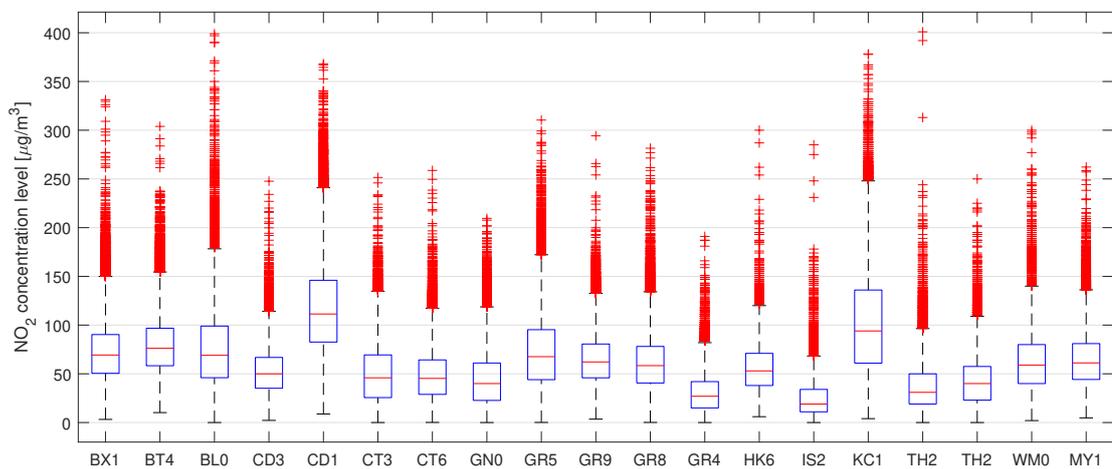
Fig. 4.9. Box plots of the collected NO_2 data from all 19 sites.

The data collected roadside sites clearly exhibit higher mean values, e.g. all greater

than 63 μg/m^3, and amplitudes than those collected from suburban or background sites. The roadside and kerbside sites are selected as target sites while the remaining sites are utilised as reference sites. The evaluation of the proposed modelling approach is performed through the use of the leave-one-out cross-validation methodology, where the pollutant concentration of each selected target site is estimated using the information from the neighbouring sites.

4.5.2 Model Development

The collected data in the current experiment are also preprocessed using the techniques applied in the previous experiments. To avoid model overfitting, 90% of the data were allocated for the training, while 10% was utilised for the testing set. For instance, data from 01/01/07 to 25/05/12 was used to train the model while the data from 26/05/12 to 31/12/12 was used to evaluate the performances of the model.

The architecture of the proposed model is a deep wavelet-based LSTM model, e.g. wavelet FC-LSTM model. On the other hand, the following models are developed to serve as benchmark models: (1) a plain feed-forward ANN model, i.e. MLP model, (2) a plain LSTM model, and (3) a wavelet-based LSTM model. Note that the benchmark models are built using the same parameter settings used to develop the proposed model.

The optimal network structure, e.g. hidden nodes and input lags, of both the LSTM and feed-forward ANN models are determined by trial and error. The optimum number of hidden layer nodes for the LSTM layer was selected between 100 to 200, and the optimum number of hidden layer nodes was searched between 10 to 45. Furthermore, A trial and error procedure was carried out by running the models 100 times to account for the sensitivity of the training algorithm to initial synaptic weights. The individual parameter setting that yields the least average forecasting error, i.e. root mean squared error (RMSE) value was selected. Finally, only 1-h ahead forecasts are to be carried out in Experiment 3.

The parameter settings of the modelling schemes are shown in Table 4.12.

Table 4.12. Parameter setting of the developed models.

Modelling process	Parameter setting
Wavelet-based LSTM model	Input node: $S_i(t)$; Input lag: 4; Output node: $\hat{S}_i(t)$;
	LSTM: single layer; LSTM node: $[100, 125]$;
	Transfer functions: tanh (state) and sigmoid (gate);
	Solver: 'adam'; Train epoch: 100; Gradient threshold: 1;
	Initial learning rate: 0.005; Learn rate drop period: 125;
	Drop factor: 0.2.
Feed-forward ANN model	Input node: $\hat{S}_i(t)$; Input lag: 12; Output node: $\hat{y}_i(t+1)$;
	Hidden layer: single; Hidden node: $[30, 40, 45]$;
	Transfer functions: tanh (hidden) and linear (output);
	Training algorithm: Levenberg-Marquardt; Train epoch: 1000
	Initial weights and bias values: random values between -1 and 1.

4.6 Summary

This Chapter has described the settings applied in developing the proposed and benchmark models in this research. The site locations and data utilised in training the ANN models were discussed. More than two years worth of hourly NO_2 measurements at several roadside locations in Central London were considered in the proposed experiments set out in Chapter 1.5. The application of the specific model development protocol described in Chapter 2.2.2 was also documented in the current chapter. The differences model configurations between experiments, e.g. predictors utilised, network architecture, hidden layer configurations, etc., were fully described and justified. Finally, the performance metrics applied in assessing the predictive and replicative validity of the results of the proposed models were discussed.

Chapter 5

Results and Discussion

5.1 Introduction

As detailed in Chapter 4, several experiments were conducted to determine the effectiveness of state-of-the-art ANN-based models in forecasting outdoor AP levels. The following sections provide the results of each step of the model development protocol presented in Figure 2.5. It is worth noting that the results of the preceding experiments were utilised in the succeeding experiments which will be thoroughly discussed in the following sections.

5.2 Experiment 1 results

In the current experiment, four predictor selection techniques have been applied to identify the best predictors from a candidate set of 21 predictors (see Table 4.4). Several variants of MLP models were then trained based on the various selected predictor subsets. In addition, benchmark MLP models based on a pre-defined set of predictors were also developed to test whether the use predictor selection techniques improve the performance of plain MLP models.

5.2.1 Predictor Selection

Table 5.1 shows the predictors that were identified by several predictor selection techniques discussed in Chapter 3 Section 3.2. Note that the complete results of the pre-

dictor selection techniques are given in Appendix F. From an initial set of 21 potential predictors, forward SR technique identified W_x, T_bg and HoD_x as insignificant predictors. NCA with regularisation technique eliminated CO, PM_{10}, PM_{10}_bg, W_x_bg, T_bg and HoD_x as predictors. SFS technique considered a combination of emissions, meteorological and temporal-based dummy variables, e.g. O_3, NO, W_x, T, O_3_bg, and HoD_x, as significant model inputs. Lastly, RT technique only identified O_3, NO, CO, $PM_{2.5}$, T, O_3_bg, and HoD_x as significant predictors. For completeness, the predictors of the benchmark models other than MLP_{ALL} were also shown in the table.

Table 5.1. Predictors identified by the predictor selection techniques.

Predictors	Predictor selection techniques						
	SR	NCA	SFS	RT	MET	POL	TIME
O_3	x	x	x	x		x	
NO	x	x	x			x	
CO	x			x		x	
PM_{10}	x					x	
$PM_{2.5}$	x	x		x		x	
W_x		x	x		x		
W_y	x	x			x		
T	x	x	x	x	x		
BP	x	x			x		
O_3_bg	x	x	x	x		x	
NO_bg	x	x		x		x	
NO_2_bg	x	x		x		x	
PM_{10}_bg	x					x	
$PM_{2.5}$_bg	x	x				x	
W_x_bg	x				x		
W_y_bg	x	x			x		
T_bg					x		
HoD_x			x				x
HoD_y	x	x		x			x
DoW_x	x	x					x
DoW_y	x	x					x

In summary, predictors such as O_3, T, and O_3_bg were identified as significant by all predictor selection techniques, while $PM_{2.5}$ was selected by at least three of the said techniques. The finding supports several results within the literature indicating that the concentration of any pollutant depends mainly on meteorological conditions and emission rates (Gorai and Mitra, 2017; Robeson and Steyn, 1990; Ziomas et al., 1995). The temporal-based dummy predictors such as HoD_x, HoD_y, DoW_x, and DoW_y

were also chosen by the said selection techniques. This finding reveals that the said predictors can provide information needed by the model to account for the variations in outdoor NO_2 emissions according to these time scales (Elangasinghe et al., 2014). In contrast, T_bg was considered irrelevant by all of the techniques, which could be explained by the presence of another variable, e.g. T, that already has a significant influence on the target variable.

5.2.2 Hidden Layer Optimisation

The results of the trial-and-error scheme for identifying the optimal number of hidden nodes according to the selected predictors are shown in Table 5.2. Note that all models share the same initial training parameters. In more detail, each model were implemented using the similar data splitting configurations, number of max iteration limits, initial weights and bias factors, and training algorithm.

Table 5.2. Final structure of the MLP-based models (where n_i, n_h and n_o denote the number of nodes in the input, hidden and output layers, respectively).

Model	Predictors utilised	$n_i - n_h - n_o$
MLP_{SR}	O_3, NO, CO, PM_{10}, $PM_{2.5}$, W_y, T, BP, O_3_bg, NO_*bg*, NO_2_*bg*, PM_{10}_bg, $PM_{2.5}$_bg, W_x_bg, W_y_bg, HoD_y, DoW_x, DoW_y	18-27-1
MLP_{NCA}	O_3, NO, $PM_{2.5}$, W_x, W_y, T, BP, O_3_bg, NO_bg, NO_2_bg, $PM_{2.5}$_bg, W_y_bg, HoD_y, DoW_x, DoW_y	16-25-1
MLP_{SFS}	O_3, NO, W_x, T, O_3_bg, HoD_x	6-28-1
MLP_{RT}	O_3, CO, $PM_{2.5}$, T, O_3_bg, NO_bg, NO_2_bg, HoD_y	8-28-1
MLP_{ALL}	All predictors	21-25-1
MLP_{MET}	W_x, W_y, T, BP, W_x_bg, W_y_bg, T_bg	7-28-1
MLP_{POL}	O_3, NO, CO, PM_{10}, $PM_{2.5}$, O_3_bg, NO_bg, NO_2_bg, PM_{10}_bg	10-27-1
MLP_{TIME}	HoD_x, HoD_y, DoW_x, DoW_y	4-20-1

The minimum and maximum number of hidden nodes were chosen according to the results of an initial study (Cabaneros et al., 2017). Table 5.2 indicates that most models exhibit minimal average RMSE values on the testing set using 25 to 28 hidden nodes. Furthermore, the findings reveal that a higher number of utilised predictors does not necessarily translate to higher number of optimal hidden nodes. This observation is demonstrated by the MLP_{NCA} and MLP_{ALL} models. The complete results are provided in Appendix G.

5.2.3 Model Validation Results

Table 5.3 shows the validation results of the developed ANN models in forecasting hourly NO_2 concentration levels in MAR site using the test sets. The third to fifth columns of the table show the performance metrics results in terms of RMSE (in $\mu g/\text{m}^3$), MAE (in $\mu g/\text{m}^3$) and IA, respectively, using the block data splitting configuration. The last three columns of the table show the same metrics but using 10-fold cross-validation.

The best forecasting results are achieved by the models that implemented predictor selection techniques. In detail, the MLP_{SFS} models yielded the best results, with IA scores ranging from 0.87 to 0.97 and from 0.93 to 0.98 using block splitting and 10-fold cross-validation, respectively. This finding indicates that a combination of emissions, e.g. O_3, O_3_bg and NO, meteorological, e.g. W_x and T, and temporal-based, e.g. HoD_x, variables can accurately approximate the hourly NO_2 levels in MAR site 0-h to 3-h in advance. The findings are in agreement with the results of previous related case studies within the literature (Cabaneros et al., 2019).

The MLP_{RT} models also provided satisfactorily superior results comparable to those from the MLP_{SFS} models. Although both MLP_{SR} and MLP_{NCA} models also exhibit highly accurate forecasts, MLP_{RT} and MLP_{SFS} models outperform the rest in terms of model parsimony. That is, the models provided accurate results while only utilising a minimal number of predictors. This is beneficial in some real-world applications where only a few predictors are available or measured accurately. Satisfactory results are also achieved by the MLP_{POL} models, indicating the significance of emission-type predictors in estimating the concentration levels of outdoor air pollutants such as NO_2 (Colls, 2001).

On the other hand, poor performance are exhibited by the MLP_{TIME} and MLP_{MET} with IA scores below 0.70 and RMSE and MAE scores higher than 36 $\mu g/\text{m}^3$ and 27 $\mu g/\text{m}^3$, respectively. This indicates that information either temporal-based or meteorological parameters alone are not enough in accurately modelling NO_2 levels.

Table 5.3. Predictive validation results of the developed models for the forecasting of hourly NO_2 levels.

Model	Horizon	Performance Criteria					
		RMSE	MAE	IA	$RMSE_{CV}$	MAE_{CV}	IA_{CV}
MLP_SR	+ 0 hr	15.862	11.753	0.964	13.641	10.114	0.969
	+ 1 hr	21.089	14.904	0.933	18.502	13.772	0.940
	+ 2 hr	24.502	17.573	0.908	22.283	16.842	0.905
	+ 3 hr	27.199	19.565	0.880	24.788	18.856	0.876
MLP_{NCA}	+ 0 hr	16.514	12.298	0.961	12.093	25.595	0.976
	+ 1 hr	21.378	15.360	0.931	17.325	25.610	0.948
	+ 2 hr	25.458	18.075	0.897	20.585	26.011	0.923
	+ 3 hr	27.783	20.264	0.873	22.476	26.395	0.904
MLP_{SFS}	+ 0 hr	13.754	10.120	0.972	11.855	8.534	0.977
	+ 1 hr	19.228	13.792	0.942	17.169	12.557	0.949
	+ 2 hr	23.551	17.248	0.905	20.355	15.102	0.925
	+ 3 hr	26.953	20.128	0.868	22.374	16.763	0.905
MLP_{RT}	+ 0 hr	15.901	12.054	0.965	12.311	8.911	0.975
	+ 1 hr	21.237	15.115	0.933	17.843	13.087	0.944
	+ 2 hr	24.879	17.925	0.901	21.793	16.313	0.911
	+ 3 hr	28.516	21.055	0.861	24.376	18.507	0.882
MLP_{ALL}	+ 0 hr	17.456	15.591	0.927	16.698	13.459	0.948
	+ 1 hr	22.159	18.434	0.899	21.708	17.182	0.912
	+ 2 hr	25.432	20.826	0.874	24.368	19.260	0.903
	+ 3 hr	27.351	23.438	0.856	26.086	20.651	0.878
MLP_{MET}	+ 0 hr	36.444	28.013	0.693	32.570	25.595	0.718
	+ 1 hr	36.028	27.771	0.701	32.424	25.610	0.722
	+ 2 hr	35.931	27.736	0.701	32.757	26.011	0.712
	+ 3 hr	36.076	27.948	0.695	33.173	26.395	0.698
MLP_{POL}	+ 0 hr	17.138	12.931	0.958	12.713	9.212	0.974
	+ 1 hr	21.077	14.922	0.931	18.732	13.843	0.938
	+ 2 hr	26.270	19.135	0.880	24.093	18.085	0.885
	+ 3 hr	30.642	23.036	0.821	27.785	21.162	0.831
MLP_{TIME}	+ 0 hr	37.575	30.590	0.657	33.219	26.500	0.701
	+ 1 hr	37.584	30.591	0.657	33.203	26.495	0.701
	+ 2 hr	37.580	30.592	0.657	33.211	26.497	0.702
	+ 3 hr	37.575	30.581	0.657	33.215	26.492	0.702

Figure 5.1 shows the plots of the first 500 samples of the 1-h ahead forecasting results of MLP_{ALL} and MLP_{SFS} models. The plots represent the results of the said models using the block data splitting scheme, e.g. the plotted values represent those hourly NO_2 concentration levels measured and modelled from 03/01/17 at 22:00 to 24/01/17 at 18:00. The plot of the forecasting results of the MLP_{SFS} model coincides

very well with the actual hourly NO_2 data, which is in agreement with the error indices in Table 5.3. In contrast, the MLP_{ALL} model clearly encounters some difficulties in approximating peak actual NO_2 levels measured at MAR site. Even the MLP_{SFS} model slightly fails in estimating several peak NO_2 concentration trends. This finding can be explained the inability of empirical models in general to capture extreme concentration levels due to the under-representation of these cases in the training data (Gong and Ordieres-Meré, 2016).
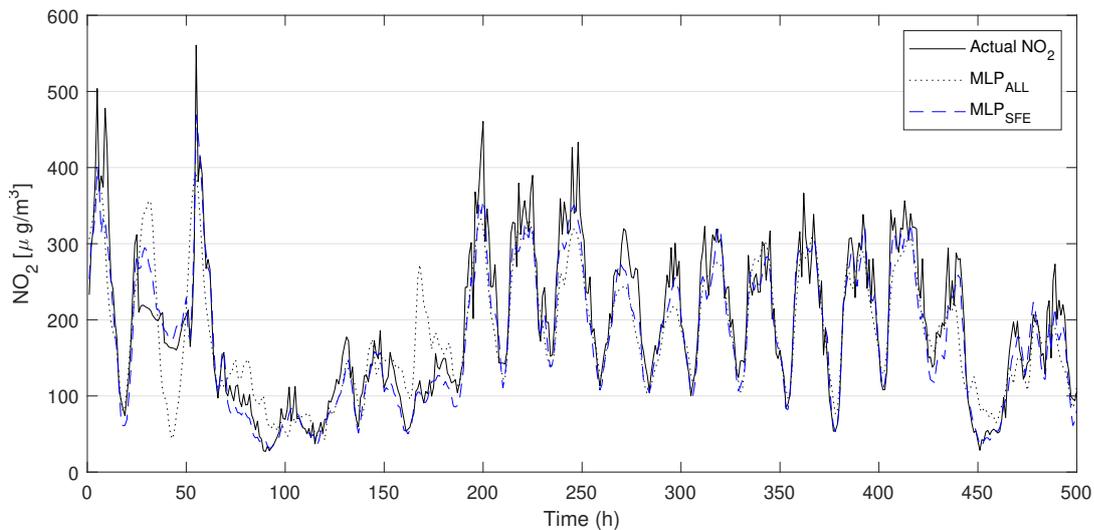
Fig. 5.1. Comparison of the 1h-ahead forecasting results of MLP_{ALL} and MLP_{SFS} models.

The findings above can also be said of the model results using the training sets (see Table 5.4). The best performance is also achieved by MLP_{SFS} models while the worst by the MLP_{TIME} models. The models are clearly able to replicate the intrinsic qualities of the training set for the actual NO_2 levels in MAR site. It is also apparent that the results in the table below are slightly better than those listed in Table 5.3. This finding expected as the training set contains more sample that enables the models to learn and perform better.

Table 5.4. Replicative validation results of the developed models

Model	Horizon	Performance Criteria					
		RMSE	MAE	IA	RMSE$_{\text{CV}}$	MAE$_{\text{CV}}$	IA$_{\text{CV}}$
MLP_SR	+ 0 hr	13.441	9.988	0.970	13.502	10.030	0.970
	+ 1 hr	18.328	13.720	0.939	18.266	13.624	0.941
	+ 2 hr	21.934	16.605	0.906	22.075	16.709	0.907
	+ 3 hr	24.483	18.678	0.875	24.476	18.649	0.879
MLP$_{\text{NCA}}$	+ 0 hr	11.526	8.376	0.978	11.686	8.525	0.978
	+ 1 hr	16.585	12.210	0.952	16.817	12.392	0.951
	+ 2 hr	19.618	14.608	0.929	19.911	14.800	0.928
	+ 3 hr	21.454	16.114	0.911	21.786	16.352	0.910
MLP$_{\text{SFE}}$	+ 0 hr	11.234	8.138	0.979	11.283	8.218	0.980
	+ 1 hr	16.394	12.054	0.953	16.531	12.156	0.953
	+ 2 hr	19.474	14.531	0.930	19.609	14.607	0.931
	+ 3 hr	21.313	16.044	0.913	21.526	16.177	0.913
MLP$_{\text{RT}}$	+ 0 hr	11.873	8.599	0.977	11.988	8.738	0.977
	+ 1 hr	17.268	12.756	0.947	17.455	12.879	0.947
	+ 2 hr	21.239	15.923	0.914	21.515	16.126	0.913
	+ 3 hr	23.725	18.038	0.885	23.902	18.171	0.887
MLP$_{\text{ALL}}$	+ 0 hr	14.134	10.056	0.950	15.218	11.147	0.940
	+ 1 hr	18.875	13.647	0.926	19.982	14.738	0.917
	+ 2 hr	21.465	15.621	0.908	22.621	17.753	0.909
	+ 3 hr	22.992	16.864	0.895	24.161	19.489	0.886
MLP$_{\text{MET}}$	+ 0 hr	31.942	25.113	0.723	32.390	25.462	0.721
	+ 1 hr	33.956	26.272	0.723	32.207	25.428	0.717
	+ 2 hr	35.220	28.574	0.714	32.555	25.849	0.696
	+ 3 hr	36.650	30.010	0.701	33.002	26.255	0.672
MLP$_{\text{POL}}$	+ 0 hr	12.253	8.855	0.975	12.498	9.087	0.975
	+ 1 hr	18.340	13.609	0.939	18.361	13.625	0.941
	+ 2 hr	23.568	17.741	0.888	23.650	17.797	0.890
	+ 3 hr	27.211	20.799	0.834	27.351	20.875	0.837
MLP$_{\text{TIME}}$	+ 0 hr	32.589	25.945	0.708	33.107	25.407	0.714
	+ 1 hr	34.594	26.950	0.698	34.111	26.413	0.703
	+ 2 hr	35.586	27.944	0.681	35.096	27.397	0.694
	+ 3 hr	39.582	30.139	0.657	37.993	29.393	0.677

The error histograms of both the 1-h forecasting results of MLP$_{\text{ALL}}$ and MLP$_{\text{SFS}}$ models are shown in Figure 5.2. The error distribution of the MLP$_{\text{SFS}}$ model is centered at around $25 \pm 5\mu\text{g/m}^3$ but with a smaller range, e.g. -60 $\mu\text{g/m}^3$ to 150 $\mu\text{g/m}^3$. This reveals that the model has less under- and over-predictions of the measured data. In contrast, the error distribution of the MLP$_{\text{ALL}}$ model tends to center around 0 $\mu\text{g/m}^3$

to 10 μg/m^3 but with a much wider range, e.g. from -150 μg/m^3 to 150 μg/m^3.
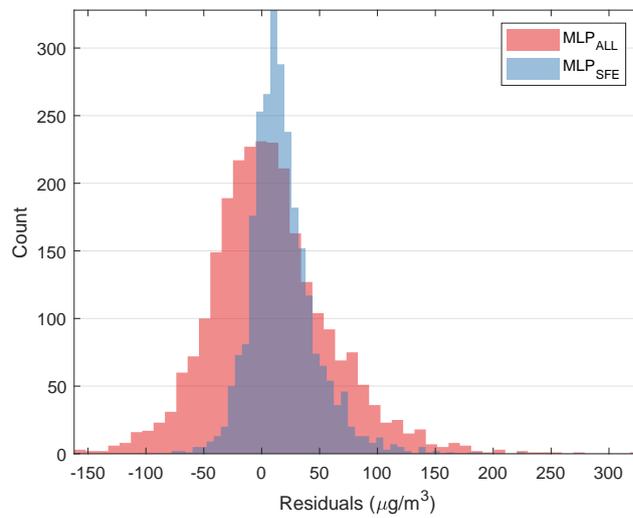
Fig. 5.2. Error histograms of the 1-h ahead forecasting results of MLP$_{\text{ALL}}$ and MLP$_{\text{SFS}}$ models.

Figure 5.3 and Figure 5.4 show the comparison of the 1-h ahead forecasting results of all MLP-based models and the actual NO$_2$ values from MAR site. The scatter plots reveal the effectiveness of the predictor selection techniques. In detail, the majority of the points yielded by the models that applied predictor selection are concentrated near the ideal fit. Conversely, the points representing the results of the MLP$_{\text{ALL}}$ model tend to disperse away from the ideal fit especially above 150 μg/m^3 values of the actual NO$_2$ data. Finally, the underwhelming performance of both MLP$_{\text{MET}}$ and MLP$_{\text{TIME}}$ models are clearly depicted in the scatter plots in which a lot of points are sparsely located at both high and low concentration ends.

(a) MLP$_{SR}$

(b) MLP$_{NCA}$

(c) MLP$_{SFS}$

(d) MLP$_{RT}$

(e) MLP$_{ALL}$

(f) MLP$_{MET}$
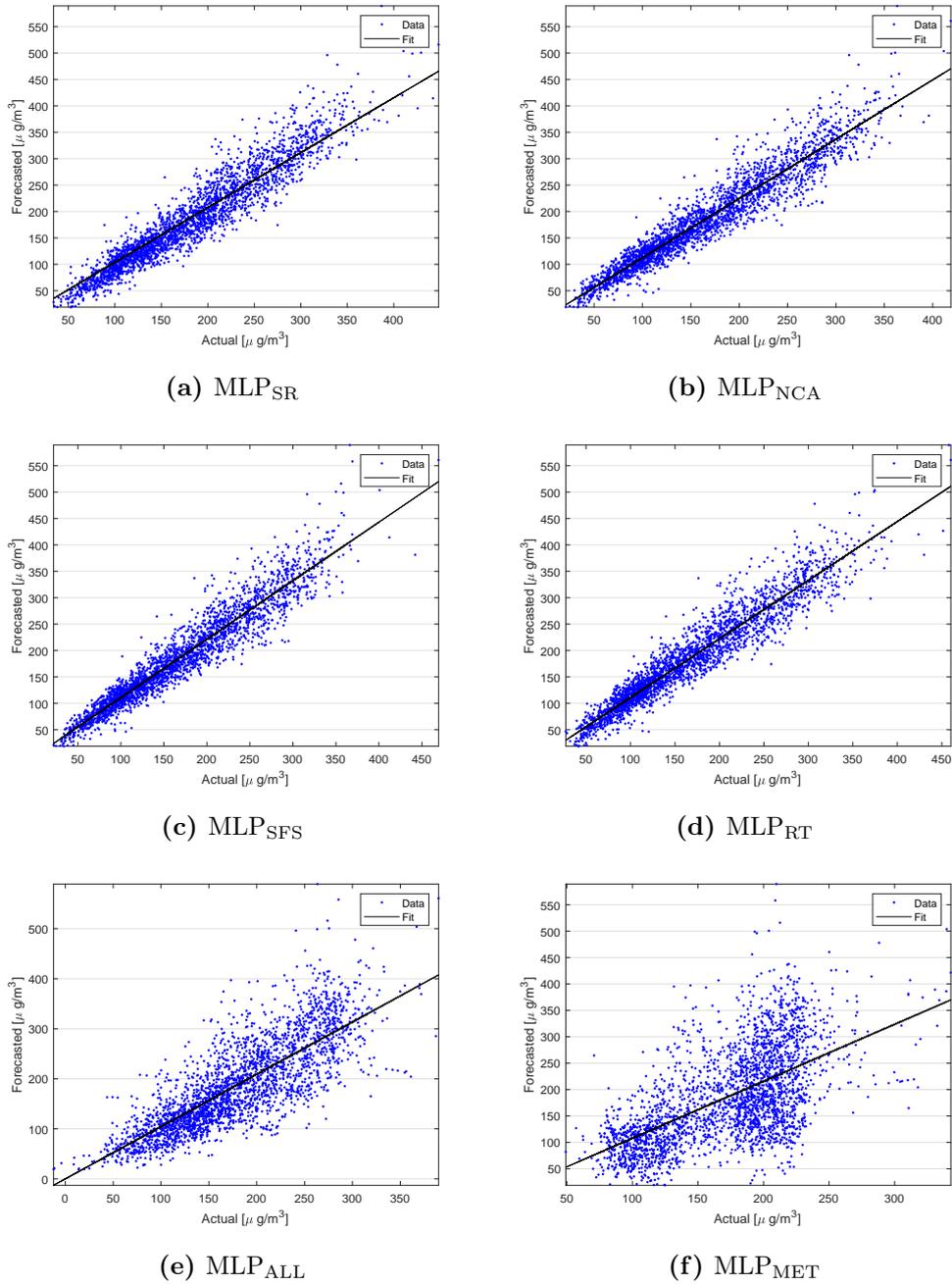
Fig. 5.3. Scatter plots of the 1-h ahead results of the developed MLP models.

Figure 5.5 and Figure 5.6 show the RMSE, MAE and IA values of all developed models across various forecasting horizon values. At the lower forecasting horizons, e.g. $h = 0$ and $h = 1$, the performance of the MLP models that employed predictor

(a) MLP$_{POL}$

(b) MLP$_{TIME}$
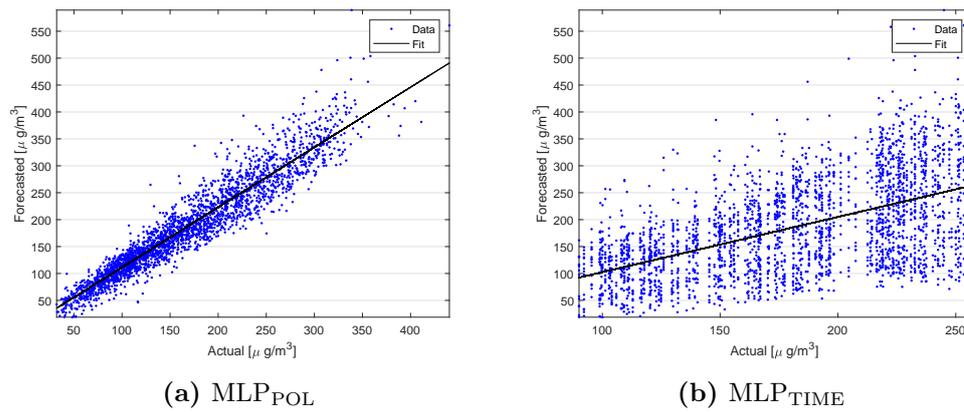
Fig. 5.4. (continued) Scatter plots of the 1-h ahead results of the developed MLP models.

selection is comparable. As with any forecasting tasks, however, the performance of all models tend to degrade as the forecasting horizon increases (Valput et al., 2019; Cabaneros et al., 2020; Coman et al., 2008). As shown in Figures 5.5a and 5.6a, the RMSE and MAE values of the models increase as forecasting horizons move from 0-h to 3-h. On the other hand, the decreasing accuracy of results of the developed models in terms of IA scores is also clearly exhibited as seen in Figure 5.6b.

(a) RMSE
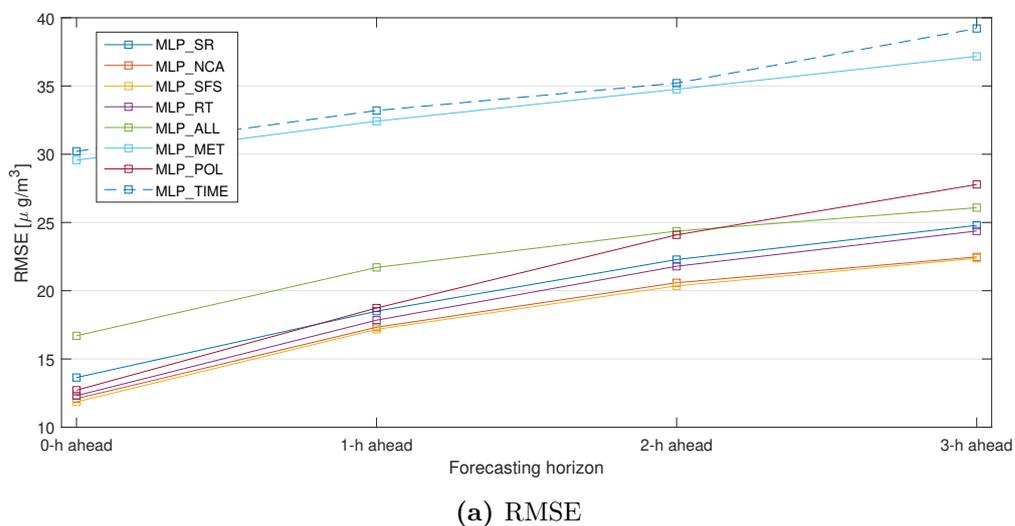
Fig. 5.5. Scatter plots of the 1-h ahead results of the developed MLP models.
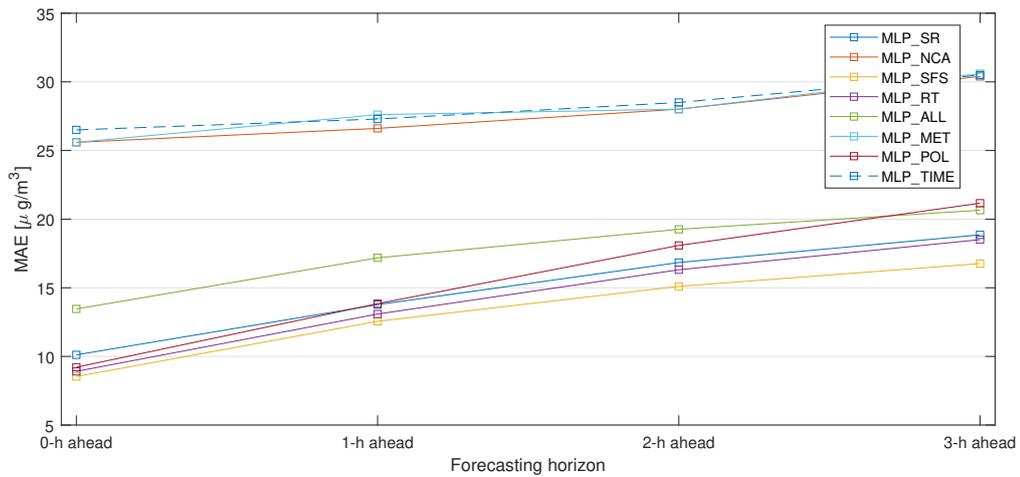
(a) MAE
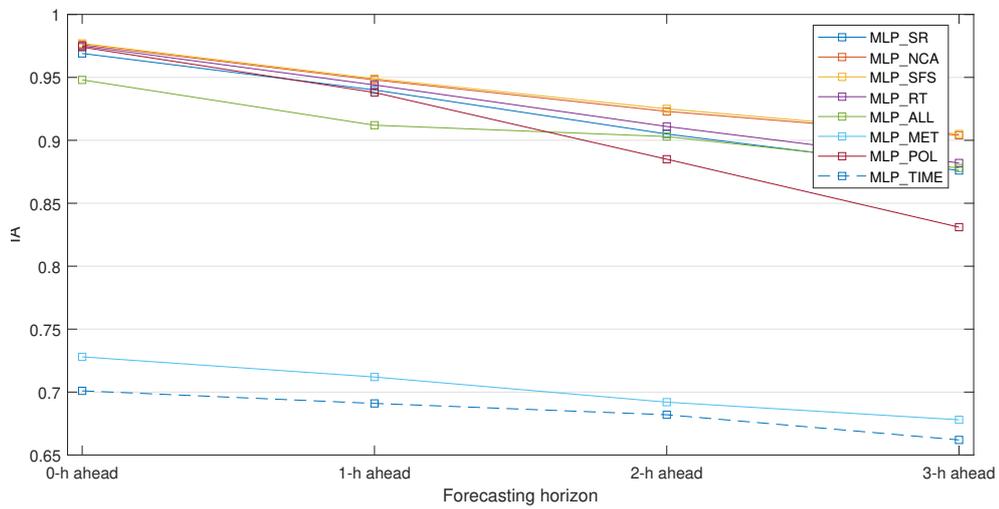
(b) IA

Fig. 5.6. (continued) Scatter plots of the 1-h ahead results of the developed MLP models.

The degradation of model performance results as the forecasting horizon increases is also illustrated in terms of error distributions. As shown in Figure 5.7, the range of errors of the MLP$_{SFS}$ model becomes larger as the forecasting horizon moves from 1-h to 3-h.
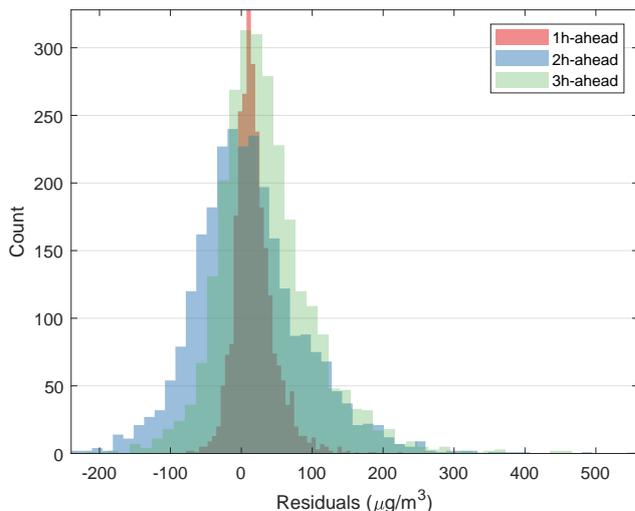
Fig. 5.7. Error histograms of the 1-h, 2-h, and 3-h ahead forecasting results of the MLP_{SFS} model.

To compare the MLP model performance with linear approaches, MLR models were also developed to serve as benchmark to the MLP models. The predictive and replicative validation results of the MLP_{SR} and MLR models are shown in Table 5.5.

Table 5.5. Predictive and replicative validation results of the MLP_{SR} and MLR models.

Model	Horizon	Predictive			Replicative		
		RMSE	MAE	IA	RMSE	MAE	IA
MLP_{SR}	+ 0 hr	15.862	11.753	0.964	13.441	9.988	0.970
	+ 1 hr	21.089	14.904	0.933	18.328	13.720	0.939
	+ 2 hr	24.502	17.573	0.908	21.934	16.605	0.906
	+ 3 hr	27.199	19.565	0.880	24.483	18.678	0.875
MLR	+ 0 hr	26.891	20.890	0.869	22.3733	17.361	0.8977
	+ 1 hr	28.717	22.279	0.842	24.9724	19.341	0.8637
	+ 2 hr	30.606	23.865	0.807	27.6189	21.408	0.8196
	+ 3 hr	32.234	25.300	0.774	29.4702	22.952	0.7809

The results indicate the superiority of the MLP models over the MLR models according to the RMSE, MAE and IA results. For instance, the RMSE and MAE values of the real-time predictions of the MLP_{SR} model are 69.53% and 77.74% lesser than those yielded by the MLR model. Moreover, the IA values of the said MLP model is 9.85%

higher than those from the MLR model. This finding clearly show the ability of MLP models in capturing the non-linearity in the relationship between several predictors and hourly NO_2 levels in MAR site that a linear model is unable to approximate (Gardner and Dorling, 1998; Elangasinghe et al., 2014). On the other hand, a sudden decrease in accuracy is exhibited by the MLR model results as the forecasting horizon moves from $h = 1$ to $h = 2$. Although the findings above indicate that all the accuracy of the MLP model forecasts suffer as the forecasting horizon increases, the performance of the developed MLR models seem to degrade more quickly. Although some studies suggest that the MLP models do not necessarily outperform traditional linear models such as the MLR model (Shahraiyni and Sodoudi, 2016; Cabaneros et al., 2019), the findings of the experiment clearly indicate the superiority of the former over the latter.

5.3 Experiment 2 Results

In the current experiment, temporal and spatiotemporal ANN models coupled with a preprocessing technique based on DWT have been developed. Experiment 2.1 focused on the development of the temporal models while the second on the spatiotemporal ones.

5.3.1 Daubechies Wavelet Configuration Determination

The results of the trial-and-error procedure for identifying the optimal configuration of the Daubechies wavelets are shown in Figure 5.8. The complete results can be found in Appendix I. For models with target sites MAR through HAM, level-5 Db5, Db4, Db4, Db6, Db6, and Db5 wavelets provided the least $\text{std}(A_j)/\text{std}(S)$ ratios, respectively. As such, the said wavelet configurations were used in the succeeding experiments. This finding means that six ANN models will be trained for every experiment. That is, the current experiment attempts to replace the modelling task of the original NO_2 time series of high variability by modelling its wavelet coefficients on different levels with lower variability. As such, five different ANN models will be built for the detailed coefficients, D_i, $(i = 1, 2, \ldots, 5)$ and one for the residual signal, A_5. The final forecast

of the NO_2 pollution level at any time point t is then calculated by applying Eq. (3.21). Since most of the wavelet coefficients exhibits lower variability, the method is argued to increase the total forecasting accuracy of the original NO_2 time series (Siwek and Osowski, 2012).
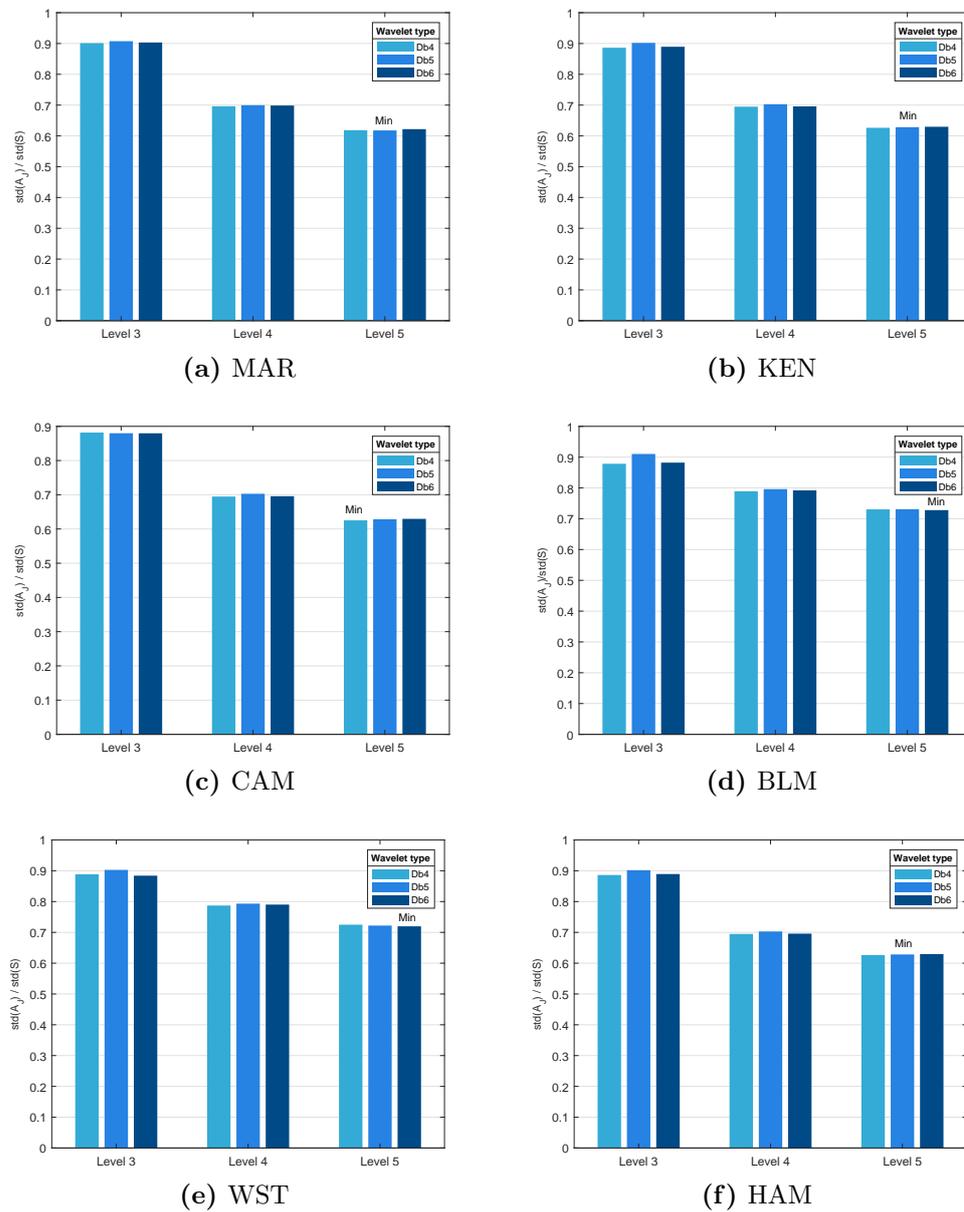
Fig. 5.8. Daubechies type and level optimisation results.

Figure 5.9 presents the results of the 5-level wavelet decomposition of the first 1000

samples of actual NO_2 data from MAR site, e.g. the upper curve, s(t), using MATLAB software. The said decomposition were obtained by applying Daubechies Db5 wavelets. A significant difference of variability can be observed after inspecting the fluctuations of D_1 through D_5 signals. That is, the higher is the wavelet level, the lower is the variation of its coefficients which translates to coefficients becoming easier to model.
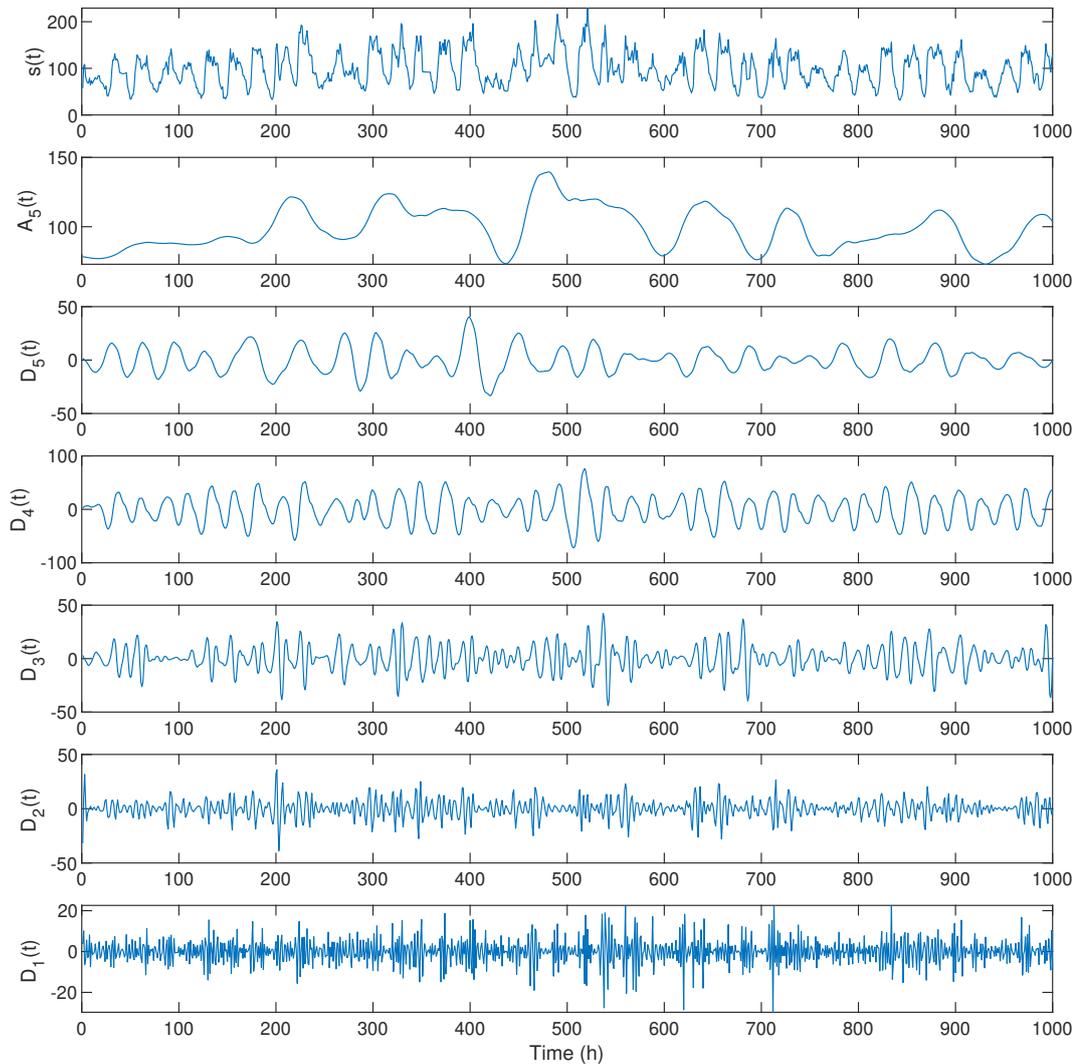
Fig. 5.9. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from MAR site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Note that their sum, together with A_5 signal, represents the original NO_2 time series. The said decomposition will be used in Experiment 2.1. The wavelet decomposition

of the actual NO_2 data measured from KEN, CAM, BLM, WST, and HAM sites are shown in Appendix J.

5.3.2 Model Validation Results of Experiment 2.1

As mentioned in Chapter 4 Section 4.4, Experiment 2 is split into two sub-experiments. The first one attempts to build a wavelet-based temporal forecasting model using the MLP model providing the best results in Experiment 1, e.g. the MLP_{SFE} model. Using the same data preprocessing schemes, predictors, hidden layer configurations and training parameters, a wavelet-based model is developed to assess the effectiveness of applying a level-5 Daubechies Db5 wavelet on the raw NO_2 data from MAR site.

The predictive and replicative validation results of models developed in Experiment 2.1 are presented in Table 5.6. Note that the computed RMSE (in $\mu g/m^3$), MAE (in $\mu g/m^3$) and IA values represent the mean values obtained in 10 repetitions of experiments at random initial training parameters.

Table 5.6. Predictive and replicative validation results of W-MLP and MLP models for the forecasting of hourly NO_2 levels.

Model	Horizon	Predictive			Replicative		
		RMSE	MAE	IA	RMSE	MAE	IA
W-MLP	+ 0 hr	11.701	9.851	0.980	9.504	7.313	0.982
	+ 1 hr	18.131	13.166	0.950	13.851	10.049	0.954
	+ 6 hr	27.418	21.295	0.827	22.817	17.739	0.829
	+ 12 hr	32.939	23.978	0.759	26.427	20.321	0.776
MLP	+ 0 hr	13.754	10.120	0.972	11.234	8.138	0.979
	+ 1 hr	19.228	13.792	0.942	16.394	12.054	0.953
	+ 6 hr	30.777	23.344	0.819	26.763	20.440	0.841
	+ 12 hr	34.498	26.557	0.748	30.288	23.503	0.771

Overall results show that the use of wavelet preprocessing translates to better forecasting results across various forecasting horizons. In detail, the W-MLP models are able to improve the accuracy of the MLP models results: decrease from 1.1 $\mu g/m^3$ to 3.36 $\mu g/m^3$ and from 0.27 $\mu g/m^3$ to 2.58 $\mu g/m^3$ in RMSE and MAE values, respectively, and increase from 0.007 to 0.011 in IA values. For the 1-h ahead forecasting, the W-MLP model reduced the RMSE and MAE values of the MLP model by up to

14%, and increased the IA values by 0.74%. A similar observation can be made when assessing the replicative validity of both proposed and benchmark models.

Similar to the results in Experiment 1, however, the forecasting accuracy of the proposed wavelet-based models in the current experiment suffer as the forecasting horizon increases. This time, forecasting horizons of 6-h and 12-h in advance were considered to test the ability of the models to provide accurate long-term forecasts. In detail, an increase from 9.29 μg/m^3 to 14.81 μg/m^3 and from 8.13 μg/m^3 to 10.81 μg/m^3 in RMSE and MAE values can be observed as the forecasting horizon move from 1-h to 6-h and 12-h, respectively. In particular, an increase by 60% and 69% in terms of RMSE and MAE values, respectively, and a 13% decrease in IA values were observed as the forecasting horizon moves from 1-h to 6-h. Nonetheless, the 6-h ahead forecasts of the W-MLP model are still fairly acceptable. Figure 5.10 illustrates the 1-h ahead forecasting results of both W-MLP and MLP models. The figure only corresponds to the first 500 samples from the testing data set for illustrative purposes.
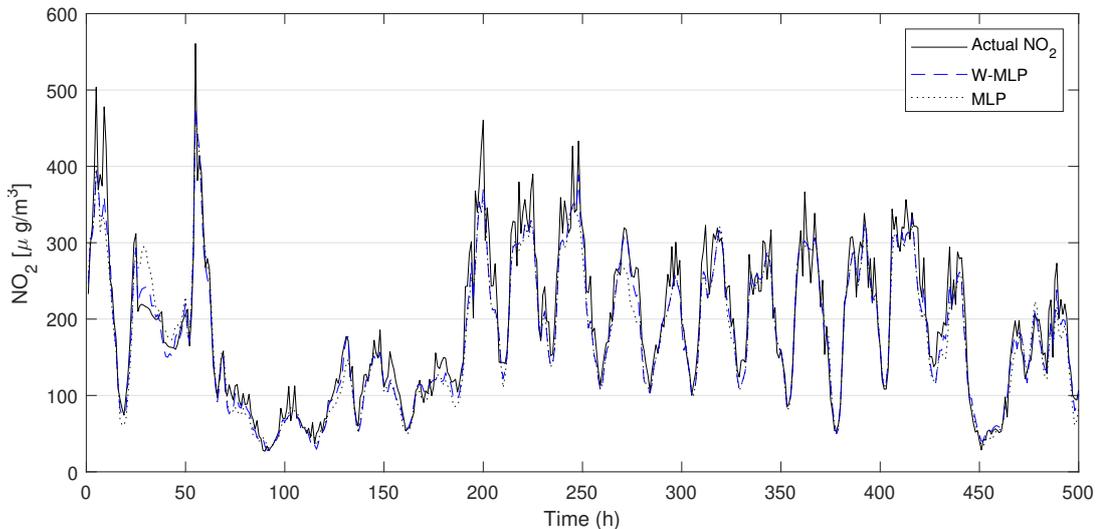
Fig. 5.10. Comparison of the 1h-ahead forecasting results of W-MLP and MLP models.

It can be seen that the 1-h forecasting results of the W-MLP model follow the hourly variations of the NO$_2$ data more accurately than those by the MLP model. Peak NO$_2$ levels are also better approximated by the W-MLP model in comparison with to plots

of the MLP model results. However, under-predictions by the W-MLP results can still be seen in a few hourly instances.

Figure 5.11 presents the distribution of 1-h ahead forecasting errors of both W-MLP and MLP models for the testing data. It can be observed that most of the errors of the W-MLP model are concentrated between the values of 0 μg/m^3 and 15 μg/m^3. Furthermore, the wavelet-based model has fewer number of significantly large errors (above 50 μg/m^3) than those from the plain MLP model. Such observations indicate that the W-MLP model performed better than the MLP model.
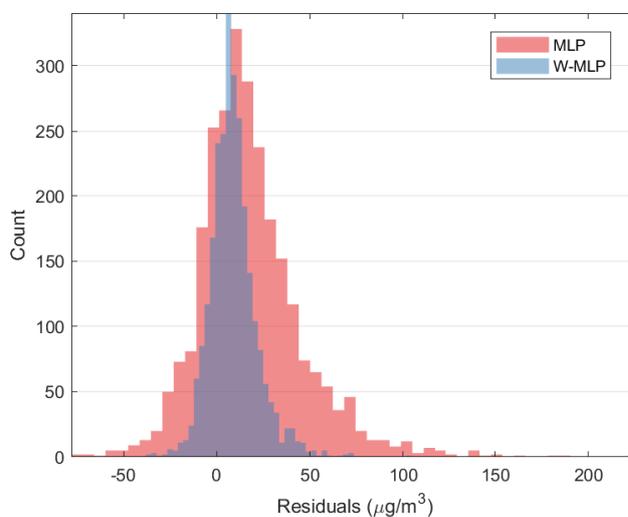
Fig. 5.11. Error histograms of the 1h-ahead forecasting results of W-MLP and MLP models.

The scatter plots of 1-h ahead forecasting results of both W-MLP and MLP models shown in Figure 5.12. It can be observed that the findings above are also in agreement with the scatter plots. Although both models suffer from both under- and over-prediction of NO$_2$ concentration values above 50 μg/m^3, the scatter plots of the W-MLP model are less dispersed around the ideal fit for value above 150 μg/m^3.

5.3.3   Hidden layer optimisation results for Experiment 2.2

The hidden layer configurations determined using the procedure described in Section 4.4 are summarised in Table 5.7. The optimal number of hidden nodes for the proposed

(a) W-MLP

(b) MLP
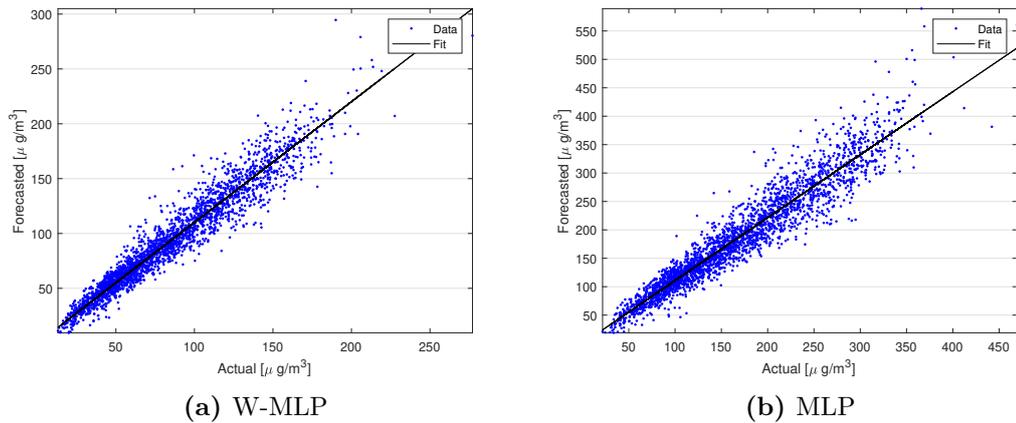
Fig. 5.12. Scatter plots of the 1-h ahead results of the W-MLP and MLP models.

LSTM models ranges from 100 to 150. More nodes tend to be needed to estimate the pollutant values at the kerbside target sites, e.g. MAR, CAM and HAM sites.

Table 5.7. Optimal hidden layer configurations of the both plain and wavelet-based MLP models.

Site /	Plain		W-MLP						W-LSTM					
Predictors	MLP	LSTM	D_1	D_2	D_3	D_4	D_5	A_5	D_1	D_2	D_3	D_4	D_5	A_5
MAR:														
All	35	150	20	30	20	25	20	30	125	115	115	110	115	110
CAM; HAM	30	125	20	25	25	20	25	25	115	120	115	115	105	105
KEN:														
All	35	130	20	20	40	20	20	40	120	115	120	110	105	100
BLM; WST	30	115	25	20	25	25	30	30	115	115	110	115	110	100
CAM:														
All	25	135	25	20	30	20	20	40	120	125	110	115	115	105
BLM; WST	25	120	20	25	25	20	25	30	115	120	115	110	110	105
BLM:														
All	25	130	20	35	20	45	20	50	125	120	120	115	110	110
KEN; WST	25	125	20	25	25	30	30	25	120	115	115	110	115	105
WST:														
All	30	130	25	20	25	20	25	40	125	125	115	120	115	115
KEN; BLM	25	125	30	20	20	25	20	25	125	120	120	110	110	100
HAM:														
All	25	145	25	25	20	35	25	40	130	125	125	120	125	115
BLM; WST	25	135	20	20	25	30	20	35	125	125	120	115	120	120

Furthermore, a lesser number of hidden nodes is needed as the decomposition level using Daubechies wavelets increases, indicating that the variability of a given time series directly influences the complexity of the model needed to estimate the time series. On the other hand, the computed number of hidden nodes for the benchmark MLP models varies from 20 to 35 for the standalone MLP models and 20 to 50 for the

wavelet-based MLP models. Similar to the results for the LSTM models, more hidden nodes are required to estimate the data collected from the kerbside urban sites. Lastly, the number of hidden nodes in most cases is less in models that use less number of predictors.

5.3.4   Correlation and Lag Analysis for Experiment 2.2

As shown in Figure 5.13, there is a high mutual correlation ($> 60\%$) between the NO_2 data collected from all sites but MAR site. The data from BLM and WST sites exhibit the highest mutual correlation index (85.2%). With the exception of MAR site, CAM and KEN sites obtained the lowest mutual correlation index (58.4%) despite their close proximity, e.g. approximately 2.55 km apart. It also appears that the data exhibiting the highest variability levels, e.g. data taken from MAR site, is least correlated with those data collected from the remaining sites. Conversely, the dataset with the least variability level, e.g. data taken from KEN site, is highly correlated with the rest of the collected data.

Ideally, the dataset that is least correlated with the rest of the datasets should be discarded. However, one of the objectives of the study is to test the spatial estimation ability of the proposed model at locations with a limited or poor quality of data. As such, variants of the wavelet-based and plain LSTM and MLP models are also built based on the predictors. Given the results provided in Figure 5.13, the top two neighbouring sites with data that are highly correlated to a given target site are chosen.

For the lag analysis of the model predictors, the number of lags for each predictor was determined using the autocorrelation function. The computed optimum number of lagged inputs varies from 1 to 2. That is, for a specific target site j, the predictors utilised are $x_i\left(t-h\right)$, $x_i\left(t-h-1\right)$, and $x_i\left(t-h-2\right)$, where $i \in [1, 6]$ and $i \neq j$.

5.3.5   Model Validation Results of Experiment 2.2

The predictive validation results of the models developed in the current experiment are presented in Table 5.8. Similar to Experiment 2.1, the values in the table represent the mean values of computed RMSE (in $\mu g/m^3$), MAE (in $\mu g/m^3$) and IA values from ten
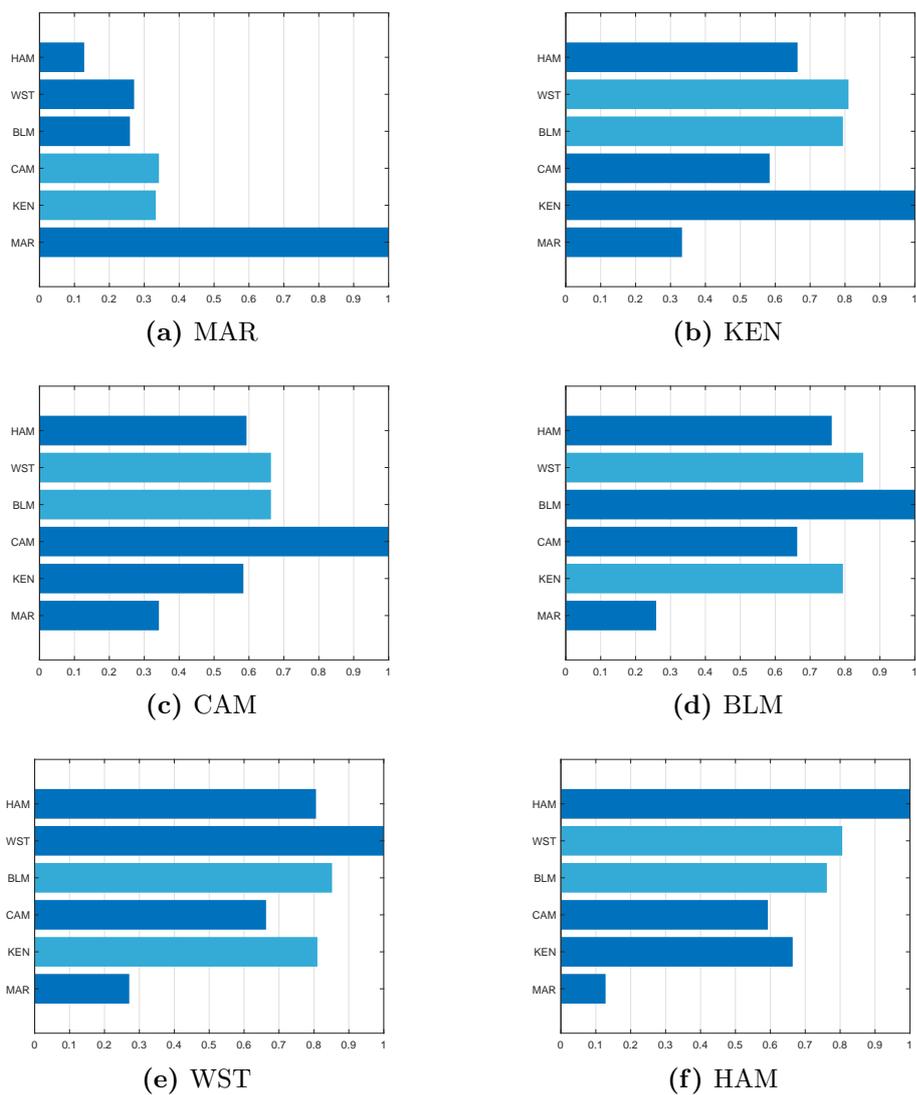
Fig. 5.13. The correlation scores between the collected NO_2 concentration data from between a target site and neighbouring sites.

repeated experiments to account for the random initial training parameters.

Table 5.8.   The predictive validation results of different models for the 1h-ahead forecasting of hourly NO_2 levels.

Model Code	Error Index	Target sites					
		MAR	KEN	CAM	BLM	WST	HAM
W-LSTM-CA	RMSE	34.791	10.204	23.012	9.110	11.311	22.052
	MAE	25.021	7.241	18.001	6.177	8.006	16.122
	IA	0.594	0.882	0.753	0.931	0.902	0.812
W-LSTM	RMSE	33.665	8.414	22.022	7.656	10.373	21.723
	MAE	24.500	5.808	16.785	4.892	7.086	15.224
	IA	0.610	0.946	0.809	0.953	0.930	0.847
LSTM-CA	RMSE	34.911	10.831	23.621	10.135	12.051	22.178
	MAE	26.331	8.333	18.731	6.819	9.563	17.019
	IA	0.592	0.883	0.754	0.922	0.887	0.799
LSTM	RMSE	34.224	9.610	22.708	9.579	10.854	22.730
	MAE	24.996	7.510	17.301	7.233	7.805	16.670
	IA	0.603	0.935	0.782	0.926	0.918	0.829
W-MLP-CA	RMSE	34.899	11.019	23.981	9.953	12.377	23.423
	MAE	26.872	8.627	18.763	7.391	9.599	18.053
	IA	0.589	0.869	0.740	0.910	0.873	0.793
W-MLP	RMSE	33.224	9.320	22.421	9.232	11.117	23.199
	MAE	24.796	7.324	17.116	7.115	8.371	18.031
	IA	0.594	0.931	0.794	0.937	0.905	0.826
MLP-CA	RMSE	35.665	11.065	24.213	10.244	12.689	23.823
	MAE	27.029	8.680	19.110	7.673	10.003	18.432
	IA	0.587	0.872	0.736	0.918	0.881	0.804
MLP	RMSE	36.654	9.664	22.924	10.186	11.201	23.695
	MAE	26.224	7.520	17.828	7.883	8.430	18.962
	IA	0.588	0.921	0.772	0.909	0.901	0.818

The overall results show that the integration of a wavelet decomposition technique and a deep learning model provides superior forecasting results. For instance, a significant increase in forecasting accuracy ranging from 0.007 to 0.063 (IA scores) is achieved by the W-LSTM models when compared to the plain LSTM models. The W-MLP models also exhibited improvement in results, e.g. 0.004 to 0.059 (IA scores), when compared to the plain MLP models. A reduction of RMSE and MAE values are also observed when the wavelet-based models are compared with the plain ones. For instance, a decrease by 1.633% and 1.984% of RMSE and MAE values, respectively, are achieved by the LSTM models for MAR site when DWT was implemented. In addition,

a decrease by 9.357% and 5.445% of RMSE and MAE values, are achieved by the MLP models for the same site when DWT was implemented.

The reduction of some utilised predictors through correlation analysis tends to degrade the overall performance of the models. For instance, the greatest decrease in accuracy, e.g. 0.064 decrease in IA, was observed by the W-LSTM model for KEN site when only two predictors are utilised. A similar observation can be made for most remaining models where decrease in IA scores ranges from 0.001 to 0.062. The results above suggest that the information derived from all selected reference sites is important in helping the ANN models approximate the NO_2 levels at a given target site. However, it is worth noting that the models with a reduced number of predictors for KEN, BLM and WST sites still provided satisfactory results. This indicates the applicability of the proposed spatiotemporal model in cases where the number of neighbouring sites is severely limited.

Consistent with the findings of several previous works (Alimissis et al., 2018; Tzanis et al., 2019), the performance of the models tend to be site-dependent. That is, the models for the background sites, e.g. KEN, BLM and WST sites, significantly outperform the models for the kerbside sites, e.g. MAR, CAM and HAM sites. It is also worth the emphasis that the performance of the plain LSTM and MLP models for KEN and BLM sites are almost similar. Factors such as the traffic characteristics, location, pollution sources, and geometry of the buildings around the target site tend to explain the results above. In summary, the ranking of the sites in terms of the model performance (in decreasing order) is as follows: BLM, KEN, WST, HAM, CAM and MAR (see Table 5.8).

The overall model results are also influenced by the level of variability of the data. That is, the models for the target sites with dataset exhibiting high variability perform poorly. For instance, the ranking of the sites in terms of standard deviation values (in increasing order) almost matches the ranking above: KEN, BLM, WST, HAM, CAM and MAR sites (see Table 4.8). Finally, the mutual relationship between data from the neighbouring and targets sites has a significant effect on model results. That is, the NO_2 level data from KEN, BLM and WST sites are highly correlated, e.g. correlation

score from 0.58 to 0.85, with the data of the remaining sites except MAR site.

Figure 5.14 presents the distribution of the 1-h ahead forecasting errors of both W-LSTM and LSTM models for all six monitoring sites.
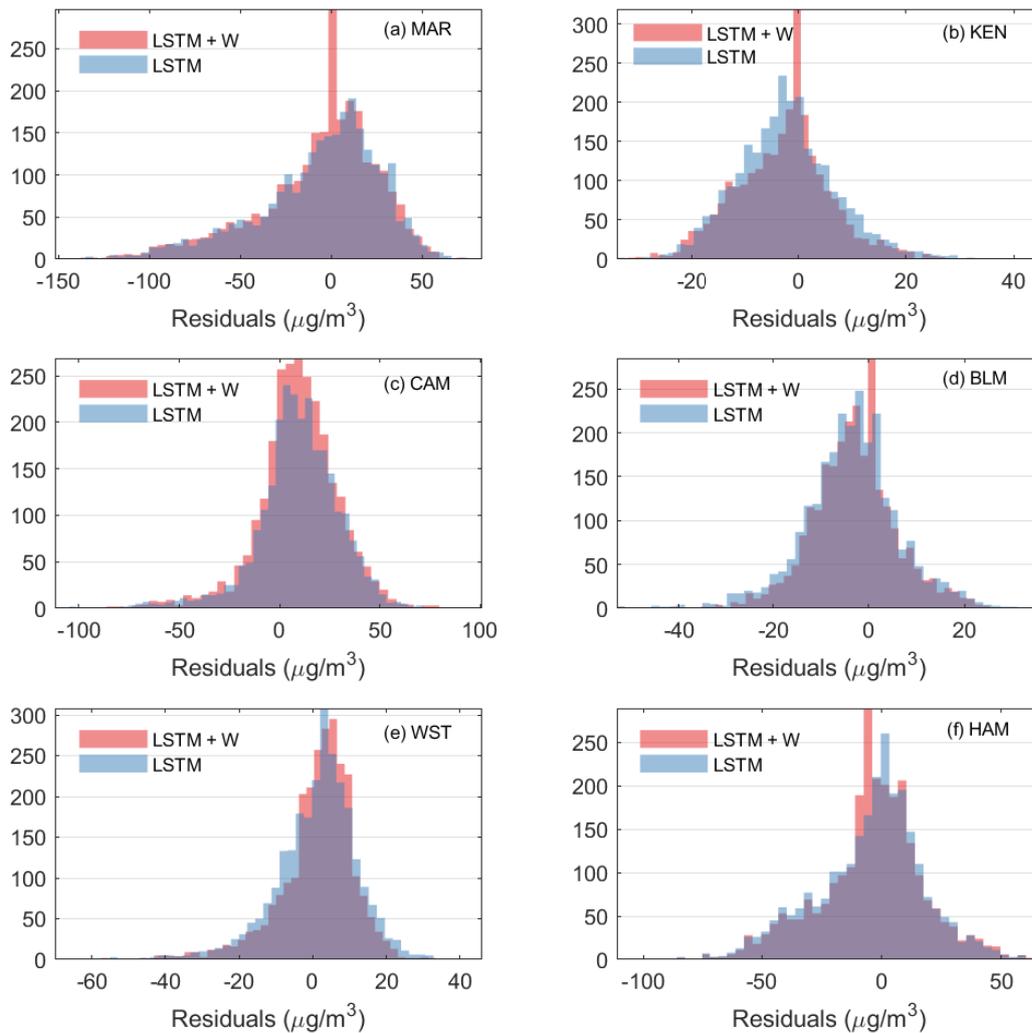
Fig. 5.14. Error histograms of the forecasting results of both proposed and benchmark LSTM models for (a) MAR, (b) KEN, (c) CAM, (d) BLM, (e) WST, and (f) HAM sites.

It can be observed that the errors of both models are concentrated around 0 $\mu g/m^3$. Consistent with the findings above, the improvement achieved by applying DWT on the base LSTM models is not apparent for those models employed for the kerbside sites. Furthermore, the site-dependency of the model results are clearly highlighted as the er-

ror values of best-performing models, e.g. models for KEN, BLM and WST sites, range from -40 $\mu g/m^3$ to 30 $\mu g/m^3$. In contrast, the error values of best-performing models, e.g. models for MAR, CAM and HAM sites, range from approximately -100 $\mu g/m^3$ to 50 $\mu g/m^3$.

Figure 5.15 presents the distribution of the 1-h ahead forecasting errors of both W-MLP and MLP models for all six monitoring sites.
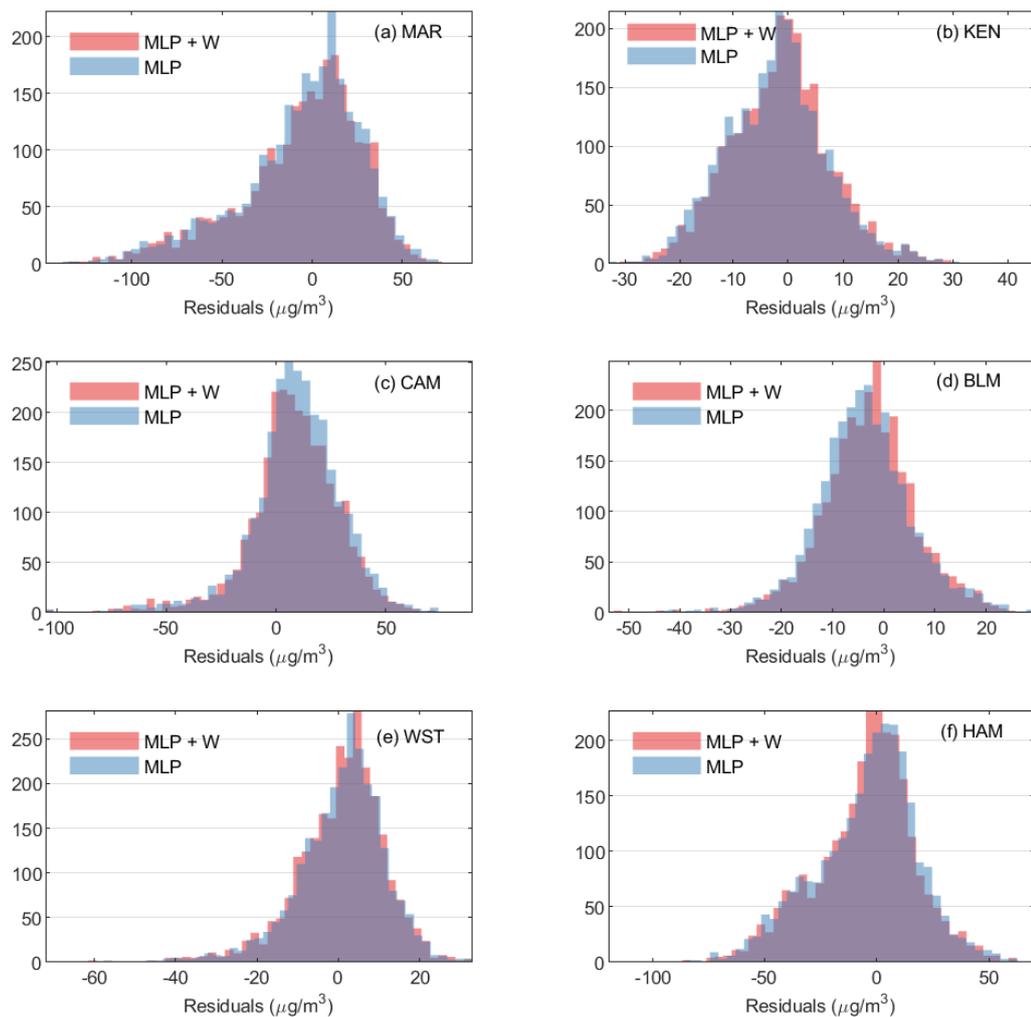
Fig. 5.15. Error histograms of the forecasting results of both proposed and benchmark MLP models for (a) MAR, (b) KEN, (c) CAM, (d) BLM, (e) WST, and (f) HAM sites.

Similar to the findings from the error distribution plots above, the error values

of the MLP model forecasts are centered at 0 μg/m^3. However, it is apparent that both W-MLP and MLP models have slightly wider range of residuals, e.g. -50 μg/m^3 to 30 μg/m^3, for the best-performing models. In contrast, the worst-performing ones have a slightly wider range of residual values, e.g. -105 μg/m^3 to 55 μg/m^3. As above, the site-dependency of the accuracy of the model results is quite apparent in the error distribution plots below.

The scatter plots of the results produced by the W-LSTM and W-MLP models are shown in Figures 5.17 and 5.16.
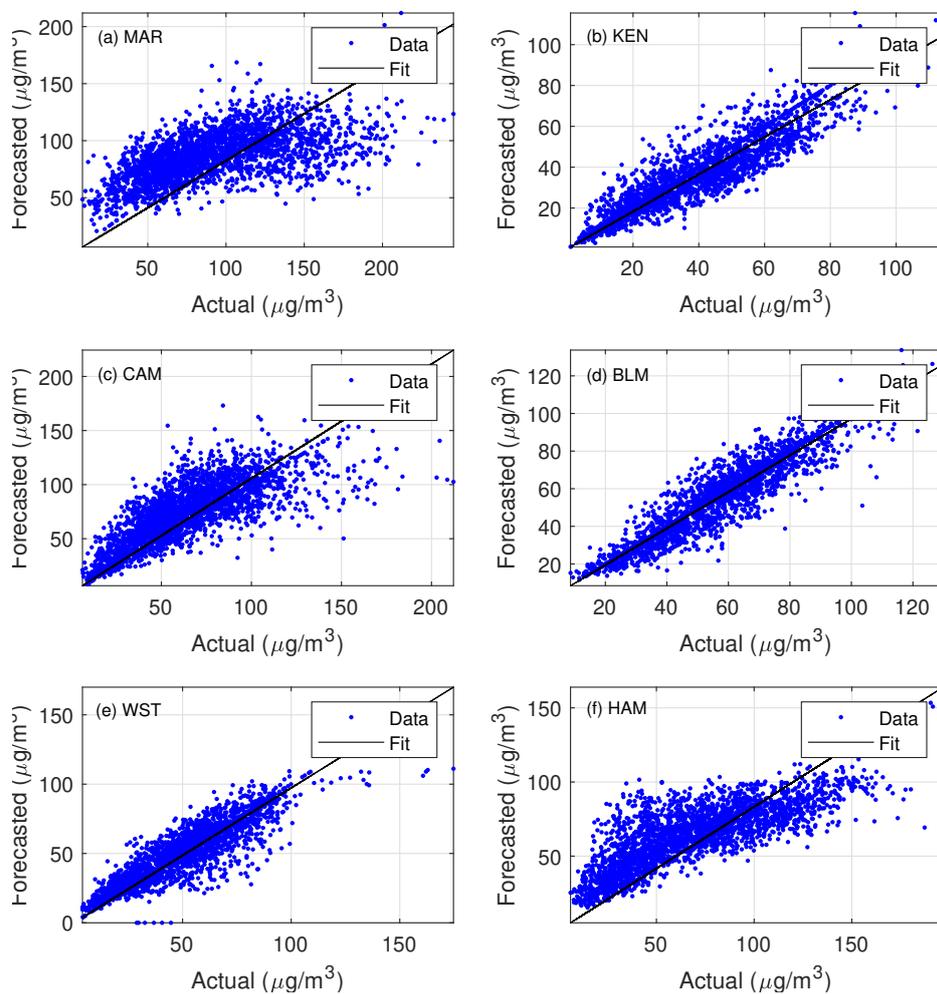
Fig. 5.16. Scatter plots of the actual NO$_2$ data and the forecasting results of the wavelet-based MLP models for (a) MAR, (b) KEN, (c) CAM, (d) BLM, (e) WST and (f) HAM sites.

Chapter 5. Results and Discussion

As shown in Figures 5.16b, 5.16d, and 5.16e, the actual and predicted NO_2 data points from the W-LSTM models for KEN, BLM and WST sites are concentrated near the ideal fit. In addition, the said models demonstrate their ability to capture the higher values of concentrations more accurately.
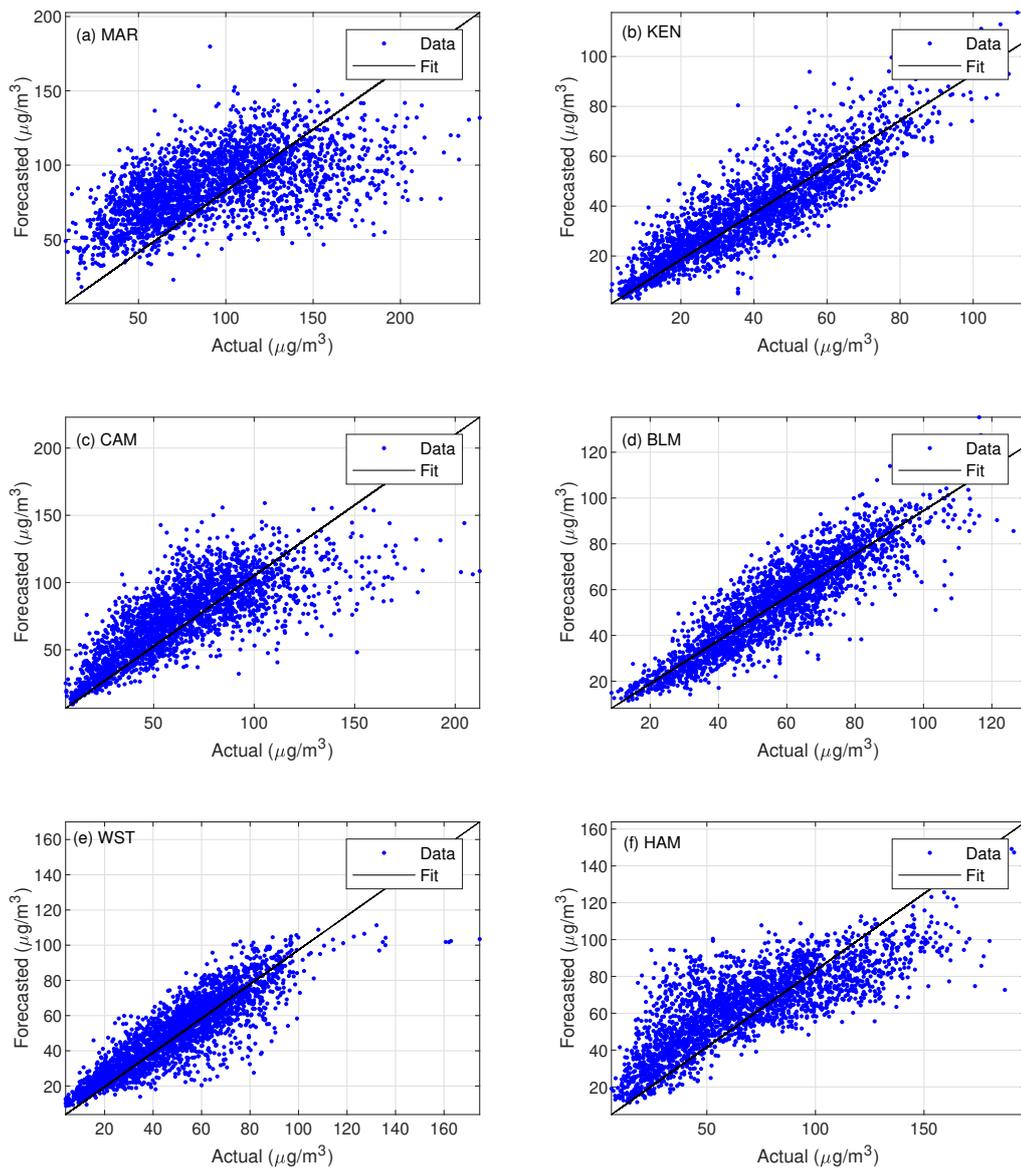
Fig. 5.17. Scatter plots of the actual NO_2 data and the forecasting results of the wavelet-based MLP models for (a) MAR, (b) KEN, (c) CAM, (d) BLM, (e) WST and (f) HAM sites.

A similar observation can be made for the results of the W-MLP models for KEN, BLM and WST sites (see Figures 5.17b, 5.17d, and 5.17e). In contrast, the poor performance of the models for MAR site is clearly depicted in Figures 5.17a and 5.16a. The scatter plot between the actual data and the results of the W-MLP model for MAR site also exhibits a very high tendency to over-and under-predict. This is also true for the W-MLP model for CAM site. In general, the plots demonstrate the suitability of the proposed wavelet-based approach for forecasting NO_2 levels at sites utilising only the data from their neighbouring sites.

5.3.6 Probability of detection and false alarm rate results of Experiment 2.2

The POD and FAR results results of selected developed models in Experiment 2.2 are summarised in Table 5.9. The W-LSTM and LSTM results for KEN, CAM and WST sites reveal the ability of the models to correctly forecast NO_2 episodes. In general, the wavelet-based LSTM models provide the best results although the wavelet-based MLP models also obtain satisfactory predictions. The application of DWT improves the ability of the plain LSTM and MLP models to correctly identify peak NO_2 pollution values. Both W-LSTM and W-MLP models for MAR site provide fairly reasonable results despite their generally poor results shown in the previous sections. The frequency of peak hourly NO_2 data from MAR site may have helped the models learn to approximate and reproduce peak values. However, the LSTM and W-LSTM models for both BLM and HAM sites correctly identified only 19% to 22% of the potential high levels of the actual NO_2 concentration data. On the other hand, all models that demonstrated good performance in correctly detecting potential episodes seem to display higher tendencies in issuing false alarms. The W-LSTM models for KEN, CAM and WST performed poorly, releasing false alerts 44% to 71% of the time. In contrast, the models for HAM and MAR sites exhibit the least tendencies in issuing false alarms, e.g. FAR of 0.03% and 19.10%, respectively.

Table 5.9. The results of different models for the 1h-ahead forecasting of hourly NO$_2$ levels.

Model Code	Alarm Results	Target sites					
		MAR	KEN	CAM	BLM	WST	HAM
W-LSTM	No. of predicted episodes	810	9	461	18	20	162
	No. of correctly predicted episodes	716	7	211	9	15	126
	POD	0.612	0.909	0.718	0.188	0.882	0.216
	No. of false alarms	147	4	233	6	10	4
	FAR	0.191	0.444	0.524	0.333	0.714	0.027
LSTM	No. of predicted episodes	740	11	463	15	20	135
	No. of correctly predicted episodes	603	4	198	9	13	82
	POD	0.532	0.801	0.623	0.196	0.565	0.148
	No. of false alarms	159	8	296	7	11	5
	FAR	0.197	0.464	0.614	0.383	0.015	
W-MLP	No. of predicted episodes	721	9	490	16	26	81
	No. of correctly predicted episodes	563	5	186	9	14	73
	POD	0.499	0.649	0.576	0.196	0.609	0.145
	No. of false alarms	168	4	294	7	7	2
	FAR	0.247	0.444	0.6204	0.438	0.350	0.030
MLP	No. of predicted episodes	680	14	471	15	20	67
	No. of correctly predicted episodes	512	5	177	7	13	63
	POD	0.477	0.625	0.543	0.152	0.565	0.129
	No. of false alarms	185	9	304	8	12	8
	FAR	0.257	0.6429	0.6242	0.533	0.462	0.100
	No. of observed episodes	1074	8	326	46	23	503

In general, the findings indicate that the wavelet-based deep learning approach improves the overall ability of the models for all sites to correctly identify actual episodes and avoid issuing false alerts. However, it should also be noted that the results in Table 5.9 are based on the predefined NO$_2$ level limit of 100 μ g/m^3. The said results vary depending on the specific limit levels set by various legislative bodies across the globe. Finally, it should be emphasised that the developed models in Experiment 2 still exhibit limitations in approximating peak NO$_2$ levels as shown in the preceding sections.

5.4 Experiment 3 Results

In the current experiment, wavelet-based spatiotemporal ANN models are developed using a two-approach involving the training of temporal estimators, e.g. LSTM models,

and spatial ensemble estimators, e.g. MLP models.

5.4.1   k-optimal neighbouring site selection

Before the proposed model was developed, the k-optimal neighbouring sites for each selected target site were determined. The step was carried out to determine the significant neighbouring sites for each target site. As shown in Figure 5.4.1, the correlation values between two monitoring sites varies from 0.11 to 0.87. The full numerical results of the correlation analysis is provided in Appendix K. Based on the findings, it follows that the number of neighbouring sites selected for each target site varies.
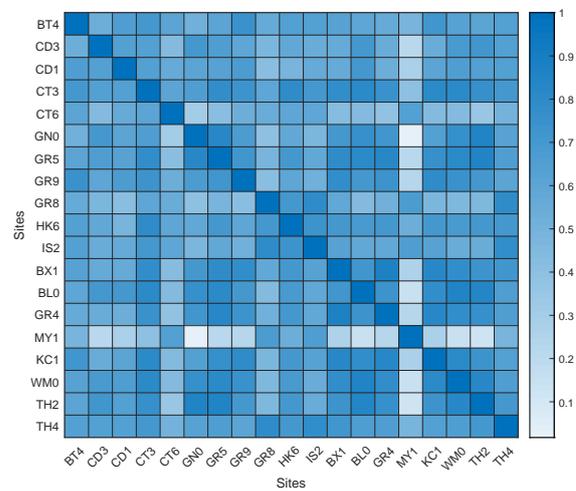
Fig. 5.18. A heat map of the Pearson correlation values between the NO_2 concentration data of the selected sites.

The results of the predictor selection approach based on correlation analysis are provided in Table 5.10. In most cases, the optimal correlation threshold value was selected to be low, e.g. equivalent to 0.65, as higher threshold values result to fewer selected neighbouring stations which might lead to poor network performance. This is particularly evident on the sites where the NO_2 data are weakly correlated to other sites.

Conversely, the use of a much lower correlation threshold value was also not as beneficial as the selection of many weakly correlated neighbouring sites leads to model

Table 5.10. Predictor selection results using the Pearson correlation values.

Site name	Correlation threshold	Selected neighbouring sites
BT4	0.65	CD1, CT3, GR9, KC1
CD1	0.65	BT4, GR9, BL0, WM0
CD3	0.65	GN0, BL0, WM0, TH2
CT6	0.6	BT4, CT3, MY1
GN0	0.65	CD3, CT3, GR5, GR9, BX1, BL0, GR4, WM0, TH2, TH4
GR5	0.75	CT3, GN0, BX1, BL0, GR4, TH2, TH4
GR8	0.65	HK6, IS2, MY1, TH4
GR9	0.65	BT4, CD1, CT3, GN0, GR5, BX1, BL0, GR4, KC1, WM0, TH4
HK6	0.65	CT3, GR5, GR8, IS2, BX1, BL0, GR4, KC1, WM0, TH2, TH4
IS2	0.7	CT3, GR8, HK6, TH4
MY1	0.6	CT6, GR8, IS2
TH2	0.65	CD3, CT3, GN0, GR5, GR9, HK6, BX1, BL0, GR4, KC1, WM0, TH4
TH4	0.65	CT3, GR5, GR9, HK6, IS2, BX1, BL0, GR4, WM0, TH2

overfitting (Hagan et al., 1995). It can also be observed that the measure of central tendency and dispersion of NO_2 data influence the correlation values between two monitoring sites. For instance, the sites exhibiting higher mean and standard deviation values, e.g. CT6 and MY1, are weakly correlated to the rest of the sites. Consequently, only three neighbouring sites were selected for each of the said sites. Conversely, the sites with the least mean and standard deviation of NO_2 concentrations, e.g. GN0, GR5, GR9, HK6, IS2, TH2 and TH4 sites, tend to be strongly correlated to many neighbouring sites. As such, approximately nine neighbouring sites were selected for the said target sites.

5.4.2   Model Validation Results for Experiment 3

Table 5.11 presents the predictive validation results of the proposed wavelet-based FC-LSTM model. Similar to the results of the previous experiments, the results presented in the table are the average RMSE (in $\mu g/m^3$), MAE (in $\mu g/m^3$) and IA values from 10 repeated experiments.

Table 5.11. The predictive validation results of the proposed wavelet-based FC-LSTM models for the forecasting of hourly NO_2 levels.

	Target Sites											
	BT4	CD1	CD3	CT6	GN0	GR5	GR8	GR9	HK6	IS2	MY1	TH2	TH4
RMSE	25.26	21.27	16.95	34.4	10.00	9.34	18.44	10.59	11.21	11.94	28.88	10.34	14.62
MAE	18.40	16.05	13.55	25.51	7.60	7.33	13.77	7.89	8.65	9.43	22.75	7.99	11.14
IA	0.78	0.89	0.87	0.81	0.96	0.94	0.92	0.94	0.93	0.90	0.86	0.96	0.90

The majority of the proposed models yielded significantly accurate forecasts, i.e. with IA scores of greater than 0.90. The best performance is achieved by the wavelet FC-LSTM models for GN0 and TH2 sites, both exhibiting the best agreement with the actual NO_2 concentrations at the respective locations. It is worth noting that GN0 and TH2 sites are among the few target sites with more than seven selected optimal neighbouring sites (see Table 5.10). This indicates the strong impact of highly correlated NO_2 data between target and neighbouring sites on model performance. For instance, the FC-LSTM model for BT4 site with NO_2 data which is weakly correlated with the NO_2 data of most neighbouring sites exhibited the worst performance. Modellers may opt to utilise the NO_2 data from all neighbouring sites but increasing the tendency of model overfitting due to the increased number of network parameters. In addition, the findings clearly indicate the site-dependency of the model results which is in agreement with previous results within the literature (Alimissis et al., 2018; Tzanis et al., 2019; Cabaneros et al., 2020).

Based on the findings above. several factors such as the variability of measured NO_2 data and its correlation to the data from neighbouring sites tend to have a massive influence on the accuracy of model results.

The comparison between the wavelet-based FC-LSTM model results and the actual NO_2 concentration values is presented in Figures 5.19 through 5.21. As shown in Figure 5.19e, 5.19f, 5.20a, 5.20b, 5.20d, 5.20f, and 5.21a, the estimated values of those models with an IA score of at least 0.90 are consistent with the actual NO_2 values.

(a) BT4

(b) CD1

(c) CD3

(d) CT6

(e) GN0

(f) GR5
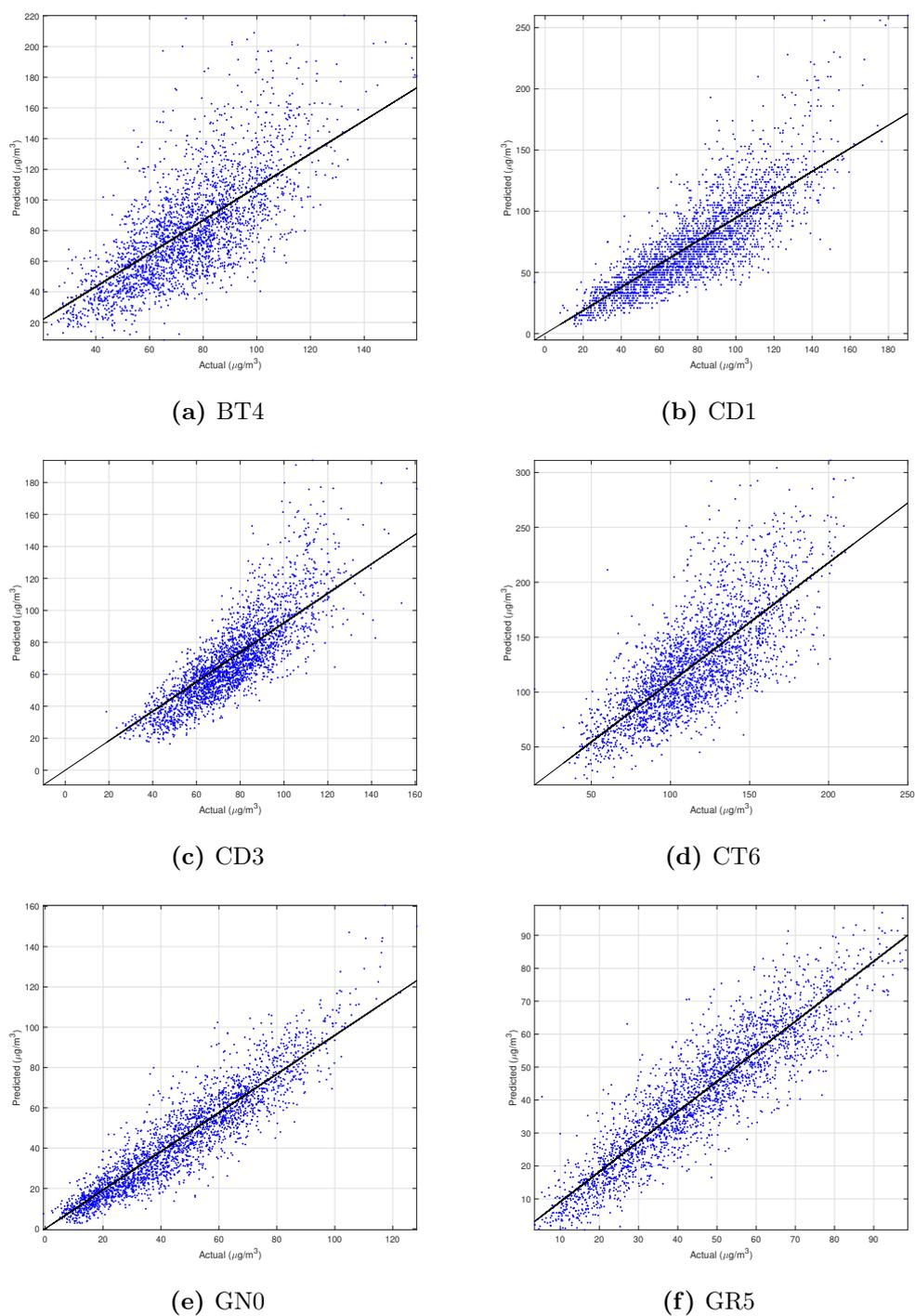
Fig. 5.19. Scatter plots of the 1-h ahead forecasting results of the Wavelet FC-LSTM models.
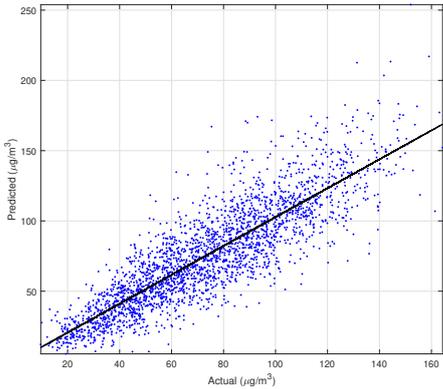
(a) GR8
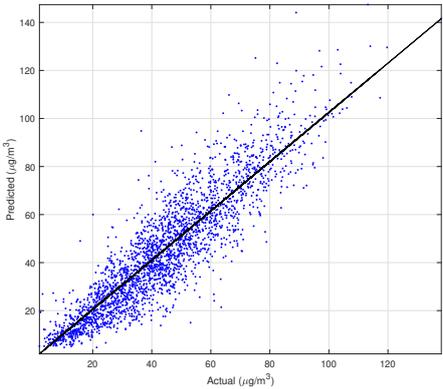
(b) GR9
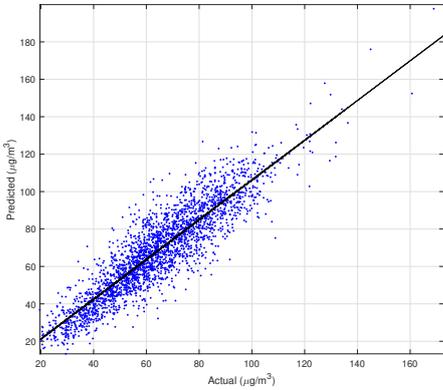
(c) HK6
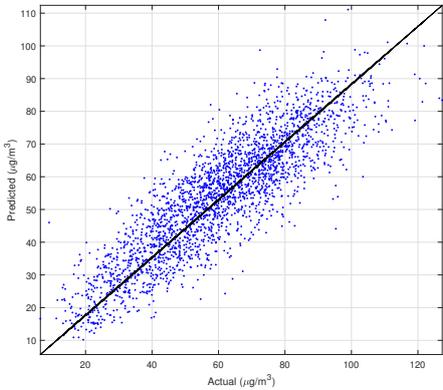
(d) IS2
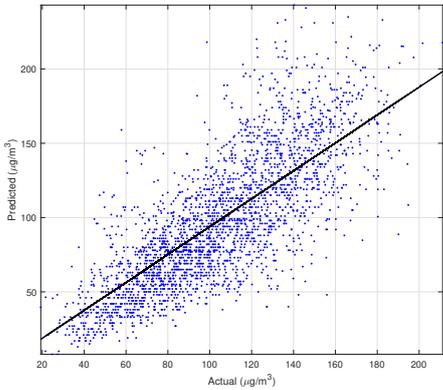
(e) MY1
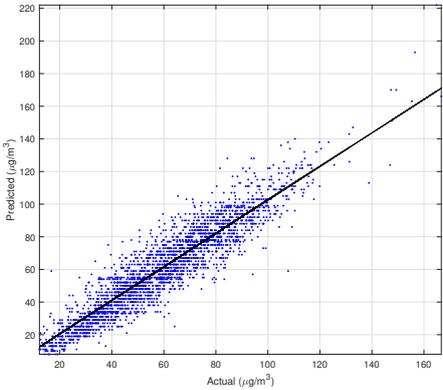
(f) TH2

Fig. 5.20. (continued) Scatter plots of the 1-h ahead forecasting results of the Wavelet FC-LSTM models.
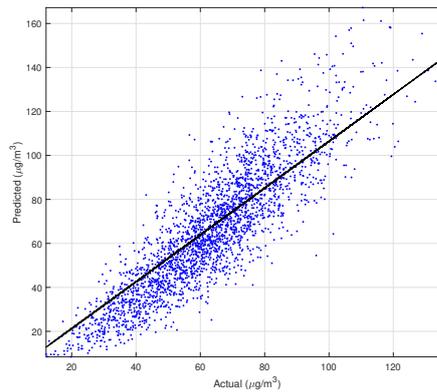
(a) TH4

Fig. 5.21. (continued) Scatter plots of the 1-h ahead forecasting results of the Wavelet FC-LSTM models.

The scatter plots clearly illustrate the ability of the said models to estimate the observed NO_2 data as the majority of the points are concentrated near the ideal fit. Conversely, the models with the worst performance exhibit strong tendencies to over- or under-estimate the actual NO_2 concentration data (see Figures 5.19a, 5.19d, and 5.20e). For instance, the poor performance of the proposed models at BT4 and MY1 sites is clearly depicted in the scatter plots, where a lot of points are sparsely located at both high and low concentration ends.

The comparison of the predictive validation results of all developed models is summarised in Figure 5.22. It can be seen that the Wavelet FC-LSTM models for all sites provided the least RMSE and MAE scores, and highest IA scores. The general results demonstrate the effectiveness of applying a preprocessing method based on DWT and a deep learning LSTM model combined with a feed-forward ANN model. Models for sites CT6, GR8 and MY1 sites seem to underperform. In contrast, all performance indices favour both the proposed and benchmarking models for GN0, GR5 and TH2 sites. The findings can be explained the variability of the NO_2 levels in those sites (Cabaneros et al., 2020).

(a) RMSE

(b) MAE

(c) IA
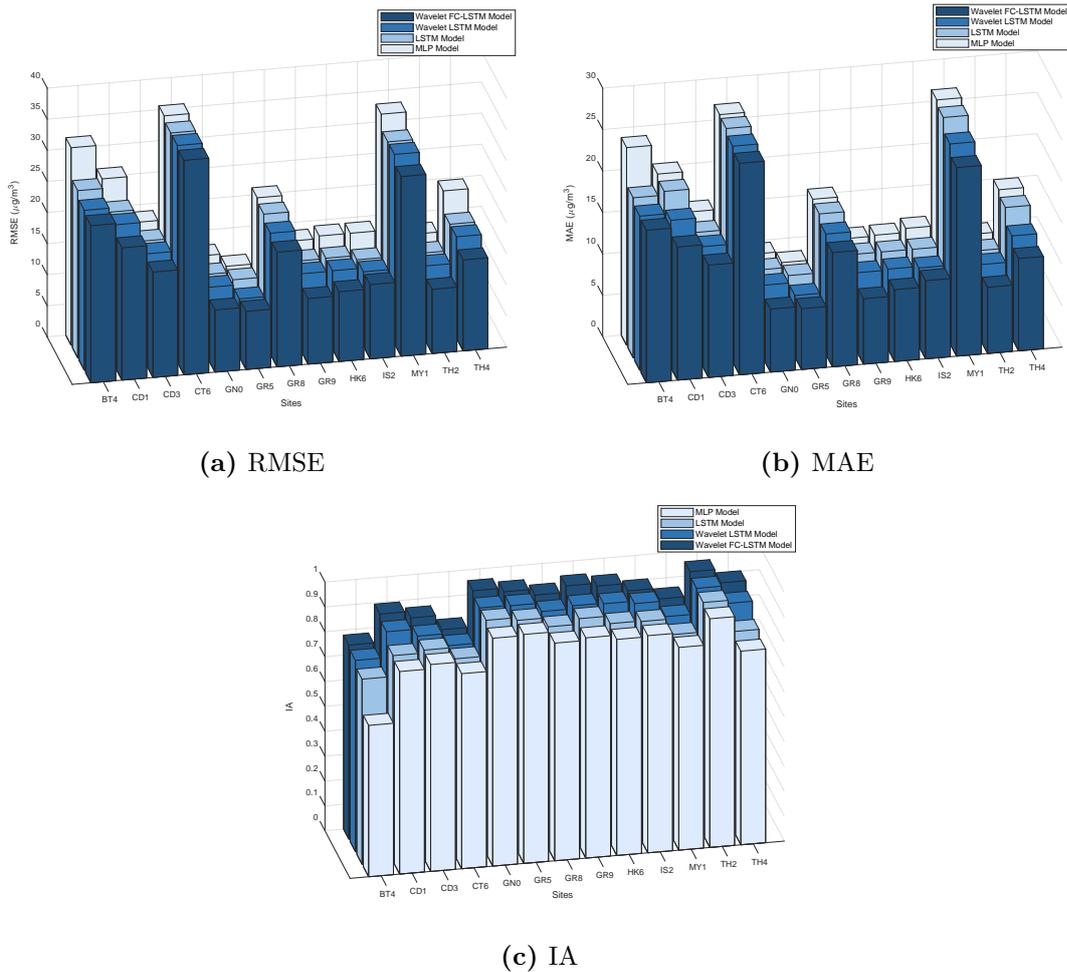
Fig. 5.22. The RMSE, MAE, and IA results of the proposed and benchmark models.

Table 5.12 provides the improvement of the performance accuracy (in %) between the wavelet FC-LSTM model and the benchmark models. Note that the said improvement of performance refers to the decrease of the RMSE and MAE values and increase of the IA values. In the comparison between the proposed and plain MLP models, the reductions of RMSE and MAE scores range from 3% to 26% and 5% to 26%, respectively, and the improvements of the IA scores range from 0.002 to 0.028. Such findings clearly indicate the improvement obtained by the benchmark models when DWT and an additional layer of MLP model serving as spatial estimator were implemented. On the other hand, the MLP model for BT4 site attained the highest improvement of accuracy. This finding is very noteworthy considering that such a site only considered a

126

very minimal number of neighbouring sites. The finding indicates that the proposed approach can help improve the performance of models with originally underwhelming forecasting results.

Table 5.12. The accuracy promotion of the benchmark models by the Wavelet FC-LSTM model.

	Proposed model vs. Plain MLP model												
	BT4	CD1	CD3	CT6	GN0	GR5	GR8	GR9	HK6	IS2	MY1	TH2	TH4
P_{RMSE} (%)	16.15	19.18	10.18	3.16	20.54	11.51	12.83	23.12	20.76	15.94	12.17	21.32	25.74
P_{IA} (%)	28.24	9.82	4.39	3.30	4.39	2.21	4.81	6.52	7.31	3.32	5.40	4.13	15.89
P_{MAE} (%)	18.89	21.02	12.18	5.02	21.48	12.37	12.19	25.57	19.19	15.83	13.78	18.63	25.29

	Proposed model vs. Plain LSTM model												
	BT4	CD1	CD3	CT6	GN0	GR5	GR8	GR9	HK6	IS2	MY1	TH2	TH4
P_{RMSE} (%)	6.00	17.07	5.77	3.16	15.71	10.41	10.17	18.45	17.30	7.21	11.07	15.67	21.67
P_{IA} (%)	4.84	8.01	3.43	2.07	1.96	1.52	3.03	3.38	5.70	2.56	5.24	2.46	11.76
P_{MAE} (%)	4.65	18.70	6.37	4.44	18.21	11.81	10.68	22.61	15.12	7.46	11.73	14.62	21.99

	Proposed model vs. Wavelet-based LSTM model												
	BT4	CD1	CD3	CT6	GN0	GR5	GR8	GR9	HK6	IS2	MY1	TH2	TH4
P_{RMSE} (%)	1.78	7.39	5.17	1.09	13.40	0.22	5.34	15.76	10.94	0.73	5.10	14.38	10.53
P_{IA} (%)	1.44	2.37	2.71	1.72	2.04	0.59	1.74	3.23	2.29	2.29	0.10	1.73	3.69
P_{MAE} (%)	1.05	9.65	5.01	2.15	15.29	1.27	5.05	17.16	10.20	0.19	5.56	3.39	9.84

In the comparison between the proposed and plain LSTM models, the reductions of RMSE and MAE scores range from approximately 3% to 22% and 4% to 23%, respectively, and the improvements of the IA scores range from 0.002 to 0.012. The said results also reveal the superior performance of deep learning LSTM models such as the LSTM model can still be improved when the original data is further analysed through a preprocessing technique such as DWT. The LSTM model for TH4 site exhibited the highest improvement when the proposed methodology was applied. The model for CT6 site observed the least improvement given how close its results are to those of the proposed model.

In the comparison between the proposed models and the wavelet-based LSTM models, the reductions of RMSE and MAE scores range from 0.22% to 16% and 0.20% to 17%, respectively, and the improvements of the IA scores range from 0.001 to 0.004.

Chapter 5. Results and Discussion

This finding reveals that the proposed addition feed-forward ANN model that acts as a spatial estimators can still improve the already outstanding results of the wavelet LSTM models. A significant improvement in performance accuracy is observed for GN0 and GR9 sites. However, the differences between the forecasts of the wavelet LSTM and proposed models are very small for GR5 and IS2 sites.

The plots of the 1-h ahead forecasting results of the proposed model with the best performance, e.g. the wavelet FC-LSTM model for GN0 and TH2 sites, and the corresponding benchmark models are presented in Figure 5.23.
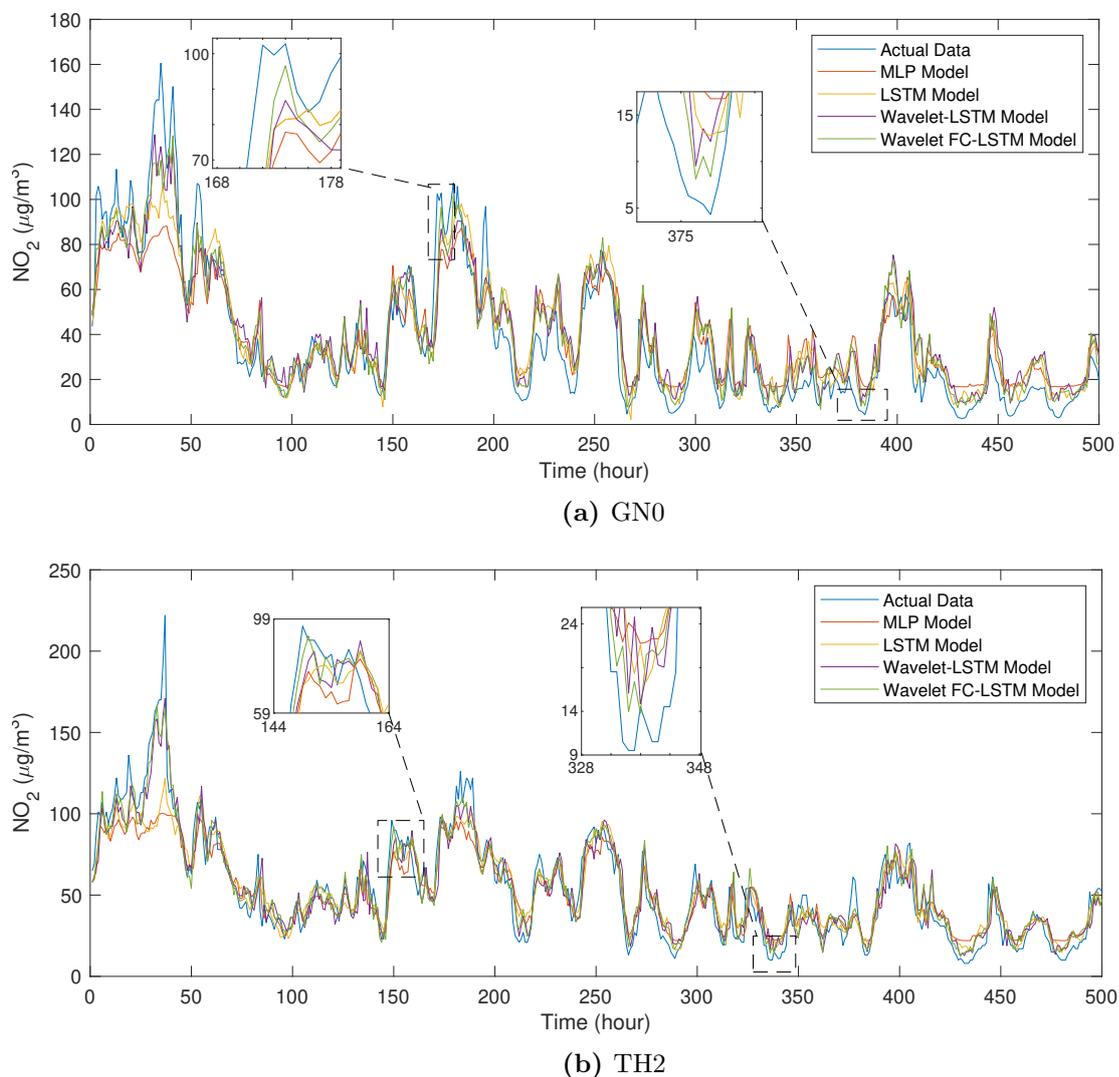
(a) GN0

(b) TH2

Fig. 5.23. Comparison of the model forecasting results at (a) GN0 and (b) TH2 sites

Note that for illustration purposes, only the last 19% of the test set, i.e. 500 hourly samples, is depicted in the figures. It can be observed that the 1-h ahead estimates of the proposed wavelet-based models coincide very well with the actual NO_2 data. This observation is in agreement with the calculated error indices presented in Table 5.12. The plots also highlight the improvement achieved by the proposed models when compared to the benchmark models. However, it is very apparent that all models encounter difficulty in approximating the peak actual NO_2 concentration measured at both sites. This observation is similar to those presented from the previous experiments. For instance, all models struggle to correctly estimate the extremely low and high actual NO_2 concentrations in almost all occasions. Nonetheless, it is very evident that the proposed models were able to provide estimates closer to those peaks than those from the benchmark models.

The plots of the forecasting results obtained for the remaining target sites are presented in Figures 5.24 and 5.25. Similar to the plots depicted in Figure 5.23, the plots of forecasting results of models for all target sites other than GN0 and TH sites reveal difficulty in estimating peak NO_2 levels. This observation is very apparent especially for models with CT6 and MY1 target sites, which suffer from high RMSE and MAE values (see Table 5.11). The superiority of the proposed Wavelet FC-LSTM models are also highlighted by the comparison plots below.
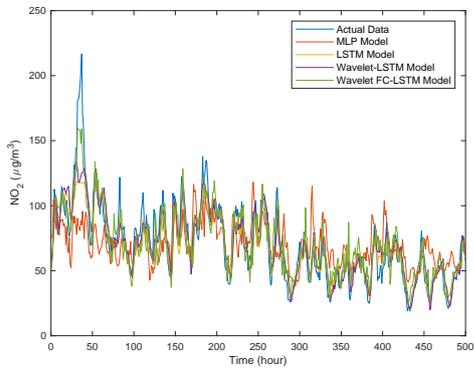
(a) BT4
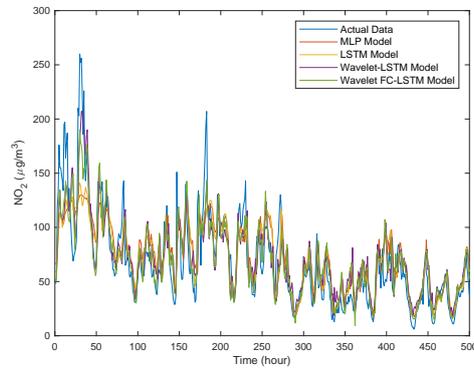
(b) CD1
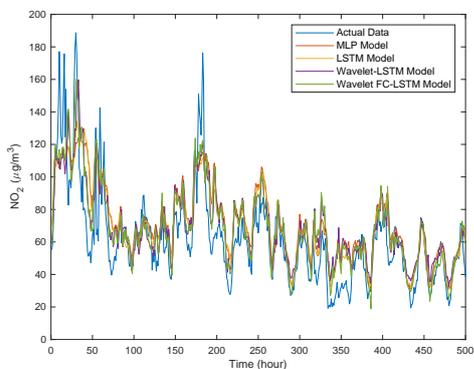
(c) CD3
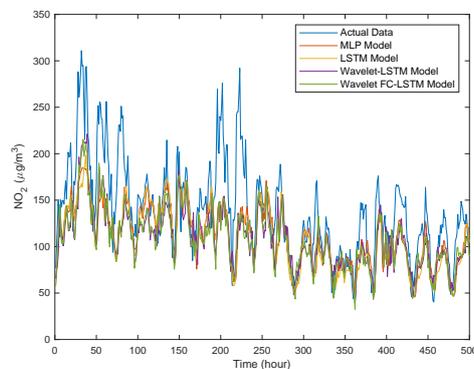
(d) CT6
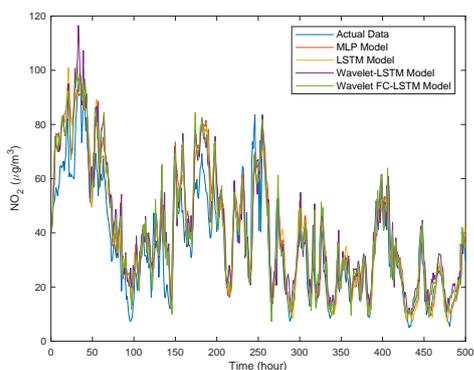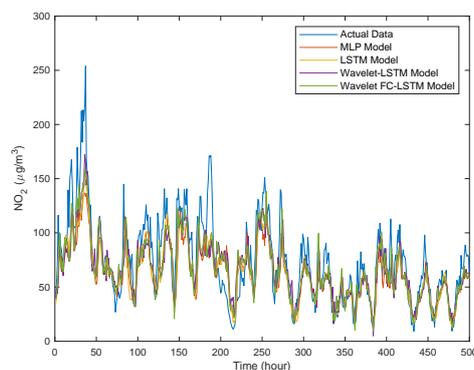
(e) GR5

(f) GR8

Fig. 5.24. Predicted and actual NO_2 concentration values at (a) BT4, (b) CD1, (c) CD3, (d) CT6, (e) GR5, and (f) GR8 sites.
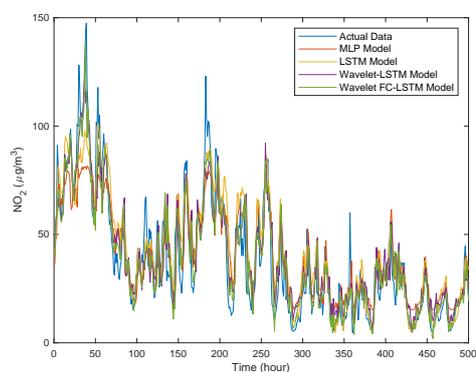
(a) GR9
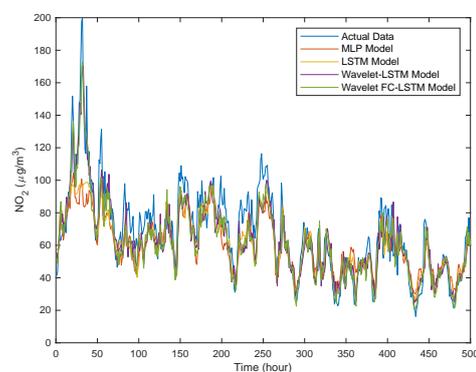
(b) HK6
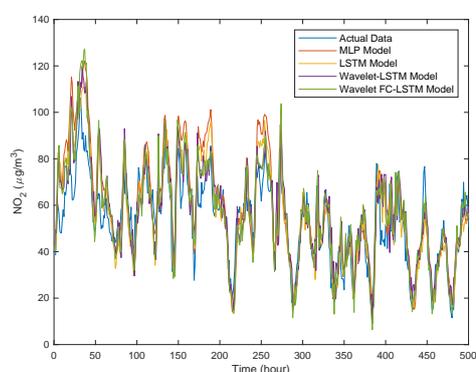
(c) IS2
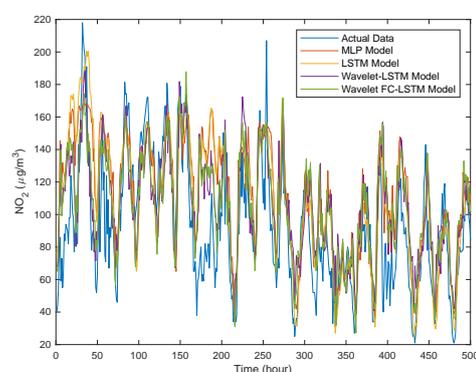
(d) MY1
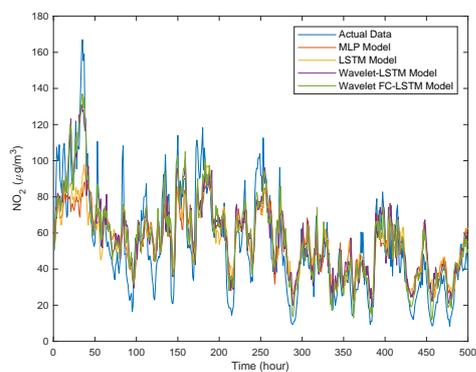
(e) TH4

Fig. 5.25. (continued). Predicted and actual NO_2 concentration values at (a) GR9, (b) HK6, (c) IS2, (d) MY1, and (e) TH4 sites.

The hourly surface distribution of the actual and estimated NO_2 concentrations across the Greater London area is shown in Figure 5.26. The spatial distribution of the NO_2 concentrations of the test dataset, e.g. 14/09/2012 to 31/12/2012, was obtained through inverse distance weighted interpolation. Details regarding the construction of the said spatial distribution can be found in Appendix L.
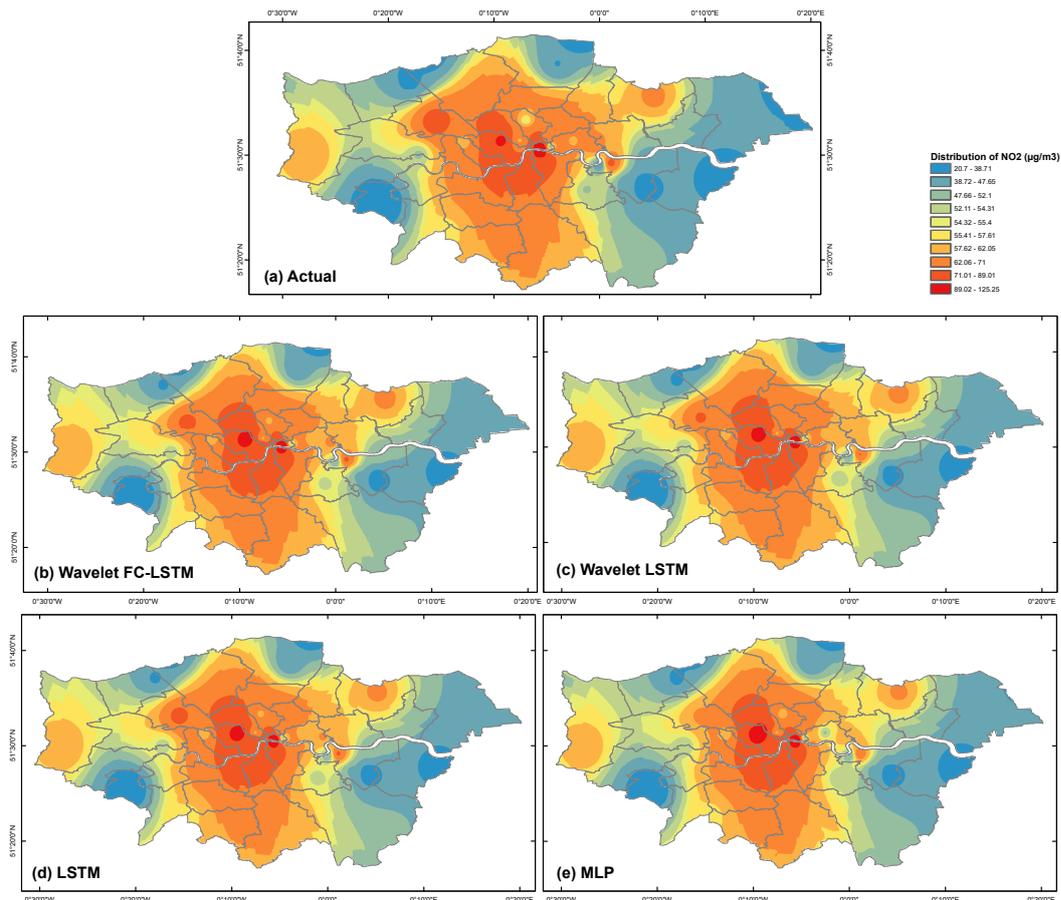
Fig. 5.26. Surface distribution of average NO_2 concentrations for (a) the actual data, (b) the proposed wavelet FC-LSTM, (c) wavelet LSTM, (d) LSTM, and (e) MLP models.

As shown in Figure 5.26a, the areas located in Central London, particularly at CT6, MY1 and GR8 sites (see Figure 4.8), suffer with mean NO_2 concentrations of more than 75 $\mu g/m^3$. Sites including the one located in London Marylebone Road, i.e. MY1 site, have been considered as one of most polluted areas in Central London often breach-

ing the NO_2 annual limits in recent years (The Guardian, 2017). Furthermore, it can be seen that the interpolation using the estimated values of the proposed model (see Figure 5.26b) is consistent with the interpolation using the actual NO_2 values in Figure 5.26a. The resulting interpolation result is significant considering only information from a selected number of optimal neighbouring sites were utilised to estimate the NO_2 level of a particular target site. Future studies may look into the effect of incorporating other predictors such as meteorological parameters on the performance of the proposed model. However, the proposed model, together with the benchmark models, tend to underestimate the surface values for locations with NO_2 concentrations between 62 μg/m^3 to 71 μg/m^3, highlighting the need to develop novel approaches to improve the ability of ANN models to estimate peak pollution levels in future works.

5.5 Summary

This Chapter has provided the results of applying temporal and spatiotemporal models based on several variants of plain and data-intensive hybrid ANN models in the short- and long-term forecasting of NO_2 levels in several outdoor locations in the Greater London area. Several criteria and visual inspection tools were applied in assessing the performance of the models. The proposed models were also validated using benchmark models based on their simpler or plain configurations.

Subsection 5.2 presented the results of employing various techniques in selecting the optimal predictors of an MLP model. Comparisons were made between the accuracy of forecasting results of models utilising various sets of predictors according to the results of three different predictor selection techniques. The developed MLP models were also compared with the MLR model across various forecasting horizons to assess the superiority of ANN models over linear statistical models.

Subsection 5.3 described the results of applying data-intensive hybrid MLP models based on DWT processing techniques. The developed models were employed in various applications, namely, the temporal and spatiotemporal modelling of NO_2 levels. Wavelet-based MLP models were validated against benchmark MLP models in tempo-

rally estimating NO$_2$ levels at a single site in Central London. In addition, wavelet-based variants of both MLP and LSTM models were validated against benchmark plain ANN models in spatially estimating NO$_2$ levels across six urban locations in Central London. The best-performing models were identified, while the influence of the target sites on the performance of each model were analysed.

Subsection 5.4 concluded the chapter with the results of applying a wavelet-based fully-connected LSTM models in estimating the spatiotemporal variations of hourly NO$_2$ levels across 19 monitoring sites in Greater London area. The proposed wavelet-based ensemble models were validated against several plain and hybrid models across multiple targets. Similar to the analyses from the preceding experiments, the influence of factors such as site locations and data variability on overall model performance were discussed and analysed.

Chapter 6

Conclusions and Future Work

6.1 Introduction

This Chapter provides the conclusions based on the numerical results presented in this research. Originally stated in Chapter 1, the primary aim of this research was to develop temporal and spatiotemporal models based on state-of-the-art ANNs in forecasting outdoor AP levels.

6.2 Research Conclusions

Each conclusion is numbered against the matching research objectives listed in Chapter 1. Each objective of this research is restated before each conclusion is provided.

1. **Conduct a critical review of existing works and state-of-the-art ANN models to identify gaps in the literature.**

   A critical review of existing literature was conducted for this research (see Chapter 2). The review covered a total of 139 peer-reviewed articles from 2000 to 2019 dealing with the use of ANN-based models to forecast urban AP concentration levels. The review highlighted the lack of systematic approaches implementing the protocol in building ANN models. The development of data-intensive hybrid and ensemble models were also emphasised given the prevailing issues in forecasting AP levels such as the peak levels and high variability of measured data. Preprocessing techniques

that reveal several inherent features from collected datasets were argued to help improve the performance of ANN models. Another identified gap that needed to be addressed is the development of parsimonious ANN models in the spatiotemporal estimation of AP levels. In more detail, models capable of estimating pollution levels in sites with limited or unreliable data using a few predictors from neighbouring sites need to be proposed and evaluated.

2. **Develop an AP forecasting model based on ANNs according to the identified model-building protocol from the literature review.**

Several models were developed according to the protocol identified from the critical review presented in Chapter 2. For instance, the use of predictor selection, various forms of data-splitting and more comprehensive model validation techniques were carried out to address the limitations of previous works identified by the said critical review. The integration of various predictor selection techniques with MLP models provided superior results to those from plain MLP models that utilised all collected predictors. In detail, the MLP models coupled with the SFS technique provided the best 0h- to 3h-ahead forecasts of hourly NO_2 levels in MAR site. For instance, the application of the SFS technique reduced the RMSE and MAE values of real-time predictions of the benchmark MLP model by 9.13% and 24.62%, respectively. Although most MLP models that combined with predictor selection techniques yielded relatively similar accurate forecasts, the general results favour the MLP_{SFS} models as the said models utilised almost half of the total number of predictors in training the MLP model and still managed to yield accurate forecasts. However, the performance of the developed models did suffer which is consistent to the results of previous works.

3. **Build and evaluate wavelet-based data-intensive hybrid ANN models for the temporal and spatiotemporal forecasting of outdoor AP levels across several monitoring sites in Central London.**

Several data-intensive hybrid ANN models were developed based on wavelet preprocessing. The proposed wavelet-based approach demonstrated an improvement on

the performance of plain MLP models in the temporal forecasting of NO_2 levels in MAR site. For instance, 39.15% and 28.58% reductions in RMSE and MAE indices, respectively, were calculated when wavelet preprocessing was applied to a plain 1-h ahead MLP forecasting model. In regards to the spatiotemporal estimation of NO_2 levels in several sites in Central London, the overall results favour the wavelet-based LSTM models over those calculated by the non-hybrid ones. For instance, reductions by 12.45% to 20.08% in terms of RMSE and MAE indices, respectively, were observed when DWT was coupled with the LSTM model for KEN site. The results above highlight the effectiveness of implementing preprocessing approaches based on wavelets in reducing the variability of the original collected NO_2 which eventually improves the model results. However, the findings of the developed wavelet-based models still exhibit difficulty in correctly estimating all peak AP values present at the selected monitoring sites. This finding is especially true for those models developed for the selected urban type sites.

4. **Build and evaluate data-intensive ensemble forms of ANN models for the spatiotemporal forecasting of outdoor AP levels across the Greater London area.**

The application of ensemble modelling approaches based on DWT and LSTM models was extended by employing them to more urban monitoring sites across the Greater London area. The proposed deep spatiotemporal forecasting models were found to provide significantly accurate estimates of NO_2 levels at several target sites utilising only the information from neighbouring sites. For instance, one of the developed models provided predictive results with an agreement score of up to 96%. Furthermore, the integration of a wavelet-preprocessing technique and deep spatiotemporal estimators represented by MLP and LSTM models has been found to improve the performance of the benchmark MLP and LSTM models by up to 28% in terms of IA scores.

Based on the conclusions set out above, the primary aim of this research has been achieved.

6.3 Contribution to knowledge

In addition to the research described in this thesis and the conclusions outlined in Subsection 6.2, the following contributions to knowledge are claimed:

1. An effective framework for forecasting outdoor AP levels using state-of-the art ANN models has been presented. Although the models in this thesis were trained using measurements from specific locations in the Greater London area, the protocol employed in performing the experiments can easily be adapted to build ANN models using datasets from other urban locations.

2. The ability of hybrid and ensemble ANN models which were based only on the NO_2 pollution data from neighbouring sites to produce accurate NO_2 forecasts has been demonstrated. The parsimonious approach offers a high theoretical significance as the influence of other explanatory variables such as meteorology, emissions, and traffic characteristics is not present. Hence, the effectiveness of the developed models relied mainly on the proposed techniques, e.g. wavelet preprocessing and ensemble modelling framework.

3. Data-intensive hybrid and ensemble temporal and spatiotemporal ANN models have been developed which demonstrated their ability to correctly reproduce NO_2 concentration values and superiority over existing hybrid ANN models.

6.4 Future Work

The following areas were identified as worthy for further investigation to extend the results that are presented within this thesis.

1. There is still a need to focus on more sophisticated hybrid modelling techniques while carefully accounting for the trade-off between model complexity and performance. In cases where reliable data and computational resources are both limited, the development of an effective yet parsimonious model should be one of the primary aims of future studies.

2. Although the proposed data-intensive hybrid and ensemble models have improved the performance of the benchmark ones, they are still limited in estimating most peak AP concentration levels from the utilised datasets. The integration of data re-sampling approaches with hybrid ANN models could be adapted and examined to address the issue of imbalanced data problem.

3. There is still a need to develop a general framework for determining the optimal amount of data to be used in calibrating any ANN model configuration. Although several studies have highlighted the use of a large number of historical data for training ANN models, there are only a small number that attempt to provide a systematic framework that can be applied to any case study.

4. The optimisation of several hidden layer configurations can be a tedious process and a clean-cut approach is still not available. One step in addressing this shortcoming would be to develop evolutionary model-based techniques that determine the optimal number of nodes while taking into account model parsimony and computational complexity.

5. The hybrid of physics-based and ANN-based models has been rarely examined within the literature. Using the same principles into building ensemble-type ANN models, the coupling of physics-based and data-driven approaches could potentially take advantage of the strengths and alleviate the weaknesses of the said involved techniques.

Appendices

Appendix A

Peer-reviewed Articles Selected for Comprehensive Review

Tables A.1 through A.4 present the full list of the selected 139 peer-reviewed articles, including details such as name of authors, year of publication, case study location, and examined air pollutant(s). Full details of the comprehensive review can be found in Cabaneros et al. (2019).

Table A.1. Details of papers reviewed

Authors (year)	Location(s)	Air pollutants examined
Kolehmainen et al. (2001)	Stockholm, Sweden	NO_2
Perez and Trier (2001)	Santiago, Chile	NO; NO_2
Chelani et al. (2002)	Delhi, India	SO_2
Abdul-Wahab and Al-Alawi (2002)	Khaldiya, Kuwait	O_3
Kukkonen et al. (2003)	Helsinki, Finland	NO_2
Lu et al. (2003)	Hong Kong	RSP
Wang et al. (2003)	Mong Kok, Hong Kong	RSP
Hasham et al. (2004)	Edmonton, Canada	NO_x
Heo and Kim (2004)	Seoul, Korea	O_3
Jiang et al. (2004)	Shanghai, China	TSP; SO_2; NO_x
Niska et al. (2004)	Helsinki, Finland	NO_2
Nunnari (2004)	Syracuse, Italy	SO_2
Olcese and Toselli (2004)	Cordoba, Argentina	?
Chelani et al. (2005)	Kolkata, India	NO_2
Hooyberghs et al. (2005)	Belgium	PM_{10}
Niska et al. (2005)a	Helsinki, Finland	NO_2; $PM_{2.5}$
Ordieres et al. (2005)	Ciudad Juarez El Paso, Mexico	$PM_{2.5}$

Appendix A. Peer-reviewed Articles Selected for Comprehensive Review

Table A.2. (continued) Details of papers reviewed

Authors (year)	Location(s)	Air pollutants examined
Agirre-Basurko et al. (2006)	Bilbao, Spain	O_3; NO_2
Grivas and Chaloulakou (2006)	Athens, Greece	PM_{10}
Nagendra and Khare (2006)	New Delhi, India	NO_2
Schlink et al. (2006)	Several EU countries	O_3
Slini et al. (2006)	Thessaloniki, Greece	PM_{10}
Brunelli et al. (2007)	Palermo, Italy	SO_2; O_3; PM_{10}; NO_2; CO
Dutot et al. (2007)	Orleans, France	O_3
Osowski and Garanty (2007)	Warsaw, Poland	CO; NO_2; SO_2; dust
Sousa et al. (2007)	Porto, Portugal	O_3
Al-Alawi et al. (2008)	Kuwait	O_3
Coman et al. (2008)	Paris, France	O_3
Díaz-Robles et al. (2008)	Temuco, Chile	PM_{10}
Ibarra-Berastegi et al. (2008)	Bilbao, Spain	SO_2; CO; NO_2; NO; O_3
Martín et al. (2008)	Algeciras, Spain	CO
Perez and Salini (2008)	Santiago, Chile	$PM_{2.5}$
Solaiman et al. (2008)	Ontario, Canada	O_3
Zito et al. (2008)	Leicestershire, UK	CO; NO_2
Ettouney et al. (2009)	Jahra, Kuwait	O_3
Galatioto and Zito (2009)	Palermo, Italy	CO; C_6H_6
Hrust et al. (2009)	Zagreb, Croatia	NO_2; O_3; CO; PM_{10}
Juhos et al. (2009)	Szeged, Hungary	NO; NO_2
Pisoni et al. (2009)	Milan, Italy	O_3
Tsai et al. (2009)	Taiwan	O_3
Demir et al. (2010)	Istanbul, Turkey	PM_{10}
Inal (2010)	Istanbul, Turkey	O_3
Jain and Khare (2010)	Delhi City, India	CO
Kurt and Oktay (2010)	Istanbul, Turkey	SO_2; CO; PM_{10}
Mahapatra (2010)	New Delhi, India	O_3
Moustris et al. (2010)	Athens, Greece	ERPI (NO_2; CO; SO_2; O_3)
Pires et al. (2010)	Oporto, Portugal	O_3
Feng et al. (2011)	Beijing, China	O_3
Paschalidou et al. (2011)	4 cities in Cyprus	PM_{10}
Prakash et al. (2011)	New Delhi, India	CO; NO_2; NO; O_3; SO_2; $PM_{2.5}$
Vlachogianni et al. (2011)	Thessaloniki, Greece Helsinki, Finland	PM_{10}; NO_x
Voukantsis et al. (2011)	Thessaloniki, Greece Helsinki, Finland	PM_{10}; $PM_{2.5}$
Barrón-adame et al. (2012)	Salamanca, Mexico	SO_2
Chattopadhyay and Chattopadhyay (2012)	Kolkata, India	O_3
Fernando et al. (2012)	Phoenix, Arizona	PM_{10}
Perez (2012)	Santiago, Chile	PM_{10}

Table A.3. (continued) Details of papers reviewed

Authors (year)	Location(s)	Air pollutants examined
Singh et al. (2012)	Lucknow, India	RSPM; SO_2; NO_2
Siwek and Osowski (2012)	Warsaw, Poland	PM_{10}
Antanasijević et al. (2013)	26 EU countries	PM_{10}
Arhami et al. (2013)	Tehran, Iran	CO; NO_x; NO; NO_2; O_3
Gennaro et al. (2013)	Northeast Spain	PM_{10}
Moustris et al. (2013)	Greater Athens Area, Greece	PM_{10}
Papaleonidas and Iliadis (2013)	Athens, Greece	O_3
Russo et al. (2013)	Lisbon, Portugal	NO_2
Ul-Saufie et al. (2013)	Negeri Sembilan, Malaysia	PM_{10}
Yan and Jian (2013)	Hangzhou, China	PM_{10}; $PM_{2.5}$
Zhang et al. (2013)	Taiyuan, China	PM_{10}
Azid et al. (2014)	Malaysia	API
Elangasinghe et al. (2014)	Auckland, New Zealand	NO_2
He et al. (2014)	Mong Kok, Hong Kong	PM_{10}; PM_1
Luna (2014)	Rio de Janeiro, Brazil	O_3
Özdemir and Taner (2014)	Kocaeli, Turkey	PM_{10}
Russo and Soares (2014)	Lisbon, Portugal	PM_{10}
Zhang et al. (2015)	Xi'an Province, China	$PM_{2.5}$
Alam and McNabola (2015)	Vienna, Austria	PM_{10}
Biancofiore et al. (2015)	Pescara, Italy	O_3
Cortina-Januchs et al. (2015)	Salamanca, Mexico	PM_{10}
Dunea et al. (2015)	Oltenia, Romania	O_3; PM_{10}; $PM_{2.5}$
Dursun and Taylan (2015)	Konya City, Turkey	SO_2
Feng et al. (2015)	Jing-Jin-Ji area, China	$PM_{2.5}$
Mishra et al. (2015)	Agra, India	NO_2
Russo et al. (2015)	Lisbon, Portugal	PM_{10}
Santos and Fernández-olmo (2015)	Cantabria Region, Spain	As; Cd; Ni; Pb
Zhu et al. (2015)	Chongqing, China	NO_x
Zou et al. (2015)	Texas, USA	$PM_{2.5}$
Abderrahim et al. (2016)	Algiers, Algeria	PM_{10}
Bai et al. (2016)	Chongqing, China	PM_{10}; SO_2; NO_2
Catalano et al. (2016)	London, United Kingdom	NO_2
Chellali et al. (2016)	Algiers, Algeria	PM_{10}
Ding et al. (2016)	Hong Kong	NO_2; NO_x; O_3; SO_2; $PM_{2.5}$
Durao et al. (2016)	Sines, Portugal	O_3
He et al. (2016)	Lanzhou, China	SO_2; NO_2; PM_{10}
Hoshyaripour et al. (2016)	Sao Paulo, Brazil	O_3
Li et al. (2016)	Beijing, China	$PM_{2.5}$
Li et al. (2017)	China	$PM_{2.5}$
Lightstone et al. (2017)	United States of America	$PM_{2.5}$
Mao et al. (2017)	Eastern China	$PM_{2.5}$
Peng et al. (2017)	Canada	O_3; PM_{10}; NO_2
Rahimi (2017)	Tabriz, Iran	NO_x; NO_2

Appendix A. Peer-reviewed Articles Selected for Comprehensive Review

Table A.4. (continued) Details of papers reviewed

Authors (year)	Location(s)	Air pollutants examined
Stamenković et al. (2017)	17 EU countries, USA, China, Japan, Russia and India	NO_x
Taylan (2017)	Jeddah, Saudi Arabia	O_3
Yeganeh et al. (2017)	Queensland, Australia	$PM_{2.5}$
Zhang and Ding (2017)	Hong Kong	NO_2; NO_x; O_3; $PM_{2.5}$; SO_2
Alimissis et al. (2018)	Athens, Greece	NO_2; NO; O_3; CO; SO_2
Antanasijević et al. (2018)	26 EU countries	SOx; NO_x; NH_3; $NMVOC$; PM_{10}
Dotse et al. (2018)	Brunei Darussalam	PM_{10}
Franceschi et al. (2018)	Bogota, Colombia	$PM_{2.5}$; PM_{10}
Freeman et al. (2018)	Kuwait	O_3
Gao et al. (2018)	Jinan, China	O_3
Huang and Kuo (2018)	Beijing, China	$PM_{2.5}$
Jiang et al. (2018)	Beijing, China	$PM_{2.5}$; SO_2; NO_2; CO; O_3
Li and Zhu (2018)	China	$PM_{2.5}$; PM_{10}; CO
Nidzgorska-Lencewicz (2018)	Tricity Agglomeration, Poland	PM_{10}
Pak et al. (2018)	Beijing, China	O_3
Radojević et al. (2018)	Belgrade, Serbia	SO_2; NO_x
Tzanis et al. (2019)	Attica, Greece	$PM_{2.5}$; PM_{10}
Ventura et al. (2019)	Rio de Janeiro, Brazil	$PM_{2.5}$
Wang and Song (2018)	Beijing, China	CO; NO_2; SO_2; O_3; PM_{10}; $PM_{2.5}$
Yeganeh et al. (2018)	Queensland, Australia	NO_2
Zhu et al. (2018)	China	$PM_{2.5}$
Bai et al. (2019)	Beijing, China	$PM_{2.5}$
Liu et al. (2019)	Beijing, China	$PM_{2.5}$; SO_2; NO_2; CO
Qi et al. (2017)	Jing-Jin-Ji, China	$PM_{2.5}$
Qin et al. (2019)	Shanghai, China	$PM_{2.5}$

Appendix B

Levenberg-Marquardt (LM) Algorithm

LM algorithm is based on the backpropagation algorithm which utilises the chain rule to compute the derivatives of the squared error according to the weights and biases in the hidden layers of an ANN. However, backpropagation algorithm suffers from long training time which limits its practicability on real-world tasks. The Newton's method addresses the said issue as it generally provides faster convergence to a solution that minimises the mean squared error. (The following notations are consistent with those in Hagan et al. (1995).)

Recall in Chapter 3 Section 3.2.5 that the performance index of a MLP model is given by:

$$F\left(\mathbf{w}\right) = \sum_{i=1}^{N} \left(y_i(\mathbf{w}) - \hat{y}_i(\mathbf{w})\right)^T \left(y_i(\mathbf{w}) - \hat{y}_i(\mathbf{w})\right) = \mathbf{e}^T\left(\mathbf{w}\right)\mathbf{e}\left(\mathbf{w}\right), \qquad \text{(B.1)}$$

where y and $\hat{y}$ are the actual data and estimated values by the MLP model, respectively, and $\mathbf{w}$ is the vector containing the weights and biases. The goal of calibrating ANN models is to minimise the performance index, $F\left(\mathbf{w}\right)$, by adjusting the values of $\mathbf{w}$.

The Newton's method for optimising the performance index $F\left(\mathbf{w}\right)$ is given by

$$\mathbf{w}_{k+1} = \mathbf{w}_k - \mathbf{A}_k^{-1}\mathbf{g}_k, \qquad \text{(B.2)}$$

Appendix B. Levenberg-Marquardt (LM) Algorithm

where $\mathbf{A}_k \equiv \nabla^2 F(w)\,|_{\mathbf{w}=\mathbf{w}_k}$ and $\mathbf{g}_k = \nabla F(\mathbf{w})\,|_{\mathbf{w}=\mathbf{w}_k}$. Finding the factors $\mathbf{A}_k$ and $\mathbf{g}_k$ requires the determination of the Jacobian and Hessian matrices, respectively.

The j-th element of the gradient can be expressed as

$$[\nabla F(\mathbf{w})]_j = \frac{\partial F(\mathbf{w})}{\partial w_j} = 2\sum_{i=1}^{N} e_i(\mathbf{w}) \frac{\partial e_i(\mathbf{w})}{\partial w_j}. \tag{B.3}$$

The gradient can therefore be expressed in matrix form as

$$\nabla F(\mathbf{w}) = 2\mathbf{J}^T(\mathbf{w})\,\mathbf{e}(\mathbf{w}), \tag{B.4}$$

where

$$\mathbf{J}(\mathbf{w}) = \begin{bmatrix} \dfrac{\partial e_1(\mathbf{w})}{\partial w_1} & \dfrac{\partial e_1(\mathbf{w})}{\partial w_2} & \cdots & \dfrac{\partial e_1(\mathbf{w})}{\partial w_n} \\[2mm] \dfrac{\partial e_2(\mathbf{w})}{\partial w_1} & \dfrac{\partial e_2(\mathbf{w})}{\partial w_2} & \cdots & \dfrac{\partial e_2(\mathbf{w})}{\partial w_n} \\[2mm] \vdots & \vdots & & \vdots \\[2mm] \dfrac{\partial e_N(\mathbf{w})}{\partial w_1} & \dfrac{\partial e_N(\mathbf{w})}{\partial w_2} & \cdots & \dfrac{\partial e_N(\mathbf{w})}{\partial w_n} \end{bmatrix} \tag{B.5}$$

matrix $\mathbf{J}(\mathbf{w})$ is the *Jacobian matrix.*

On the other hand, the factor $\nabla^2 F(w)$ is the Hessian matrix. The k, j element of the Hessian matrix can be expressed as

$$[\nabla^2 F(\mathbf{w})]_{k,j} = \frac{\partial^2 F(\mathbf{w})}{\partial w_k \partial w_j} = 2\sum_{i=1}^{N}\left\{ \frac{\partial e_i(\mathbf{w})}{\partial w_k}\frac{\partial^2 e_i(\mathbf{w})}{\partial w_j} + e_i(\mathbf{w})\frac{\partial e_i(\mathbf{w})}{\partial w_k \partial w_k}\right\}. \tag{B.6}$$

The Hessian matrix can then be written in matrix form as

$$\nabla^2 F(\mathbf{w}) = 2J^T(\mathbf{w})\,\mathbf{J}(\mathbf{w}) + 2\mathbf{S}(\mathbf{w}), \tag{B.7}$$

where

$$\mathbf{S}(\mathbf{w}) = \sum_{i=1}^{N} e_i(\mathbf{w})\,\nabla^2 e_i(\mathbf{w}). \tag{B.8}$$

148

Appendix B. Levenberg-Marquardt (LM) Algorithm

If $\mathbf{S}\left(\mathbf{w}\right)$ is assumed to be small, the Hessian matrix can be approximated as

$$\nabla^2 F\left(\mathbf{w}\right) \cong 2\mathbf{J}^T\left(\mathbf{w}\right)\mathbf{J}\left(\mathbf{w}\right). \tag{B.9}$$

Substituting Eq. (B.9) and Eq. (B.4) into Eq. (B.2), the expression for the *Gauss-Newton* method is obtained:

$$\begin{aligned}
\mathbf{w}_{k+1} &= \mathbf{w}_k - \left[2\mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{J}\left(\mathbf{w}_k\right)\right]^{-1} 2\mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{e}\left(\mathbf{w}_k\right) \\
&= \mathbf{w}_k - \left[\mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{J}\left(\mathbf{w}_k\right)\right]^{-1} \mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{e}\left(\mathbf{w}_k\right).
\end{aligned} \tag{B.10}$$

However, one disadvantage of the Gauss-Newton method is that the matrix $\mathbf{H} = \mathbf{J}^T\mathbf{J}$ may not be invertible. The issue can be addressed by the following modification of the Hessian matrix:

$$\mathbf{G} = \mathbf{H} + \mu\mathbf{I}. \tag{B.11}$$

This leads to the LM algorithm:

$$\mathbf{w}_{k+1} = \mathbf{x}_k - \left[\mathbf{J}^T\left(\mathbf{x}_k\right) + \mu_k\mathbf{I}\right]^{-1}\mathbf{J}^T\left(\mathbf{x}_k\right)\mathbf{e}\left(\mathbf{w}_k\right) \tag{B.12}$$

or

$$\Delta\mathbf{w}_k = -\left[\mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{J}\left(\mathbf{w}_k\right) + \mu_k\mathbf{I}\right]^{-1}\mathbf{J}^T\left(\mathbf{w}_k\right)\mathbf{e}\left(\mathbf{w}_k\right). \tag{B.13}$$

The LM algorithm initialises by setting μ_k to some small value, e.g. $\mu_k = 0.01$. If the chosen parameter yields a smaller $F\left(\mathbf{w}\right)$, then the step is repeated by multiplying μ_k by some factor $\vartheta > 1$, e.g. $\vartheta = 10$. Otherwise, then the same step is repeated by dividing μ_k by ϑ. The iterations of the LM algorithm are summarised as follows:

Appendix B. Levenberg-Marquardt (LM) Algorithm

Algorithm 5 Levenberg-Marquardt Algorithm

---

1. Present all network inputs and compute the network outputs and errors.

2. Compute the sum of squared errors over all inputs, $F(\mathbf{w})$, using Eq. (B.1).

3. Compute the Jacobian matrix using Eq. (B.5)

4. Solve Eq. (B.13) to obtain $\Delta \mathbf{w}_k$.

5. Recompute the sum of squared errors using $\mathbf{w}_k + \Delta \mathbf{w}_k$. If the new sum of squares is smaller than that computed in Step 1, then divide μ by ϑ, let $\mathbf{w}_{k+1} = \mathbf{w}_k + \Delta \mathbf{w}_k$ and go back to Step 1. Otherwise, multiply μ by ϑ and go back to Step 4.

Appendix C

DEFRA Data Selector Online Interface

Figure C.1 shows the online interface of Department of Environment Food & Rural Affairs' data selector page: `https://uk-air.defra.gov.uk/`. The page contains all information needed to access and download data such as monitoring data, descriptive statistics, exceedance statistics and more from the UK-AIR database.
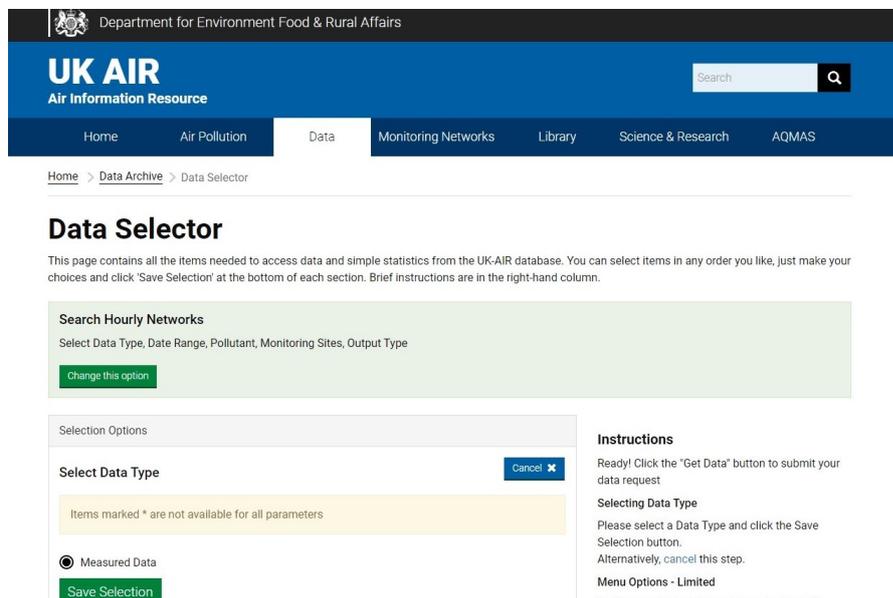
Fig. C.1. The online interface of DEFRA's Data Selector (DEFRA, 2017).

Appendix D

Data Augmentation (DA) Algorithm

Data Augmentation (DA) algorithm is a popular missing data imputation technique (Tanner and Wong, 1987). The following notations are consistent with those in Tanner and Wong (1987).

DA algorithm initially imputes missing data in a random manner according to assumed values of the parameters, and then draws new parameters from a Bayesian posterior distribution based on the observed and imputed values. The desired posterior density is represented as:

$$p\left(\theta \mid y\right) = \int_Z p\left(\theta \mid z, y\right) p\left(z \mid y\right) dz, \qquad (D.1)$$

where $p\left(\theta \mid y\right)$ denotes the posterior density of the parameter θ given the data y, $p\left(z \mid y\right)$ denotes the predictive density of the latent data z given y, and $p\left(\theta \mid z, y\right)$ denotes the conditional density of θ given the augmented data $x = (z, y)$. The predictive density of z can be related to the desired posterior density by

$$p\left(z \mid y\right) = \int_\theta p\left(z \mid \phi, y\right) p\left(\phi \mid y\right) d\phi. \qquad (D.2)$$

In the above equations, the sample space for the latent data z is denoted by Z and the parameter space for θ is denoted by Θ. Substituting Eq. (D.2) into Eq. (D.1) and

Appendix D. Data Augmentation (DA) Algorithm

interchanging the order of integration, it can be seen that $p(\theta \mid y)$ must satisfy the integral equation

$$g(\theta) = \int K(\theta, \phi) g(\phi) d\phi, \tag{D.3}$$

where

$$K(\theta, \phi) = \int p(\phi \mid z, y) p(z \mid \phi, y) dz. \tag{D.4}$$

Let T be the integral transformation that transforms any integrable function f into another integrable function Tf by the equation

$$Tf(\theta) = \int K(\theta, \phi) f(\phi) d\phi. \tag{D.5}$$

The method of successive substitution for solving Eq. (D.3) thereby suggest an iterative method for solving $p(\theta \mid y)$. Namely, begin with any initial approximation $g_0(\theta)$ to $p(\theta \mid y)$, and successively calculate

$$g_{i+1}(\theta) = (Tg_i)(\theta). \tag{D.6}$$

Unfortunately, the integration in Eq. (D.1), Eq. (D.2) and Eq. (D.6) is challenging to perform analytically. Monte Carlo simulation can be employed to perform the integration. As such, Eq. (D.1) motivates the following iterative scheme: Given the current approximation g_i to $p(\theta \mid y)$,

(a) generate a sample $z^{(1)}, \ldots, z^{(m)}$ from the current approximation to the predictive density $p(z \mid y)$; and

b) update the current approximation to $p(\theta \mid y)$ to be the mixture of conditional densities of θ given the augmented data patterns generated in (a), that is,

$$g_{i+1}(\theta) = m^{-1} \sum_{j=1}^{m} p\left(\theta \mid z^{(j)}, y\right). \tag{D.7}$$

The said procedure of alternately simulation missing data values and parameters forms a Markov chain that eventually converges to a distribution (Tanner and Wong, 1987).

Appendix D. Data Augmentation (DA) Algorithm

The iterations of the DA algorithm are summarised as follows:

Algorithm 6 Data Augmentation Algorithm

---

1. Define parameter matrix X

2. Define $X_k^{\#}$ and X^* as the missing and known parts of X at iteration k, respectively

3. Estimate initial values of the parameter set Θ

4. For iteration k, alternately do the following steps:

    a) Imputation: draw $X_k^{\#}$ from their conditional distribution given X^* and Θ_{k-1}

    b) Posterior: draw Θ_k from their posterior distribution given X^* and $X_k^{\#}$

Appendix E

Time Series Plots of Predictors in Experiment 1
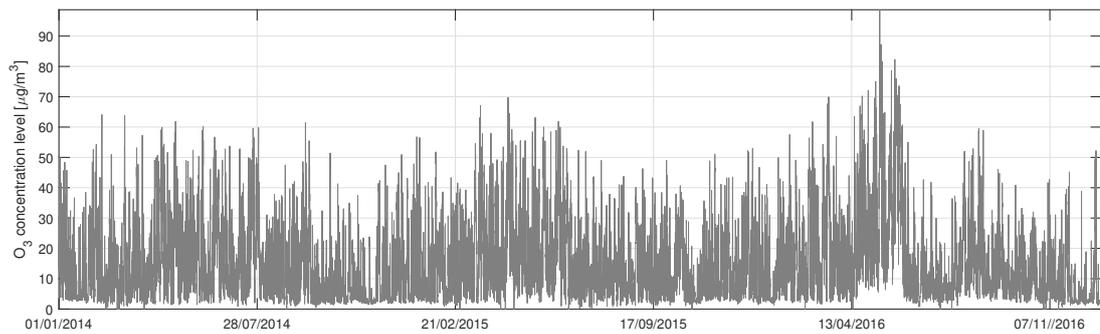
(a) O_3
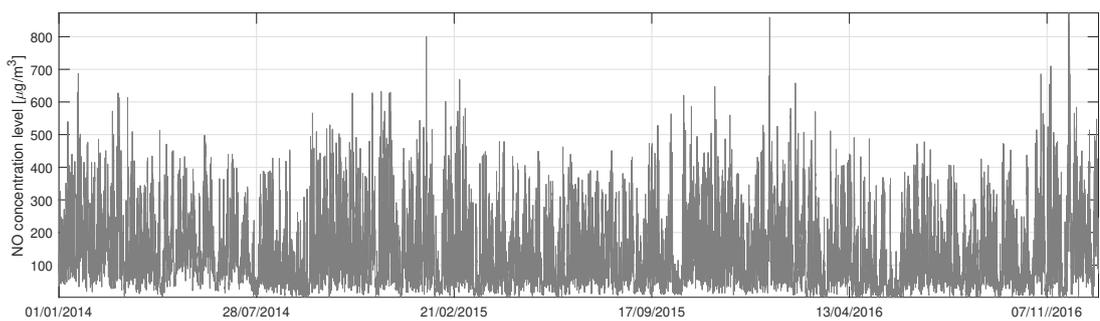
(b) NO

Fig. E.1. (Plots of predictors (in hours) measured in MAR site during from January 2014 to December 2016.
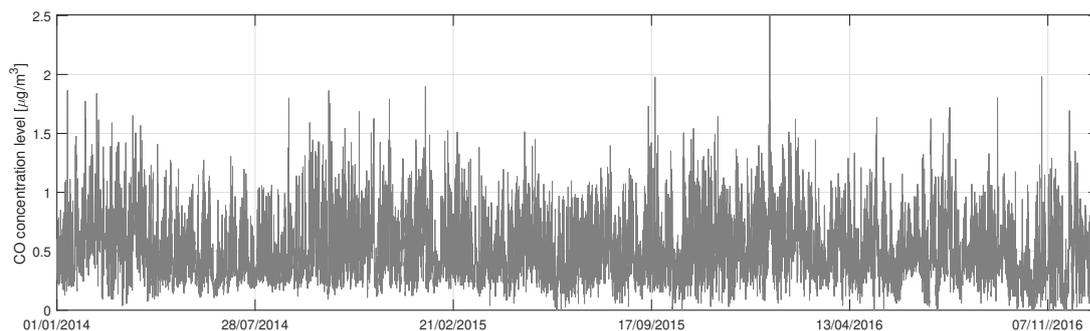
(a) CO
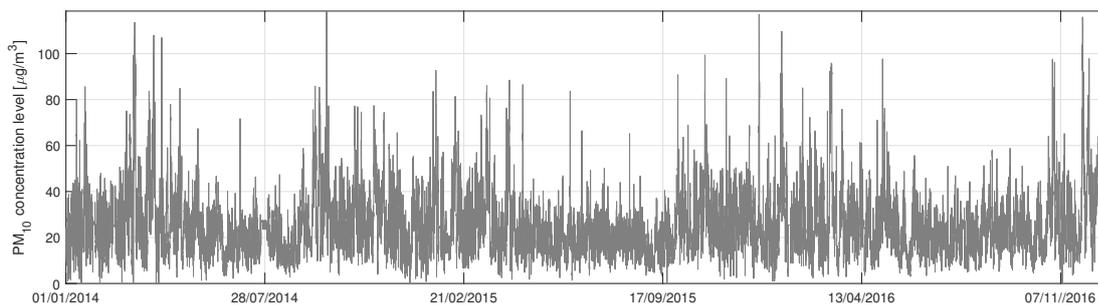
(b) PM$_{10}$
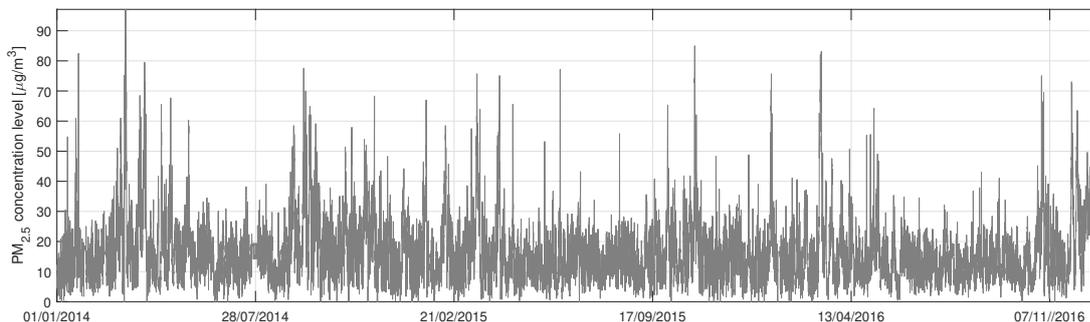
(c) PM$_{2.5}$
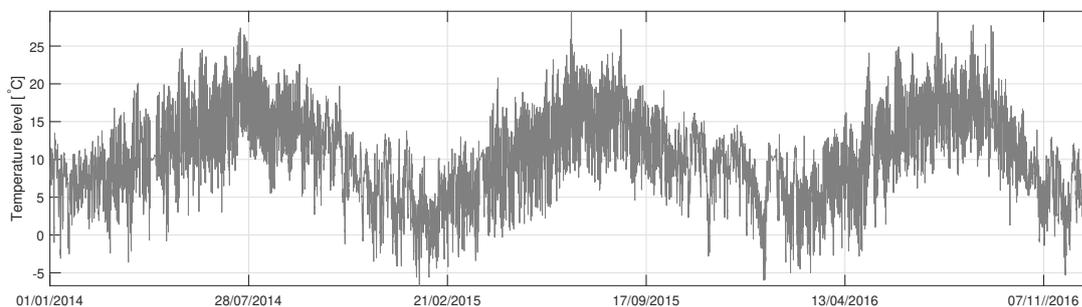
(d) T

Fig. E.2. (continued) Plots of predictors (in hours) measured in MAR site during from January 2014 to December 2016.

158
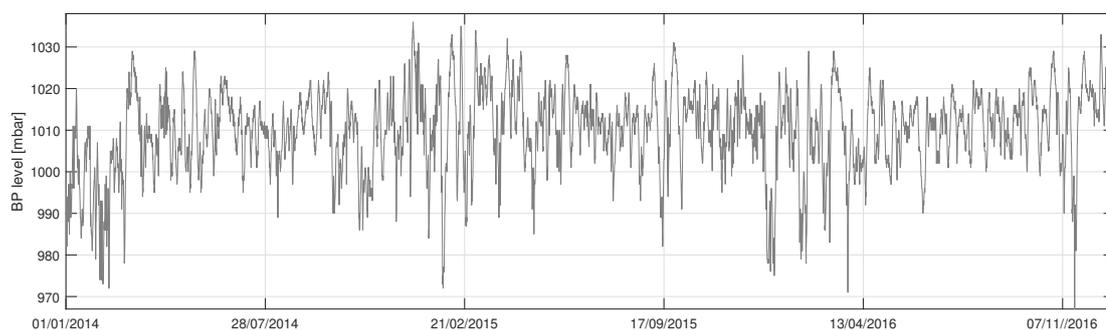
(a) BP

Fig. E.3. (continued) Plots of predictors (in hours) measured in MAR site during from January 2014 to December 2016.
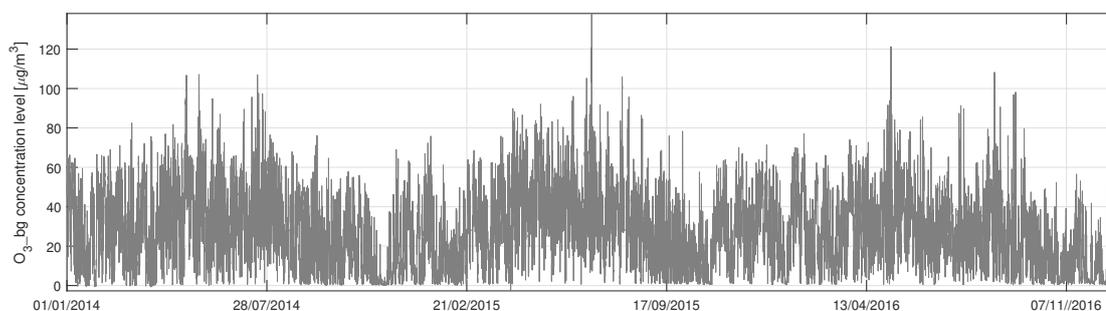
(a) O_3_bg
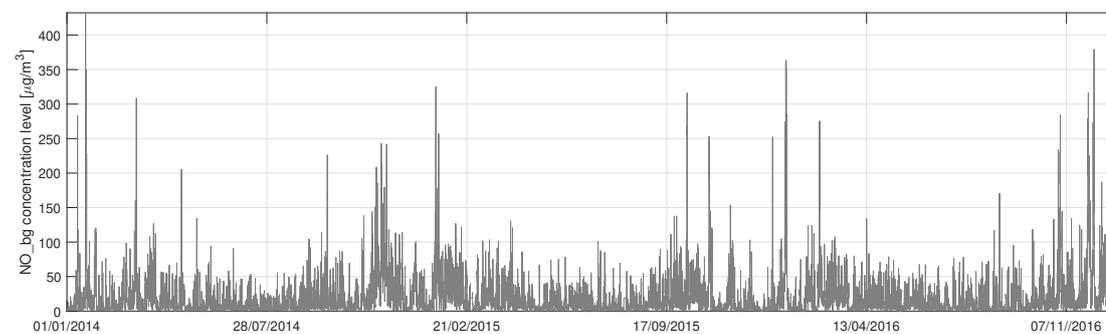
(b) NO_bg

Fig. E.4. Plots of predictors (in hours) measured in BLM site during from January 2014 to December 2016.

(a) NO$_2$_bg

(b) PM$_{10}$_bg

(c) PM$_{2.5}$_bg

Fig. E.5. (continued) Plots of predictors (in hours) measured in BLM site during from January 2014 to December 2016.

(a) T_bg

Fig. E.6. (continued) Plots of predictors (in hours) measured in BLM site during from January 2014 to December 2016.
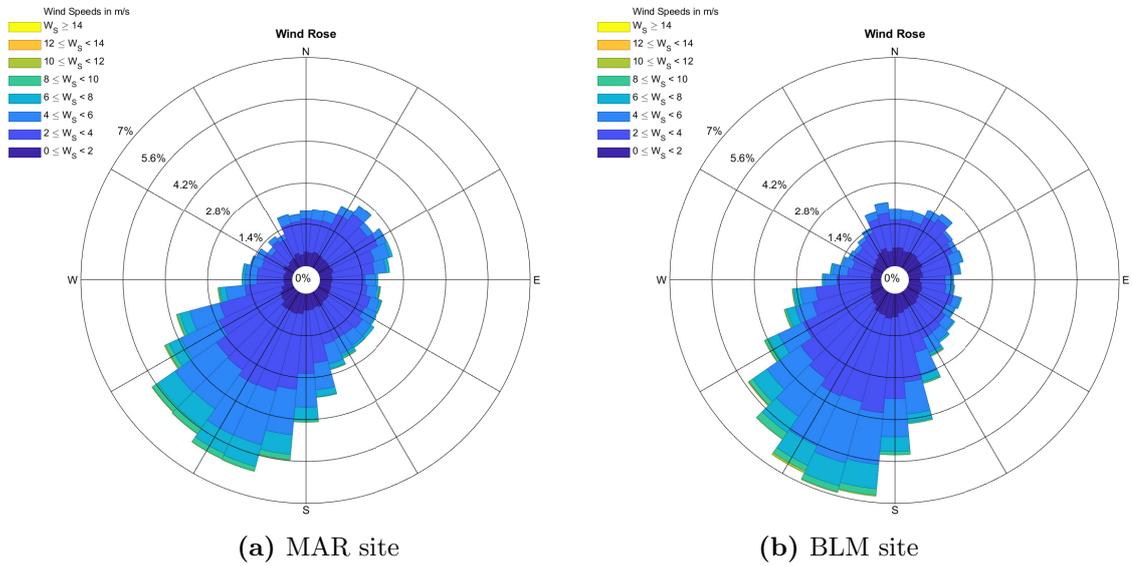
(a) MAR site   **(b)** BLM site

Fig. E.7. Wind rose of the predictors collected in MAR and BLM sites.

Appendix F

Predictor Selection Results in Experiment 1

Figure F.1 presents the results of the regression trees implementation for selecting the optimal predictors. The results were obtained by employing the `stepwiseglm` built-in function in MATLAB which uses both forward and backward stepwise regression. A predictor is added to or removed from the regression model according to the change in the SSE value.
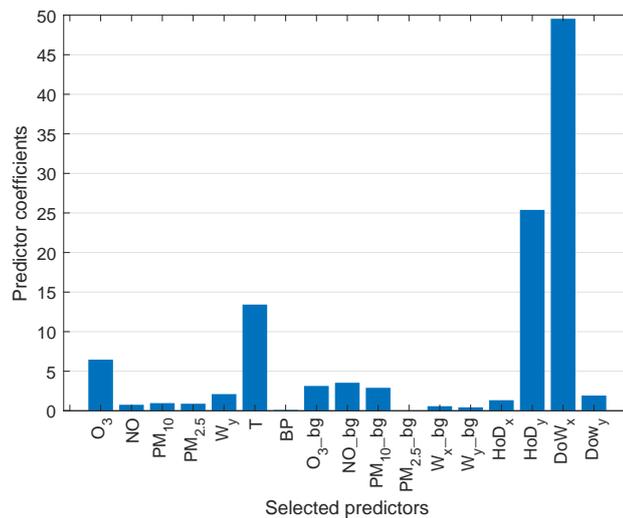
Fig. F.1. Predictor selection results using Stepwise Regression.

The results indicate that all but four predictors, e.g. T_bg, HOD_x_bg and DoW_y_bg, were selected. Note that CO got the largest magnitude of coefficient value, e.g. -122.95, and is therefore not included in Figure F.1 for illustration purposes as it masks the coefficient values of rest of the predictors.

The results of the NCA with regularisation method are shown in Figure F.2. The results were obtained by employing the `fsrnca` built-in function in MATLAB which performs predictor selection for regression based on a diagonal adaptation of NCA with regularisation. The regularisation parameter λ was initially calibrated using five-fold cross-validation with MSE as the loss function.
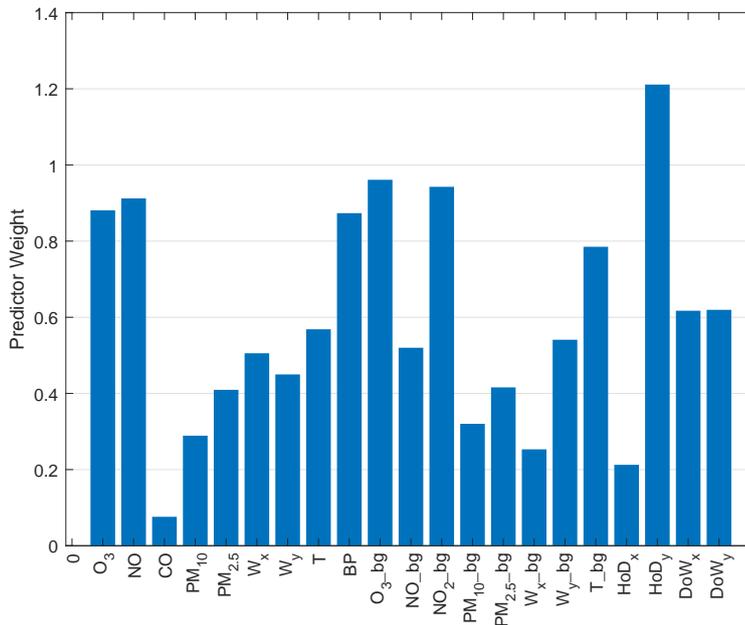
Fig. F.2. Predictor selection results using Neighbourhood Component Analysis with Regularisation.

The importance weights of most predictors are above 0.3. For the experiment, only the predictors with weights greater than 0.4 were selected, amounting to a total of 16 predictors. The selection of the predictor weight criterion is usually undertaken arbitrarily, and it is usually the prerogative of the modeller to determine which value reduces the total count of predictors without fully removing all of predictors.

The results of the SFS technique are shown in Figure F.3. The results were ob-

tained by employing the `sequentialfs` built-in function in MATLAB which sequentially selecting predictors from an initial set of candidate predictors until there is no improvement in the prediction. The said improvement is assessed in terms of the MSE values after performing 10-fold cross-validation.
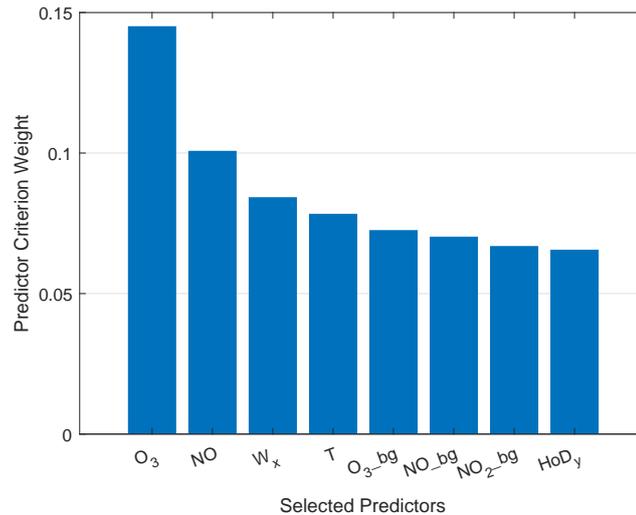
Fig. F.3. Predictor selection results using Sequential Forward Selection.

The optimal predictors are automatically selected by the technique (in this case, a total count of 6 predictors). In contrast to all other predictor selection techniques, SFS selected the least number of predictors which minimise the MSE value. The said technique also took the longest time to implement. This is no surprise considering SFS is model-based technique and involved the training of MLP models multiple times (Maier et al., 2010; Galelli et al., 2014).

The results of the predictor selection technique based on regression trees are shown in Figure F.4. The results were obtained by employing the `fitrtree` built-in function in MATLAB which returns a regression tree based on the initial candidate set of predictors. The `predictorImportance` built-in function was then employed to compute the estimates of predictor weights of the resulting regression tree by summing the change in the MSE due to splits on every predictor and dividing the sum of the number of

branch nodes.
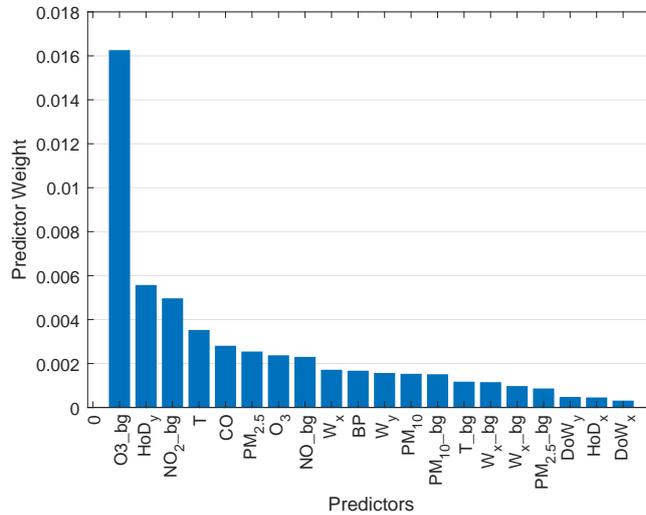
Fig. F.4. Predictor selection results using Regression Tree Analysis.

Only predictors with importance weights above 0.002 were selected. For instance, O_3_bg is considered as the most important predictor, followed by HoD_y. Notice that several background predictors were selected by the RT technique as predictors collected from the same site may not exhibit significant variances from the predictand, NO_2, to trigger binary splits.

Appendix G

Hidden Layer Optimisation Results in Experiment 1

Table G.1. Optimal hidden layer configurations of developed MLP models.

Model / n_h	RMSE [$\mu g/m^3$]										
	20	21	22	23	24	25	26	27	28	29	30
MLP$_{SR}$	21.07	21.23	21.26	21.07	21.07	21.17	21.14	**21.04**	21.13	21.20	21.22
MLP$_{NCA}$	21.70	21.86	21.72	21.81	21.74	**21.64**	21.66	21.79	21.71	21.77	21.74
MLP$_{SFS}$	20.16	20.18	20.18	20.15	20.15	20.12	20.14	20.13	**20.10**	20.10	20.12
MLP$_{RT}$	20.59	20.15	20.54	20.46	20.52	20.32	20.11	20.15	**20.05**	20.14	20.16
MLP$_{ALL}$	20.71	20.57	20.67	20.78	20.54	**20.46**	20.61	20.88	20.68	20.67	20.70
MLP$_{MET}$	36.12	36.10	36.27	36.17	36.03	36.12	36.06	36.05	**35.99**	36.12	36.12
MLP$_{POL}$	21.41	21.35	21.37	21.38	21.33	21.42	21.44	**21.28**	21.44	21.38	21.47
MLP$_{TIME}$	**36.22**	36.23	36.23	36.23	36.22	36.22	36.23	36.23	36.23	36.23	36.23

Appendix H

Time Series Plots of Predictors in Experiment 2.2
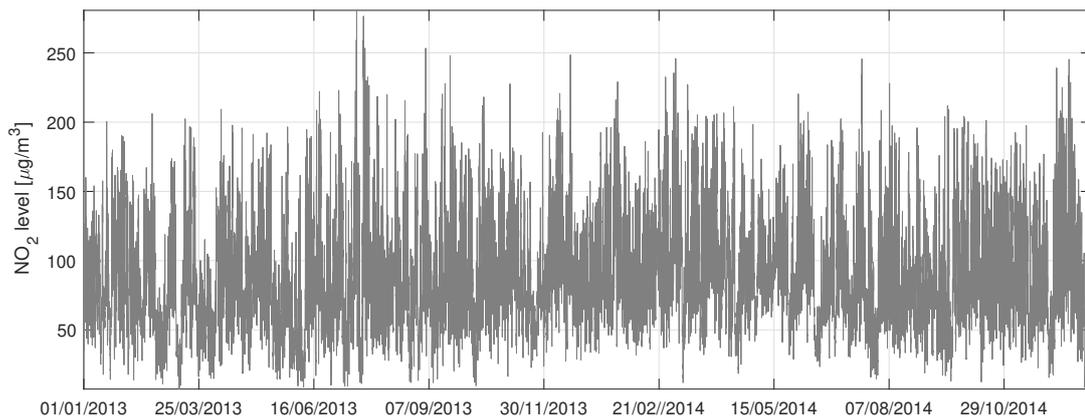
(a) MAR site

Fig. H.1. Plots of hourly NO_2 concentration data measured from January 2013 to December 2014.

(a) KEN site

(b) CAM site

(c) BLM site

Fig. H.2. Plots of hourly NO$_2$ concentration data measured from January 2013 to December 2014.

Appendix H. Time Series Plots of Predictors in Experiment 2.2
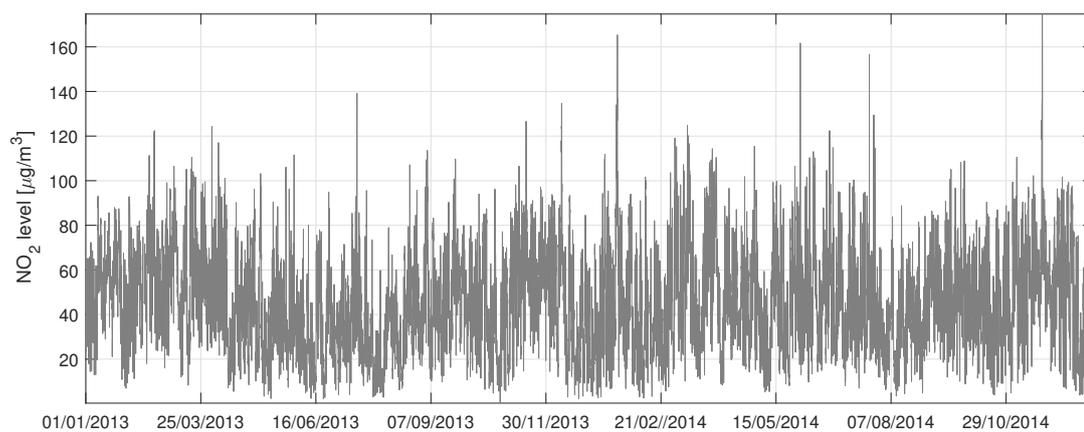
(a) WST site
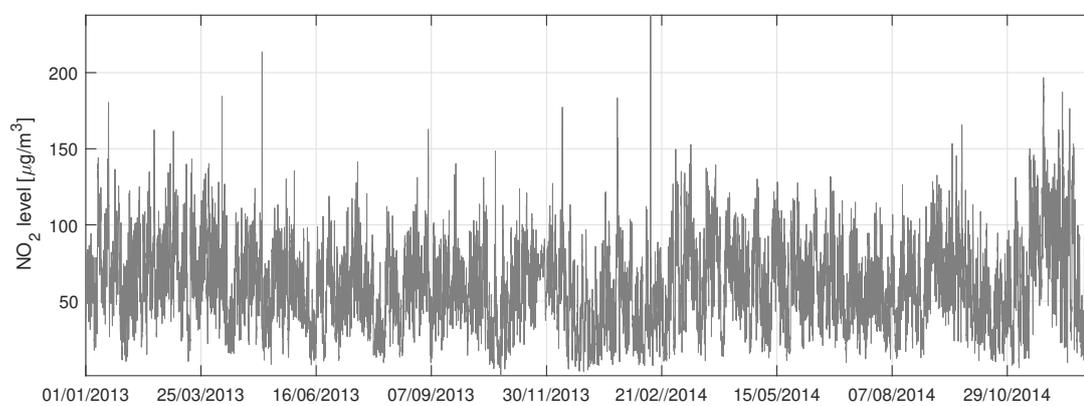
(b) HAM site

Fig. H.3. Plots of hourly NO$_2$ concentration data measured from January 2013 to December 2014.

Appendix I

Optimisation Results for the selection of Daubechies Wavelet Types in Experiment 2

Appendix I. Optimisation Results for the selection of Daubechies Wavelet Types in Experiment 2

Table I.1. The $\mathrm{std}\left(A_j\right)/\mathrm{std}\left(S\right)$ ratios based on various Daubechies wavelet configurations in Experiment 2

Site	Level	Wavelet type		
		Daubechies Db4	Daubechies Db5	Daubechies Db6
MAR	3	0.9013	0.9075	0.9032
	4	0.6962	0.6995	0.6985
	5	0.6185	0.6178	0.6219
KEN	3	0.8862	0.9019	0.8894
	4	0.7982	0.8025	0.8014
	5	0.7263	0.7265	0.7233
CAM	3	0.8818	0.8798	0.8797
	4	0.6949	0.7029	0.6959
	5	0.6273	0.6283	0.6296
BLM	3	0.8784	0.9102	0.8821
	4	0.7896	0.7961	0.7920
	5	0.7304	0.7309	0.7289
WST	3	0.8885	0.9028	0.8884
	4	0.7872	0.7933	0.7902
	5	0.7250	0.7221	0.7197
HAM	3	0.9157	0.9283	0.9176
	4	0.8332	0.8344	0.8322
	5	0.7607	0.7582	0.8322

Appendix J

Discrete Wavelet Decomposition Results in Experiment 2.2

Figure J.1 through Figure J.5 show the 5-level wavelet decomposition of the first 1000 samples of the collected NO_2 data from KEN, CAM, BLM, WST and HAM sites. The said samples were transformed using various Daubechies wavelets.
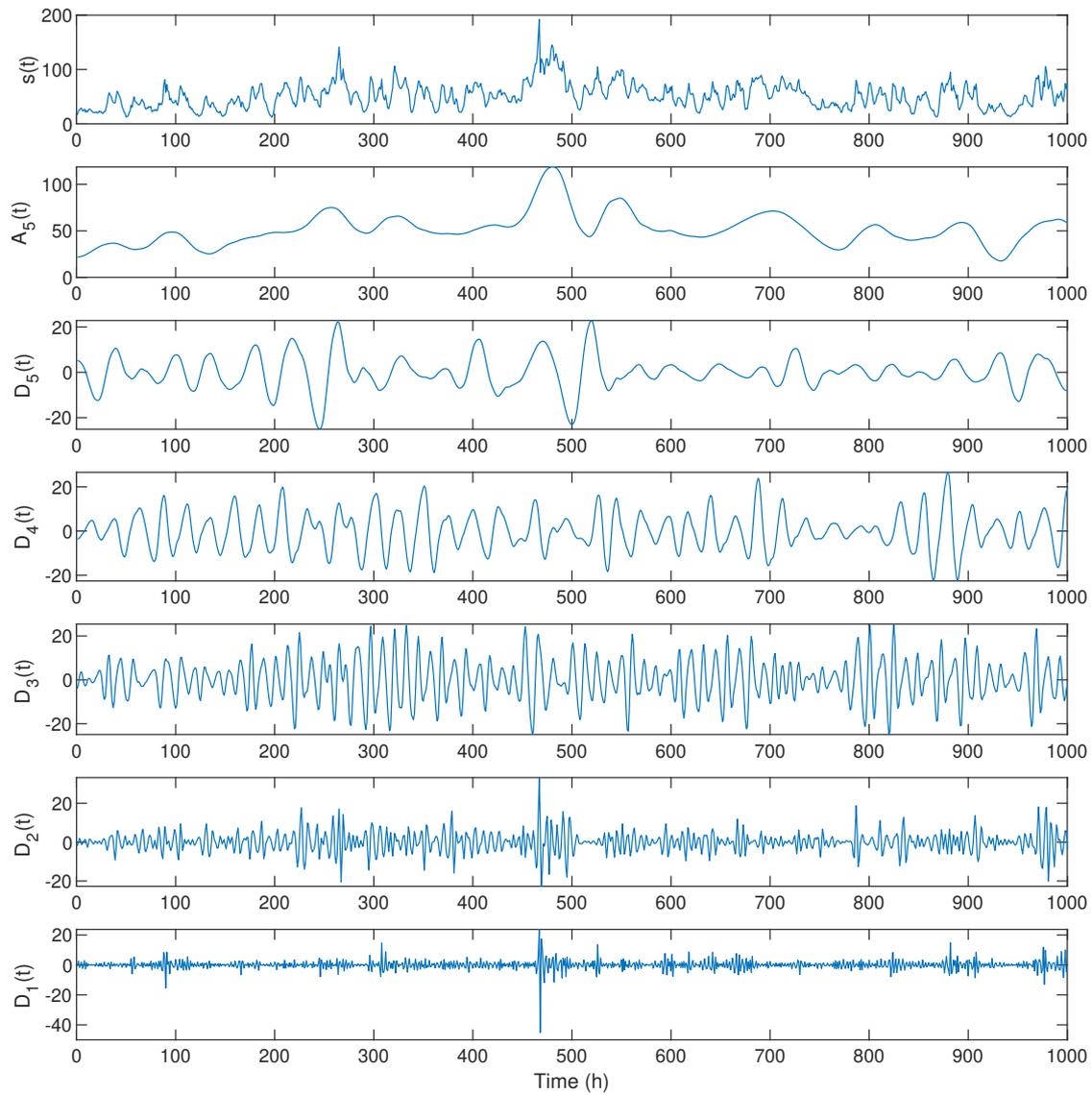
Fig. J.1. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from KEN site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Fig. J.2. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from CAM site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Fig. J.3. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from BLM site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Fig. J.4. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from WST site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Fig. J.5. Wavelet decomposition of the first 1000 samples of the collected NO_2 data from HAM site, s(t); D_1 to D_5 represent the detailed coefficients, and A_5 the coarse approximation of s(t) on the fifth level.

Appendix K

Correlation Analysis Results in Experiment 3

Appendix K. Correlation Analysis Results in Experiment 3

Table K.1. Pearson correlation scores between NO_2 data collected from 19 monitoring sites in Greater London area.

Sites	Pearson Correlation Coefficient																		
	BT4	CD3	CD1	CT3	CT6	GN0	GR5	GR9	GR8	HK6	IS2	BX1	BL0	GR4	MY1	KC1	WM0	TH2	TH4
BT4	1.000	0.554	0.669	0.717	0.613	0.529	0.618	0.759	0.556	0.642	0.647	0.634	0.611	0.558	0.482	0.726	0.626	0.613	0.655
CD3	0.554	1.000	0.631	0.637	0.450	0.686	0.633	0.592	0.482	0.597	0.548	0.554	0.696	0.538	0.222	0.545	0.672	0.704	0.643
CD1	0.669	0.631	1.000	0.633	0.554	0.604	0.621	0.675	0.408	0.481	0.552	0.565	0.706	0.520	0.265	0.609	0.652	0.630	0.635
CT3	0.717	0.637	0.633	1.000	0.609	0.651	0.778	0.738	0.585	0.794	0.694	0.784	0.819	0.746	0.389	0.817	0.808	0.764	0.707
CT6	0.613	0.450	0.554	0.609	1.000	0.286	0.396	0.522	0.547	0.590	0.594	0.409	0.422	0.353	0.643	0.422	0.413	0.325	0.499
GN0	0.529	0.686	0.604	0.651	0.286	1.000	0.818	0.665	0.385	0.587	0.463	0.702	0.767	0.714	0.012	0.637	0.748	0.856	0.625
GR5	0.618	0.633	0.621	0.778	0.396	0.818	1.000	0.714	0.482	0.709	0.576	0.794	0.814	0.821	0.214	0.768	0.802	0.853	0.663
GR9	0.759	0.592	0.675	0.738	0.522	0.665	0.714	1.000	0.412	0.601	0.527	0.777	0.708	0.735	0.234	0.795	0.738	0.690	0.614
GR8	0.556	0.482	0.408	0.585	0.547	0.385	0.482	0.412	1.000	0.701	0.802	0.585	0.430	0.510	0.675	0.468	0.443	0.449	0.784
HK6	0.642	0.597	0.481	0.794	0.590	0.587	0.709	0.601	0.701	1.000	0.739	0.709	0.692	0.694	0.526	0.694	0.676	0.693	0.698
IS2	0.647	0.548	0.552	0.694	0.594	0.463	0.576	0.527	0.802	0.739	1.000	0.625	0.588	0.580	0.655	0.634	0.529	0.535	0.782
BX1	0.634	0.554	0.565	0.784	0.409	0.702	0.794	0.777	0.585	0.709	0.625	1.000	0.730	0.865	0.262	0.826	0.773	0.739	0.719
BL0	0.611	0.696	0.706	0.819	0.422	0.767	0.814	0.708	0.430	0.692	0.588	0.730	1.000	0.729	0.150	0.787	0.848	0.833	0.656
GR4	0.558	0.538	0.520	0.746	0.353	0.714	0.821	0.735	0.510	0.694	0.580	0.865	0.729	1.000	0.225	0.820	0.772	0.752	0.649
MY1	0.482	0.222	0.265	0.389	0.643	0.012	0.214	0.234	0.675	0.526	0.655	0.262	0.150	0.225	1.000	0.272	0.138	0.108	0.480
KC1	0.726	0.545	0.609	0.817	0.422	0.637	0.768	0.795	0.468	0.694	0.634	0.826	0.787	0.820	0.272	1.000	0.800	0.728	0.644
WM0	0.626	0.672	0.652	0.808	0.413	0.748	0.802	0.738	0.443	0.676	0.529	0.773	0.848	0.772	0.138	0.800	1.000	0.812	0.657
TH2	0.613	0.704	0.630	0.764	0.325	0.856	0.853	0.690	0.449	0.693	0.535	0.739	0.833	0.752	0.108	0.728	0.812	1.000	0.688
TH4	0.655	0.643	0.635	0.707	0.499	0.625	0.663	0.614	0.784	0.698	0.782	0.719	0.656	0.649	0.480	0.644	0.657	0.688	1.000

Appendix K. Correlation Analysis Results in Experiment 3

Appendix L

Construction of the AP Level Spatial Distribution Map in Experiment 3

As mentioned in Chapter 3 Section 5.4.2, the spatial distribution of the hourly NO_2 concentration estimates of the proposed and benchmark models in Experiment 3 (see Figure 5.26) was produced by utilising inverse distance weighted interpolation. The entire procedure was carried using ArcGIS software version 10.6 (Environmental Systems Research Institute, 2020).

To generate a full surface distribution of hourly NO_2 concentration of the test dataset in Experiment 3, information from several sites surrounding the 13 examined and selected monitoring sites across the Greater London area. Details regarding the surrounding sites are provided in Table L.1. Fixed NO_2 values collected from the said surrounding sites have been used while the selected sites have varying ones based on the results of the proposed and benchmark models.

Appendix L. Construction of the AP Level Spatial Distribution Map in Experiment 3

Table L.1. Air quality monitoring sites and coordinates utilised for generating the AP spatial distribution in Figure 5.26.

Site name	Latitude / Longitude
Bexley - Slade Green	51° 27 ′ 57.71 ″ N, 0° 11 ′ 5.30 ″ E
Camden - Bloomsbury	51° 31 ′ 20.24 ″ N, 0° 7 ′ 33.20 ″ W
City of London - Sir John Cass	51° 30 ′ 49.86 ″ N, 0° 4 ′ 39.97 ″ W
Greenwich -Eltham	51° 27 ′ 9.29 ″ N, 0° 4 ′ 14.76 ″ E
Kensington and Chelsea - North Ken	51° 31 ′ 15.78 ″ N, 0° 12 ′ 48.57 ″ W
Westminster - Horseferry Road	51° 29 ′ 40.81 ″ N, 0° 7 ′ 54.95 ″ E
Thurrock - Stanford-le-Hope	51° 31 ′ 5.38 ″ N, 0° 26 ′ 22.37 ″ E
Castle Point - Hadleigh	51° 33 ′ 14.63 ″ N, 0° 36 ′ 21.82 ″ E
Brentwood - Brentwood Town Hall	51° 37 ′ 7.27 ″ N, 0° 18 ′ 27.89 ″ E
Redbridge - Fullwell Cross	51° 35 ′ 38.17 ″ N, 0° 5 ′ 7.86 ″ E
Harrow - Stanmore	51° 37 ′ 2.38 ″ N, 0° 17 ′ 55.59 ″ E
Hounslow and Ealing - Gunnersbury Avenue	51° 30 ′ 2.43 ″ N, 0° 17 ′ 3.78 ″ W
Windsor and Maidenhead - Clarence Road	51° 28 ′ 48.64 ″ N, 0° 37 ′ 25.72 ″ W
Reigate and Banstead - Horley	51° 9 ′ 57.09 ″ N, 0° 10 ′ 3.84 ″ W
Sevenoaks - Bat and Ball	51° 17 ′ 19.56 ″ N, 0° 11 ′ 36.32 ″ E
Reigate and Banstead - Reigate High St.	51° 14 ′ 15.60 ″ N, 0° 12 ′ 21.64 ″ W
Sutton - Worcester Park	51° 22 ′ 40.52 ″ N, 0° 12 ′ 25.49 ″ W
Lewisham - Catford	51° 26 ′ 43.68 ″ N, 0° 1 ′ 12.96 ″ W
Enfield - Bush Hill Park	51° 38 ′ 42.13 ″ N, 0° 3 ′ 58.25 ″ W
Hillingdon - Keats Way	51° 29 ′ 46.71 ″ N, 0° 27 ′ 38.97 ″ E
Borehamwoord Meadow Park	51° 39 ′ 40.42 ″ N, 0° 16 ′ 14.41 ″ W
Chatham Roadside	51° 22 ′ 27.35 ″ N, 0° 32 ′ 52.69 ″ E
London Teddington	51° 25 ′ 31.03 ″ N, 0° 20 ′ 44.18 ″ W

Bibliography

WHO, WHO — Ambient (outdoor) air quality and health, 2016. URL: `http://www.who.int/mediacentre/factsheets/fs313/en/`.

OECD, Policy Highlights - The economic consequences of outdoor air pollution, Technical Report, Organisation for Economic Co-operation and Development, 2016.

B. Durak, A cartoon drawing of a biological neuron, 2020. URL: `http://cs231n.github.io/neural-networks-1/`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1364815210000411`. doi:10.1016/J.ENVSOFT.2010.02.003.

S. M. S. Cabaneros, J. K. S. Calautit, B. R. Hughes, Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO2 Pollution, Energy Procedia 142 (2017) 3524–3530. URL: `https://www.sciencedirect.com/science/article/pii/S187661021735974X`. doi:10.1016/J.EGYPRO.2017.12.240.

DEFRA, The Air Quality Data Validation and Ratification Process, 2017. URL: `https://uk-air.defra.gov.uk/assets/documents/Data{_}Validation{_}and{_}Ratification{_}Process{_}Apr{_}2017.pdf`.

DEFRA, AIR QUALITY EXPERT GROUP Nitrogen Dioxide in the United Kingdom, 2004. URL: `www.defra.gov.uk/environment/airquality`.

Bibliography

King's College London, London Air Quality Network - King's College London, 2019. URL: `http://www.londonair.org.uk/LondonAir/Default.aspx`.

G. Oliveri Conti, B. Heibati, I. Kloog, M. Fiore, M. Ferrante, A review of AirQ Models and their applications for forecasting the air pollution health outcomes, Environmental Science and Pollution Research 24 (2017) 6426–6445. doi:`10.1007/s11356-016-8180-1`.

N. Fann, C. M. Fulcher, B. J. Hubbell, The influence of location, source, and emission type in estimates of the human health benefits of reducing a ton of air pollution, Air Quality, Atmosphere & Health 2 (2009) 169–176. doi:`10.1007/s11869-009-0044-0`.

M. Crippa, G. Janssens-Maenhout, F. Dentener, D. Guizzardi, K. Sindelarova, M. Muntean, R. Van Dingenen, C. Granier, Forty years of improvements in European air quality: regional policy-industry interactions with global impacts, Atmos. Chem. Phys 16 (2016) 3825–3841. URL: `www.atmos-chem-phys.net/16/3825/2016/`. doi:`10.5194/acp-16-3825-2016`.

A. Baklanov, O. Hänninen, L. H. Slørdal, J. Kukkonen, N. Bjergene, B. Fay, S. Finardi, S. C. Hoe, M. Jantunen, A. Karppinen, A. Rasmussen, A. Skouloudis, R. S. Sokhi, J. H. Sørensen, V. Ødegaard, Integrated systems for forecasting urban meteorology, air pollution and population exposure, Atmos. Chem. Phys. Atmos. Chem. Phys. 7 (2007) 855–874. URL: `www.atmos-chem-phys.net/7/855/2007/`.

F. A. Gers, D. Eck, J. Schmidhuber, Applying LSTM to Time Series Predictable through Time-Window Approaches, in: G. Dorffner, H. Bischof, K. Hornik (Eds.), Artif. Neural Networks — ICANN 2001, Springer Berlin Heidelberg, Berlin, 2001, pp. 669–676.

Nagl, Thielen, Schneider, DIRECTORATE GENERAL FOR INTERNAL POLICIES: Implementation of the Ambient Air Quality Directive, Technical Report, 2016. URL: `http://www.europarl.europa.eu/RegData/etudes/STUD/2016/578986/IPOL{_}STU(2016)578986{_}EN.pdf`.

188

Bibliography

H. Chen, Q. Li, J. S. Kaufman, J. Wang, R. Copes, Y. Su, T. Benmarhnia, Effect of air quality alerts on human health: a regression discontinuity analysis in Toronto, Canada, Lancet Planet. Heal. 2 (2018) e19–e26. URL: `https://www.sciencedirect.com/science/article/pii/S2542519617301857?via{%}3Dihub`. doi:10.1016/S2542-5196(17)30185-7.

J. McLaren, I. Williams, The impact of communicating information about air pollution events on public health, Sci. Total Environ. 538 (2015) 478–491. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0048969715304903`. doi:10.1016/j.scitotenv.2015.07.149.

F. A. Gers, D. Eck, J. Schmidhuber, Applying LSTM to Time Series Predictable through Time-Window Approaches, in: G. Dorffner, H. Bischof, K. Hornik (Eds.), Artif. Neural Networks — ICANN 2001, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 669–676.

F. J. Kelly, G. W. Fuller, H. A. Walton, J. C. Fussell, Monitoring air pollution: Use of early warning systems for public health, Respirology 17 (2012) 7–19. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1440-1843.2011.02065.x`. doi:10.1111/j.1440-1843.2011.02065.x. arXiv:`https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1843.2011.02065.x`.

J. Colls, Air Pollution 2nd Edition, Spon Press, 29 West 35th Street, New York, NY 10001, 2001.

M. Gardner, S. Dorling, Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, Atmos. Environ. 32 (1998) 2627–2636. URL: `http://www.sciencedirect.com/science/article/pii/S1352231097004470`. doi:10.1016/S1352-2310(97)00447-0.

B. Sportisse, Fundamentals in Air Pollution: From Processes to Modelling, Springer Science+Business Media, 2010. URL: `www.springer.com`. doi:10.1007/978-90-481-2970-6.

Bibliography

M. Gardner, The Advantages of Artificial Neural Network and Regression Tree Based Air Quality Models, Ph.D. thesis, University of East Anglia, 1999.

Z. Jacobson, Development and Application of a New Air Pollution Modeling System - II. Aerosol Module Structure and Design, Atmos. Environ. 31 (1997) 131–144. doi:10.16011/j.cnki.jjwt.2015.07.004.

M. E. Chang, C. Cardelino, Application of the Urban Airshed Model to forecasting next-day peak ozone concentrations in Atlanta, Georgia., J. Air Waste Manag. Assoc. 50 (2000) 2010–2024. URL: http://search.ebscohost.com/login.aspx?direct=true{&}db=cmedm{&}AN=11111345{&}site=ehost-live. doi:10.1080/10473289.2000.10464219.

M. T. Chuang, Y. Zhang, D. Kang, Application of WRF/Chem-MADRID for real-time air quality forecasting over the Southeastern United States, Atmos. Environ. 45 (2011) 6241–6250. URL: http://dx.doi.org/10.1016/j.atmosenv.2011.06.071. doi:10.1016/j.atmosenv.2011.06.071.

S. F. Mueller, J. W. Mallard, Contributions of natural emissions to ozone and PM2.5 as simulated by the community multiscale air quality (CMAQ) model, Environ. Sci. Technol. 45 (2011) 4817–4823. doi:10.1021/es103645m.

K. Ganev, M. Prodanova, D. Syrakov, N. Miloshev, Air pollution transport in the Balkan region and country-to-country pollution exchange between Romania, Bulgaria and Greece, Ecol. Modell. 217 (2008) 255–269. URL: https://www.sciencedirect.com/science/article/pii/S030438000800255X. doi:10.1016/J.ECOLMODEL.2008.06.029.

I. Lagzi, R. Meszaros, G. Gelybo, A. Leelossy, Atmospheric Chemistry, Eötvös Loránd University, Hungary, 2013. URL: http://elte.prompt.hu/sites/default/files/tananyagok/AtmosphericChemistry/index.html.

H.-d. H. W.-z. Lu, Y. Xue, Prediction of particulate matters at urban intersection by using multilayer perceptron model based on principal components, Stoch. En-

Bibliography

viron. Res. Risk Assess. (2015) 2107–2114. URL: `http://dx.doi.org/10.1007/s00477-014-0989-x`. doi:`10.1007/s00477-014-0989-x`.

D. Luecken, W. Hutzell, G. Gipson, Development and analysis of air quality modeling simulations for hazardous air pollutants, Atmospheric Environment 40 (2006) 5087–5096. URL: `https://www.sciencedirect.com/science/article/pii/S1352231006000689`. doi:`10.1016/J.ATMOSENV.2005.12.044`.

J. Chave, S. Levin, Scale and Scaling in Ecological and Economic Systems, Technical Report, 2003.

National Research Council, Models in environmental regulatory decision making, National Academies Press, 2007.

D. Valput, R. Navares, J. L. Aznarte, Forecasting hourly NO 2 concentrations by ensembling neural networks and mesoscale models, Neural Computing and Applications (2019). URL: `https://doi.org/10.1007/s00521-019-04442-z`. doi:`10.1007/s00521-019-04442-z`.

Y. Xu, P. Du, J. Wang, Research and application of a hybrid model based on dynamic fuzzy synthetic evaluation for establishing air quality forecasting and early warning system: A case study in China, Environ. Pollut. 223 (2017) 435–448. doi:`10.1016/j.envpol.2017.01.043`.

K. De Ridder, U. Kumar, D. Lauwaet, L. Blyth, W. Lefebvre, Kalman filter-based air quality forecast adjustment, Atmos. Environ. 50 (2012) 381–384. doi:`10.1016/j.atmosenv.2012.01.032`.

R. Beelen, G. Hoek, E. Pebesma, D. Vienneau, K. de Hoogh, D. J. Briggs, Mapping of background air pollution at a fine spatial scale across the European Union, Sci. Total Environ. 407 (2009) 1852–1867. URL: `http://dx.doi.org/10.1016/j.scitotenv.2008.11.048`. doi:`10.1016/j.scitotenv.2008.11.048`.

A.-L. 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.

Bibliography

Softw. 22 (2007) 1261–1269. URL: https://www.sciencedirect.com/science/article/pii/S1364815206001976. doi:10.1016/J.ENVSOFT.2006.08.002.

H. J. S. Fernando, M. C. Mammarella, G. Grandoni, P. Fedele, R. Di Marco, R. Dimitrova, P. Hyde, Forecasting PM10 in metropolitan areas: Efficacy of neural networks., Environ. Pollut. 163 (2012) 62–7. URL: http://www.ncbi.nlm.nih.gov/pubmed/22325432. doi:10.1016/j.envpol.2011.12.018.

S. Gocheva-Ilieva, A. V. Ivanov, D. S. Voynikova, D. T. Boyadzhiev, Time series analysis and forecasting for air pollution in small urban area: an sarima and factor analysis approach, Stochastic Environmental Research and Risk Assessment 28 (2014) 1045–1060. URL: https://search.proquest.com/docview/1513298134?accountid=14116, copyright - Springer-Verlag Berlin Heidelberg 2014; Last updated - 2014-08-22.

H. T. Shahraiyni, S. Sodoudi, Statistical modeling approaches for pm10 prediction in urban areas; A review of 21st-century studies, Atmosphere (Basel). 7 (2016) 10–13. doi:10.3390/atmos7020015.

H. Abderrahim, M. R. Chellali, A. Hamou, Forecasting PM 10 in Algiers : efficacy of multilayer perceptron networks, Environ. Sci. Pollut. Res. 23 (2016) 1634–1641. doi:10.1007/s11356-015-5406-6.

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.

IEEE Spectrum, The 2018 Top Programming Languages - IEEE Spectrum, 2018. URL: https://spectrum.ieee.org/at-work/innovation/the-2018-top-programming-languages.

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.

192

Bibliography

G. Corani, Air quality prediction in Milan : feed-forward neural networks , pruned neural networks and lazy learning, Ecol. Modell. 185 (2005) 513–529. doi:`10.1016/ j.ecolmodel.2005.01.008`.

L. A. Zadeh, Fuzzy logic, Computer 21 (1988) 83–93.

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. URL: `https://linkinghub.elsevier.com/retrieve/pii/ S1364815218306352`. doi:`10.1016/j.envsoft.2019.06.014`.

M. T. Hagan, H. B. Demuth, M. H. Beale, Neural Network Design, Bost. Massachusetts PWS 2 (1995) 734. URL: `http://books.google.ru/books? id=bUNJAAAACAAJ{%}5Cnhttp://ecee.colorado.edu/academics/schedules/ ECEN5120.pdf`. doi:`10.1007/1-84628-303-5`.

K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks 2 (1989) 359–366. doi:`10.1016/0893-6080(89) 90020-8`. `arXiv:arXiv:1011.1669v3`.

H. R. Maier, G. C. Dandy, Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications, Environ. Model. Softw. 15 (2000) 101–124. URL: `https://www.sciencedirect.com/science/ article/pii/S1364815299000079`. doi:`10.1016/S1364-8152(99)00007-9`.

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). URL: `https://doi.org/10.1016/j.atmosenv.2018.07.058`. doi:`10.1016/ j.atmosenv.2018.07.058`.

H. Pfeiffer, G. Baumbach, L. Sarachaga-ruiz, S. Kleanthous, O. Poulida, E. Beyaz, Neural modelling of the spatial distribution of air pollutants, Atmos. Environ. 43 (2009) 3289–3297. URL: `http://dx.doi.org/10.1016/j.atmosenv.2008.05.073`. doi:`10.1016/j.atmosenv.2008.05.073`.

Bibliography

C. G. Tzanis, A. Alimissis, K. Philippopoulos, D. Deligiorgi, Applying linear and nonlinear models for the estimation of particulate matter variability *, Environ. Pollut. (2019). URL: `https://doi.org/10.1016/j.envpol.2018.11.080`. doi:`10.1016/j.envpol.2018.11.080`.

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. URL: `http://dx.doi.org/10.1016/j.envsci.2016.03.008`. doi:`10.1016/j.envsci.2016.03.008`.

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`.

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`.

S. Osowski, K. Garanty, Forecasting of the daily meteorological pollution using wavelets and support vector machine, Eng. Appl. Artif. Intell. 20 (2007) 745–755. doi:`10.1016/j.engappai.2006.10.008`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S0160412019300650`. doi:`10.1016/J.ENVINT.2019.104909`.

L. N. Araujo, J. T. Belotti, T. A. Alves, Y. d. S. Tadano, H. Siqueira, Ensemble method based on Artificial Neural Networks to estimate air pollution health risks, Environmental Modelling and Software (2020). doi:`10.1016/j.envsoft.2019.104567`.

Bibliography

W. Date, City 'supersites' to boost air quality monitoring, 2018. URL: `https://airqualitynews.com/2018/01/23/city-supersites-boost-air-quality-monitoring/`.

Z. Jinran, Monitoring network to be further expanded - China - Chinadaily.com.cn, 2017. URL: `https://www.chinadaily.com.cn/china/2017-04/07/content{_}28827498.htm`.

A. Upadhay, Air Pollution Crisis: Does India Have Enough Air Quality Monitoring Stations? — Air Pollution, 2019. URL: `https://swachhindia.ndtv.com/air-quality-monitoring-india-importance-challenges-practices-31871/`.

RCP, Every breath we take: the lifelong impact of air pollution, Working Party Report, Technical Report February, The Royal College of Physicians, 2016.

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`.

N. Kumar, A. Middey, P. S. Rao, Prediction and examination of seasonal variation of ozone with meteorological parameter through artificial neural network at NEERI, Nagpur, India, Urban Clim. 20 (2017) 148–167. URL: `http://dx.doi.org/10.1016/j.uclim.2017.04.003`. doi:`10.1016/j.uclim.2017.04.003`.

D. Dominick, M. Talib Latif, H. Juahir, A. Z. Aris, S. Zain, An assessment of influence of meteorological factors on PM10 and NO2 at selected stations in Malaysia, Sustain. Environ. Res. 22 (2012) 305–315.

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`.

WHO, Ambient air pollution, 2018. URL: `http://www.who.int/airpollution/ambient/en/`.

195

Bibliography

A. N. Al–Dabbous, P. Kumar, A. R. Khan, Prediction of airborne nanoparticles at roadside location using a feed–forward artificial neural network, Atmos. Pollut. Res. 8 (2017) 446–454. URL: `http://linkinghub.elsevier.com/retrieve/pii/S1309104216304172`. doi:`10.1016/j.apr.2016.11.004`.

I. Juhos, L. Makra, B. Tóth, The behaviour of the multi-layer perceptron and the support vector regression learning methods in the prediction of NO and NO2 concentrations in Szeged, Hungary, Neural Comput. Appl. 18 (2009) 193–205. URL: `http://dx.doi.org/10.1007/s00521-007-0171-1`. doi:`10.1007/s00521-007-0171-1`.

A. Rahimi, Short-term prediction of NO2 and NO x concentrations using multi-layer perceptron neural network: a case study of Tabriz, Iran, Ecol. Process. 6 (2017) 4. URL: `http://ecologicalprocesses.springeropen.com/articles/10.1186/s13717-016-0069-x`. doi:`10.1186/s13717-016-0069-x`.

X. Mao, T. Shen, X. Feng, Prediction of hourly ground-level PM2.5concentrations 3 days in advance using neural networks with satellite data in eastern China, Atmos. Pollut. Res. 8 (2017) 1005–1015. URL: `https://doi.org/10.1016/j.apr.2017.04.002`. doi:`10.1016/j.apr.2017.04.002`.

C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, T. Chi, A novel spatiotemporal convolutional long short-term neural network for air pollution prediction, Sci. Total Environ. 654 (2019) 1091–1099. URL: `https://doi.org/10.1016/j.scitotenv.2018.11.086`. doi:`10.1016/j.scitotenv.2018.11.086`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1364815216309483`. doi:`10.1016/J.ENVSOFT.2016.11.017`.

D. Antanasijević, V. Pocajt, A. Perić-Grujić, M. Ristić, Multiple-input–multiple-output general regression neural networks model for the simultaneous estimation of traffic-

related air pollutant emissions, Atmos. Pollut. Res. 9 (2018) 388–397. doi:`10.1016/`
`j.apr.2017.10.011`.

M. S. Alam, A. McNabola, Exploring the modeling of spatiotemporal varia-
tions in ambient air pollution within the land use regression framework: Esti-
mation of PM10 concentrations on a daily basis, J. Air Waste Manage. As-
soc. 65 (2015) 628–640. URL: `http://www.tandfonline.com/doi/full/10.1080/`
`10962247.2015.1006377`. doi:`10.1080/10962247.2015.1006377`.

B. Yeganeh, M. G. Hewson, S. Clifford, A. Tavassoli, L. D. Knibbs, L. Morawska, Esti-
mating the spatiotemporal variation of NO2concentration using an adaptive neuro-
fuzzy inference system, Environ. Model. Softw. 100 (2018) 222–235. URL: `https:`
`//doi.org/10.1016/j.envsoft.2017.11.031`. doi:`10.1016/j.envsoft.2017.11.`
`031`.

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. URL: `https://www.sciencedirect.com/science/article/pii/`
`S1352231018302292`. doi:`10.1016/J.ATMOSENV.2018.04.004`.

H. Junninen, H. Niska, K. Tuppurainen, J. Ruuskanen, Methods for imputation of
missing values in air quality data sets, Atmos. Environ. 38 (2004) 2895–2907. doi:`10.`
`1016/j.atmosenv.2004.02.026`.

T. D. Pigott, A Review of Methods for Missing Data, Educ. Res. Eval. 7 (2001)
353–383.

A. Ã. Plaia, A. L. Bondì, Single imputation method of missing values in environmental
pollution data sets, Atmos. Environ. 40 (2006) 7316–7330. doi:`10.1016/j.atmosenv.`
`2006.06.040`.

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

Bibliography

S. Samarasinghe, Neural Networks for Applied Sciences and Engineering, Taylor & Francis, 2006. URL: `http://books.google.com.br/books?id=EyFeUiJibooC`. doi:`10.1017/CBO9781107415324.004`. arXiv:`arXiv:1011.1669v3`.

S. Makridakis, E. Spiliotis, V. Assimakopoulos, The M4 Competition: Results, findings, conclusion and way forward, Int. J. Forecast. 34 (2018) 802–808. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0169207018300785`. doi:`10.1016/j.ijforecast.2018.06.001`.

H. T. Shahraiyni, S. Sodoudi, A. Kerschbaumer, U. Cubasch, New technique for ranking of air pollution monitoring stations in the urban areas based upon spatial representativity (Case study: PM monitoring stations in Berlin), Aerosol Air Qual. Res. 15 (2015) 743–748. doi:`10.4209/aaqr.2014.12.0317`.

N. Sharma, K. K. Chaudhry, C. V. C. Rao, Vehicular pollution modeling using artificial neural network technique : A review, J. Sci. Ind. Res. (India). 64 (2005) 637–647.

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. URL: `http://linkinghub.elsevier.com/retrieve/pii/S0048969712014295`. doi:`10.1016/j.scitotenv.2012.10.110`.

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. URL: `http://dx.doi.org/10.1016/j.scitotenv.2012.03.076`. doi:`10.1016/j.scitotenv.2012.03.076`.

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`.

G. Ibarra-Berastegi, A. Elias, A. Barona, J. Saenz, A. Ezcurra, J. Diaz de Argandoña, From diagnosis to prognosis for forecasting air pollution using neu-

ral networks: Air pollution monitoring in Bilbao, Environ. Model. Softw. 23 (2008) 622–637. URL: `https://www.sciencedirect.com/science/article/pii/S1364815207001740`. doi:10.1016/J.ENVSOFT.2007.09.003.

J. Ordieres, E. Vergara, R. Capuz, R. Salazar, Neural network prediction model for fine particulate matter (PM2.5) on the US–Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua), Environ. Model. Softw. 20 (2005) 547–559. URL: `https://www.sciencedirect.com/science/article/pii/S1364815204000830`. doi:10.1016/J.ENVSOFT.2004.03.010.

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. URL: `https://www.sciencedirect.com/science/article/pii/S004896971401081X`. doi:10.1016/J.SCITOTENV.2014.07.051.

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.

U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, Two-days ahead prediction of daily maximum concentrations of SO2, O3, PM10, NO2, CO in the urban area of Palermo, Italy, Atmos. Environ. 41 (2007) 2967–2995. doi:10.1016/j.atmosenv.2006.12.013.

F. Biancofiore, M. Verdecchia, P. Di, B. Tomassetti, E. Aruffo, M. Busilacchio, S. Bianco, S. Di, C. Colangeli, Science of the Total Environment Analysis of surface ozone using a recurrent neural network, Sci. Total Environ. 514 (2015) 379–387. doi:10.1016/j.scitotenv.2015.01.106.

F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, S. Bianco, S. Di Tommaso, C. Colangeli, G. Rosatelli, P. Di Carlo, Recursive neural network model for analysis and forecast of PM10 and PM2.5, Atmos. Pollut. Res. 8 (2017) 652–659. doi:10.1016/j.apr.2016.12.014.

Bibliography

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. URL: `https://www.sciencedirect.com/science/article/pii/S1364815201000615`. doi:`10.1016/S1364-8152(01)00061-5`.

X. Li, L. Peng, X. Yao, S. Cui, Y. Hu, C. You, T. Chi, Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation, Environ. Pollut. 231 (2017) 997–1004. URL: `https://www.sciencedirect.com/science/article/pii/S0269749117307534`. doi:`10.1016/J.ENVPOL.2017.08.114`.

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. URL: `https://doi.org/10.1080/10962247.2018.1459956`. doi:`10.1080/10962247.2018.1459956`.

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`.

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

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. URL: `http://link.springer.com/article/10.1007/s10666-011-9270-6{%}5Cnhttp://link.springer.com/content/pdf/10.1007{%}2Fs10666-011-9270-6.pdf`. doi:`10.1007/s10666-011-9270-6`.

P. Jiang, C. Li, R. Li, H. Yang, An innovative hybrid air pollution early-warning system based on pollutants forecasting and Extenics evaluation, Knowledge-Based Syst. (2018). URL: `https://www.sciencedirect.com/science/article/pii/S0950705118305288`. doi:`10.1016/J.KNOSYS.2018.10.036`.

Bibliography

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. URL: `https://doi.org/10.1016/j.scitotenv.2018.01.195`. doi:`10.1016/j.scitotenv.2018.01.195`.

C. J. Huang, P. H. Kuo, A deep cnn-lstm model for particulate matter (Pm2.5) forecasting in smart cities, Sensors (Switzerland) 18 (2018). doi:`10.3390/s18072220`.

U. Pak, C. Kim, U. Ryu, K. Sok, S. Pak, A hybrid model based on convolutional neural networks and long short-term memory for ozone concentration prediction, Air Qual. Atmos. Heal. 11 (2018) 883–895. doi:`10.1007/s11869-018-0585-1`.

D. Qin, J. Yu, G. Zou, R. Yong, Q. Zhao, B. Zhang, A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM _{2.5} Concentration, IEEE Access 7 (2019) 20050–20059. URL: `https://ieeexplore.ieee.org/document/8632898/`. doi:`10.1109/ACCESS.2019.2897028`.

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). URL: `https://doi.org/10.1016/j.chemosphere.2018.12.128`. doi:`10.1016/j.chemosphere.2018.12.128`.

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`.

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

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

Bibliography

G. B. Humphrey, H. R. Maier, W. Wu, N. J. Mount, G. C. Dandy, R. J. Abrahart,
   C. W. Dawson, Improved validation framework and R-package for artificial neural
   network models, Environ. Model. Softw. 92 (2017) 82–106. URL: `https://www.`
   `sciencedirect.com/science/article/pii/S136481521630963X`. doi:`10.1016/J.`
   `ENVSOFT.2017.01.023`.

S. I. Gass, Feature Article-Decision-Aiding Models: Validation, Assessment,
   and Related Issues for Policy Analysis, Operations Research 31 (1983) 603–
   631. URL: `http://pubsonline.informs.org603-631.https//doi.org/10.1287/`
   `opre.31.4.603http://www.informs.org`. doi:`10.1287/opre.31.4.603`.

C. Dawson, R. Abrahart, L. See, HydroTest: A web-based toolbox of evaluation metrics
   for the standardised assessment of hydrological forecasts, Environ. Model. Softw.
   22 (2007) 1034–1052. URL: `https://www.sciencedirect.com/science/article/`
   `pii/S1364815206001642`. doi:`10.1016/J.ENVSOFT.2006.06.008`.

J. Kukkonen, L. Partanen, A. Karppinen, J. Ruuskanen, H. Junninen, M. Kolehmainen,
   H. Niska, S. Dorling, T. Chatterton, R. Foxall, G. Cawley, Extensive evaluation of
   neural network models for the prediction of NO2 and PM10 concentrations, compared
   with a deterministic modelling system and measurements in central Helsinki, Atmos.
   Environ. 37 (2003) 4539–4550. doi:`10.1016/S1352-2310(03)00583-1`.

A. Vlachogianni, P. Kassomenos, A. Karppinen, S. Karakitsios, J. Kukkonen, Science of
   the Total Environment Evaluation of a multiple regression model for the forecasting of
   the concentrations of NO x and PM 10 in Athens and Helsinki, Sci. Total Environ. 409
   (2011) 1559–1571. URL: `http://dx.doi.org/10.1016/j.scitotenv.2010.12.040`.
   doi:`10.1016/j.scitotenv.2010.12.040`.

A. K. Paschalidou, S. Karakitsios, S. Kleanthous, P. A. Kassomenos, Forecasting
   hourly PM10 concentration in Cyprus through artificial neural networks and multiple
   regression models: Implications to local environmental management, Environ. Sci.
   Pollut. Res. 18 (2011) 316–327. doi:`10.1007/s11356-010-0375-2`.

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. URL: http://www.atmospolres.com/articles/Volume6/issue3/APR-15-060.pdf. doi:10.5094/APR.2015.060. arXiv:1411.0701.

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.

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

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.

A. Z. Ul-Saufie, A. S. Yahaya, N. A. Ramli, N. Rosaida, H. A. Hamid, Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA), Atmos. Environ. 77 (2013) 621–630. URL: https://www.sciencedirect.com/science/article/pii/S1352231013003609. doi:10.1016/J.ATMOSENV.2013.05.017.

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. URL: http://dx.doi.org/10.1016/j.apr.2016.01.004. doi:10.1016/j.apr.2016.01.004.

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. URL: https://www.sciencedirect.com/science/article/pii/S1352231015001491. doi:10.1016/J.ATMOSENV.2015.02.030.

Bibliography

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). URL: `https://doi.org/10.1016/j.chemosphere.2019.01.121`. doi:`10.1016/j.chemosphere.2019.01.121`.

Y. Qi, Q. Li, H. Karimian, D. Liu, A hybrid model for spatiotemporal forecasting of PM 2.5 based on graph convolutional neural network and long short-term memory, Sci. Total Environ. 664 (2019) 1–10. URL: `https://doi.org/10.1016/j.scitotenv.2019.01.333`. doi:`10.1016/j.scitotenv.2019.01.333`.

J. Ma, Y. Ding, V. J. L. Gan, C. Lin, Z. Wan, Spatiotemporal Prediction of PM2.5 Concentrations at Different Time Granularities Using IDW-BLSTM, IEEE Access 7 (2019) 107897–107907. URL: `https://ieeexplore.ieee.org/document/8784234/`. doi:`10.1109/ACCESS.2019.2932445`.

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 classi fi cation, Atmos. Environ. 45 (2011) 1979–1985. URL: `http://dx.doi.org/10.1016/j.atmosenv.2011.01.022`. doi:`10.1016/j.atmosenv.2011.01.022`.

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`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1364815216310490`. doi:`10.1016/J.ENVSOFT.2017.11.031`.

A. Papaleonidas, L. Iliadis, Neurocomputing techniques to dynamically forecast spatiotemporal air pollution data, Evolving Systems (2013) 221–233. doi:`10.1007/s12530-013-9078-5`.

Bibliography

A. Kurt, A. B. Oktay, Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks, Expert Syst. Appl. 37 (2010) 7986–7992. doi:`10.1016/j.eswa.2010.05.093`.

X. Li, L. Peng, Y. Hu, J. Shao, T. Chi, Deep learning architecture for air quality predictions, Environ. Sci. Pollut. Res. 23 (2016) 22408–22417. doi:`10.1007/s11356-016-7812-9`.

W. Tong, L. Li, ·. X. Zhou, ·. A. Hamilton, K. Zhang, Deep learning PM 2.5 concentrations with bidirectional LSTM RNN, Air Qual. Atmos. Heal. 12 (2019) 411–423. URL: `https://doi.org/10.1007/s11869-018-0647-4`. doi:`10.1007/s11869-018-0647-4`.

T. Li, H. Shen, Q. Yuan, X. Zhang, L. Zhang, Estimating Ground-Level PM2.5 by Fusing Satellite and Station Observations: A Geo-Intelligent Deep Learning Approach, Geophys. Res. Lett. 44 (2017) 11,985–11,993. doi:`10.1002/2017GL075710`. `arXiv:1707.03558`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1364815213003198`. doi:`10.1016/J.ENVSOFT.2013.12.016`.

G. J. Bowden, H. R. Maier, G. C. Dandy, Real-time deployment of artificial neural network forecasting models: Understanding the range of applicability, Water Resour. Res. 48 (2012) 1–16. doi:`10.1029/2012WR011984`.

S. Galelli, G. B. Humphrey, H. R. Maier, A. Castelletti, G. C. Dandy, M. S. Gibbs, An evaluation framework for input variable selection algorithms for environmental data-driven models, Environ. Model. Softw. 62 (2014) 33–51. URL: `https://www.sciencedirect.com/science/article/pii/S1364815214002394`. doi:`10.1016/J.ENVSOFT.2014.08.015`.

Bibliography

W. Wu, R. J. May, H. R. Maier, G. C. Dandy, A benchmarking approach for comparing data splitting methods for modeling water resources parameters using artificial neural networks, Water Resour. Res. 49 (2013) 7598–7614. doi:`10.1002/2012WR012713`.

J. M. Hunter, H. R. Maier, M. S. Gibbs, E. R. Foale, N. A. Grosvenor, N. P. Harders, T. C. Kikuchi-Miller, Framework for developing hybrid process-driven, artificial neural network and regression models for salinity prediction in river systems, Hydrol. Earth Syst. Sci. 22 (2018) 2987–3006. URL: `https://www.hydrol-earth-syst-sci.net/22/2987/2018/`. doi:`10.5194/hess-22-2987-2018`.

G. B. Kingston, H. R. Maier, M. F. Lambert, Bayesian model selection applied to artificial neural networks used for water resources modeling, Water Resources Research 44 (2008). URL: `http://doi.wiley.com/10.1029/2007WR006155`. doi:`10.1029/2007WR006155`.

D. Wilks, Statistical Methods in the Atmospheric Sciences, 2nd ed., Academic Press, Burlington, MA, 2006.

W. Yang, K. Wang, W. Zuo, Neighborhood component feature selection for high-dimensional data, J. Comput. 7 (2012) 162–168. doi:`10.4304/jcp.7.1.161-168`.

L. Breiman, J. H. Friedman, R. A. Olshen, C. J. Stone, Classification and regression trees, Wadsworth International Group, Belmont, California, 1984.

E. C. Merkle, V. A. Shaffer, Binary recursive partitioning: Background, methods, and application to psychology, British Journal of Mathematical and Statistical Psychology 64 (2011) 161–181. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1348/000711010X503129`. doi:`10.1348/000711010X503129`. arXiv:`https://onlinelibrary.wiley.com/doi/pdf/10.1348/000711010X503129`.

K. S. Parmar, R. Bhardwaj, Wavelet and statistical analysis of river water quality parameters, Appl. Math. Comput. 219 (2013) 10172–10182. URL: `https://www.sciencedirect.com/science/article/pii/S0096300313003706`. doi:`10.1016/J.AMC.2013.03.109`.

Bibliography

I. Daubechies, Ten lectures on wavelets., SIAM Press, Philadelphia, USA., 1988.

Y. Nievergelt, Wavelets Made Easy, reprint of the 2001 edition ed., Birkhauser, NY, 2013.

S. Hochreiter, J. Schmidhuber, Long Short-Term Memory, Neural Comput. 9 (1997) 1735–1780. doi:`10.1162/neco.1997.9.8.1735`. `arXiv:1206.2944`.

J. Wang, G. Song, A Deep Spatial-Temporal Ensemble Model for Air Quality Prediction, Neurocomputing 314 (2018) 198–206. URL: `https://www.sciencedirect.com/science/article/pii/S0925231218307859`. doi:`10.1016/J.NEUCOM.2018.06.049`.

A. Charron, R. M. Harrison, Comparison between SMPS, Nano-SMPS and Ephiphaniometer data at an urban background site (Bloomsbury) and a roadside site (Marylebone Road), 2005. URL: `https://uk-air.defra.gov.uk/assets/documents/reports/cat05/0506061412{_}Comparison{_}between{_}SMPS-Nano-SMPS.pdf`.

C. J. Willmott, S. M. Robeson, K. Matsuura, A refined index of model performance, International Journal of Climatology 32 (2012) 2088–2094. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.2419`. doi:`10.1002/joc.2419`. `arXiv:https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.2419`.

C. J. Willmott, Some Comments on the Evaluation of Model Performance, Bulletin American Meteorological Society 63 (1982) 1309–1313.

USEPA, AERMOD Implementation Guide, 2016. URL: `https://www3.epa.gov/ttn/scram/models/aermod/aermod{_}implementation{_}guide.pdf`.

I. The MathWorks, MATLAB Documentation - MathWorks United Kingdom, 2019. URL: `https://uk.mathworks.com/help/matlab/index.html`.

Environmental Systems Research Institute, ArcGIS Desktop, 2020.

DEFRA, Marylebone Road, London - Non-Automatic Data- Defra, UK, 2020. URL: `https://uk-air.defra.gov.uk/data/maryleboneroad`.

Bibliography

The Guardian, London breaches annual air pollution limit for 2017 in just five days — Environment — The Guardian, 2017. URL: `https://www.theguardian.com/environment/2017/jan/06/london-breaches-toxic-air-pollution-limit-for-2017-in-just-five-days`.

WHO, Health Aspects of Air Pollution with Particulate Matter, Ozone and Nitrogen Dioxide Report on a WHO Working Group OZONE-adverse effects NITROGEN DIOXIDE-adverse effects AIR POLLUTANTS, ENVIRONMENTAL-adverse effects META-ANALYSIS AIR-standards GUIDELINES, Technical Report, World Health Organisation, 2003. URL: `http://www.euro.who.int/{_}{_}data/assets/pdf{_}file/0005/112199/E79097.pdf`.

M. A. Tanner, W. H. Wong, The calculation of posterior distributions by data augmentation, Journal of the American Statistical Association 82 (1987) 528–540. URL: `http://www.jstor.org/stable/2289457`.

S. Moorcroft, B. Marner, Review of the air quality monitoring network in london ref: Gla 80090, Air Quality Consultants Ltd: Greater London Authority (2011).

B. Beckerman, M. Jerrett, J. R. Brook, D. K. Verma, M. A. Arain, M. M. Finkelstein, Correlation of nitrogen dioxide with other traffic pollutants near a major expressway, Atmos. Environ. 42 (2008) 275–290. doi:`10.1016/j.atmosenv.2007.09.042`.

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`.

S. Robeson, D. Steyn, Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations, Atmospheric Environment. Part B. Urban Atmosphere 24 (1990) 303 – 312. URL: `http://www.sciencedirect.com/science/article/pii/095712729090036T`. doi:`https://doi.org/10.1016/0957-1272(90)90036-T`.

Bibliography

I. C. Ziomas, D. Melas, C. S. Zerefos, A. F. Bais, A. G. Paliatsos, Forecasting peak pollutant levels from meteorological variables, Atmospheric Environment 29 (1995) 3703 – 3711. URL: `http://www.sciencedirect.com/science/article/pii/135223109500131H`. doi:`https://doi.org/10.1016/1352-2310(95)00131-H`.

S. M. Cabaneros, J. K. Calautit, B. Hughes, Short- and long-term forecasting of ambient air pollution levels using wavelet-based non-linear autoregressive artificial neural networks with exogenous inputs, International Journal of Environmental Impacts: Management, Mitigation and Recovery 3 (2020) 143–154. doi:`10.2495/ei-v3-n2-143-154`.

A. Coman, A. Ionescu, Y. Candau, Hourly ozone prediction for a 24-h horizon using neural networks, Environ. Model. Softw. 23 (2008) 1407–1421. URL: `https://www.sciencedirect.com/science/article/pii/S1364815208000650`. doi:`10.1016/J.ENVSOFT.2008.04.004`.

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. URL: `https://linkinghub.elsevier.com/retrieve/pii/S0304380020300892`. doi:`10.1016/j.ecolmodel.2020.109017`.

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`.

P. Perez, A. Trier, Prediction of NO and NO 2 concentrations near a street with heavy traf c in Santiago, Chile. Atmos. Environ. 35 (2001) 1783–1789.

S. Abdul-Wahab, S. Al-Alawi, Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks, Environ. Model. Softw. 17 (2002) 219–228. URL: `https://www.sciencedirect.com/science/article/pii/S1364815201000779`. doi:`10.1016/S1364-8152(01)00077-9`.

W. Z. Lu, W. J. Wang, X. K. Wang, Z. B. Xu, A. Y. T Leung, Using Improved Neural

Bibliography

Network Model To Analyze Rsp, NOx and NO2 Levels in Urban Air in Mong Kok, Hong Kong, Environ. Monit. Assess. 87 (2003) 235–254.

W. Wang, Z. Xu, J. Weizhen Lu, Three improved neural network models for air quality forecasting, Eng. Comput. 20 (2003) 192–210. URL: `http://www.emeraldinsight.com/doi/10.1108/02644400310465317`. doi:`10.1108/02644400310465317`.

F. A. Hasham, W. B. Kindzierski, S. J. Stanley, Modeling of hourly NO x concentrations using artificial neural networks 1, J. Environ. Eng. Sci. 3 (2004) 111–119. doi:`10.1139/S03-084`.

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

D. Jiang, Y. Zhang, X. Hu, Y. Zeng, J. Tan, D. Shao, Progress in developing an ANN model for air pollution index forecast, Atmos. Environ. 38 (2004) 7055–7064. doi:`10.1016/j.atmosenv.2003.10.066`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S0952197604000119`. doi:`10.1016/J.ENGAPPAI.2004.02.002`.

G. Nunnari, Modelling air pollution time-series by using wavelet functions and genetic algorithms, Soft Comput. 8 (2004) 173–178. doi:`10.1007/s00500-002-0260-0`.

L. E. Olcese, B. M. Toselli, A method to estimate emission rates from industrial stacks based on neural networks, Chemosphere 57 (2004) 691–696. doi:`10.1016/j.chemosphere.2004.07.045`.

A. B. Chelani, R. N. Singh, S. Devotta, Nonlinear dynamical characterization and prediction of ambient nitrogen dioxide concentration, Water. Air. Soil Pollut. 166 (2005) 121–138. doi:`10.1007/s11270-005-7384-7`.

J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens, O. Brasseur, A neural network

forecast for daily average PM10concentrations in Belgium, Atmos. Environ. 39 (2005) 3279–3289. doi:`10.1016/j.atmosenv.2005.01.050`.

H. Niska, M. Rantam??ki, 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`.

G. Grivas, A. Ã. Chaloulakou, Artificial neural network models for prediction of PM 10 hourly concentrations , in the Greater Area of Athens , Greece, Atmos. Environ. 40 (2006) 1216–1229. doi:`10.1016/j.atmosenv.2005.10.036`.

S. S. Nagendra, M. Khare, Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions, Ecol. Modell. 190 (2006) 99–115. URL: `http://linkinghub.elsevier.com/retrieve/pii/S0304380005002280`. doi:`10.1016/j.ecolmodel.2005.01.062`.

U. Schlink, O. Herbarth, M. Richter, S. Dorling, G. Nunnari, G. Cawley, E. Pelikan, Statistical models to assess the health effects and to forecast ground-level ozone, Environ. Model. Softw. 21 (2006) 547–558. URL: `https://www.sciencedirect.com/science/article/pii/S1364815204003196`. doi:`10.1016/J.ENVSOFT.2004.12.002`.

T. Slini, A. Kaprara, K. Karatzas, N. Moussiopoulos, PM10 forecasting for Thessaloniki, Greece, Environ. Model. Softw. 21 (2006) 559–565. URL: `https://www.sciencedirect.com/science/article/pii/S1364815204003202`. doi:`10.1016/J.ENVSOFT.2004.06.011`.

S. Sousa, F. Martins, M. Alvim-Ferraz, M. Pereira, Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, Environ. Model. Softw. 22 (2007) 97–103. URL: `https://www.sciencedirect.com/science/article/pii/S1364815205002240`. doi:`10.1016/J.ENVSOFT.2005.12.002`.

Bibliography

S. M. Al-Alawi, S. A. Abdul-Wahab, C. S. Bakheit, Combining principal component regression and artificial neural networks for more accurate predictions of ground-level ozone, Environ. Model. Softw. 23 (2008) 396–403. doi:`10.1016/j.envsoft.2006.08.007`.

M. L. Martín, I. J. Turias, F. J. González, P. L. Galindo, F. J. Trujillo, C. G. Puntonet, J. M. Gorriz, Prediction of CO maximum ground level concentrations in the Bay of Algeciras, Spain using artificial neural networks, Chemosphere 70 (2008) 1190–1195. doi:`10.1016/j.chemosphere.2007.08.039`.

P. Perez, G. Salini, PM2.5forecasting in a large city: Comparison of three methods, Atmos. Environ. 42 (2008) 8219–8224. URL: `http://dx.doi.org/10.1016/j.atmosenv.2008.07.035`. 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`.

R. S. Ettouney, S. A. Wahab, A. S. Elkilani, Emissions inventory , ISCST , and neural network modelling of air pollution in Kuwait Emissions inventory , ISCST , and neural network modelling of air pollution in Kuwait, Int. J. Environ. Stud. 7233 (2009). doi:`10.1080/00207230902859929`.

F. Galatioto, P. Zito, Traffic Parameters Estimation to Predict Road Side Pollutant Concentrations using Neural Networks, Environ. Model. Assess. (2009) 365–374. doi:`10.1007/s10666-007-9129-z`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1352231009006669?via{%}3Dihub`. doi:`10.1016/J.ATMOSENV.2009.07.048`.

E. Pisoni, M. Farina, C. Carnevale, L. Piroddi, Forecasting peak air pollution levels using NARX models, Eng. Appl. Artif. Intell. 22 (2009) 593–602. URL: `http://`

212

Bibliography

dx.doi.org/10.1016/j.engappai.2009.04.002. doi:10.1016/j.engappai.2009.
04.002.

C.-h. Tsai, L.-c. Chang, H.-c. Chiang, Forecasting of ozone episode days by cost-sensitive neural network methods, Sci. Total Environ. 407 (2009) 2124–2135. URL: http://dx.doi.org/10.1016/j.scitotenv.2008.12.007. doi:10.1016/j. scitotenv.2008.12.007.

G. Demir, K. Ozcan, O. N. Ucan, C. Bayat, An Artificial Neural Network (ANN) based model for short-term predictions of daily mean PM10 concentrations . An Artificial Neural Network-based model, J. Environ. Prot. Ecol. 11(3) (2010) 1163–1171.

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.

A. Mahapatra, Prediction of daily ground-level ozone concentration, Environ. Monit. Assess. 170 (2010) 159–170. doi:10.1007/s10661-009-1223-z.

K. P. Moustris, I. C. Ziomas, A. G. Paliatsos, 3-day-ahead forecasting of regional pollution index for the pollutants NO2, CO, SO2, and O3 using artificial neural networks in athens, Greece, Water. Air. Soil Pollut. 209 (2010) 29–43. doi:10.1007/ s11270-009-0179-5.

J. C. M. Pires, M. C. M. A. Ferraz, M. C. Pereira, F. G. Martins, F. G. Martins, Atmospheric Pollution Research Evolutionary procedure based model to predict ground – level ozone concentrations, Atmos. Pollut. Res. 1 (2010) 215–219. URL: http://dx.doi.org/10.5094/APR.2010.028. doi:10.5094/APR.2010.028.

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 PM 10 and PM 2 . 5 concentrations using arti fi cial neural networks , in Thessaloniki and Helsinki, Sci. Total Environ. 409 (2011) 1266–1276. URL: http://dx.doi.org/10.1016/j.scitotenv.2010.12.039. doi:10.1016/j.scitotenv.2010.12.039.

Bibliography

J. M. Barrón-adame, M. G. Cortina-januchs, A. Vega-corona, D. Andina, Expert Systems with Applications Unsupervised system to classify SO 2 pollutant concentrations in Salamanca , Mexico, Expert Syst. Appl. 39 (2012) 107–116. doi:10.1016/j.eswa.2011.05.083.

S. Chattopadhyay, G. Chattopadhyay, Modeling and Prediction of Monthly Total Ozone Concentrations by Use of an Artificial Neural Network Based on Principal Component Analysis, Pure Appl. Geophys. 169 (2012) 1891–1908. doi:10.1007/s00024-011-0437-5.

P. Perez, Combined model for PM10 forecasting in a large city, Atmos. Environ. 60 (2012) 271–276. URL: http://dx.doi.org/10.1016/j.atmosenv.2012.06.024. doi:10.1016/j.atmosenv.2012.06.024.

G. D. Gennaro, L. Trizio, A. Di, J. Pey, N. Pérez, M. Cusack, A. Alastuey, X. Querol, Science of the Total Environment Neural network model for the prediction of PM 10 daily concentrations in two sites in the Western Mediterranean, Sci. Total Environ. 463-464 (2013) 875–883. URL: http://dx.doi.org/10.1016/j.scitotenv.2013.06.093. doi:10.1016/j.scitotenv.2013.06.093.

K. P. Moustris, I. K. Larissi, P. T. Nastos, K. V. Koukouletsos, A. G. Paliatsos, Development and Application of Artificial Neural Network Modeling in Forecasting PM 10 Levels in a Mediterranean City (2013). doi:10.1007/s11270-013-1634-x.

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

K. Yan, L. Jian, Neurocomputing Identification of significant factors for air pollution levels using a neural network based knowledge discovery system, Neurocomputing 99 (2013) 564–569. URL: http://dx.doi.org/10.1016/j.neucom.2012.06.003. doi:10.1016/j.neucom.2012.06.003.

Bibliography

H. Zhang, Y. Liu, R. Shi, Q. Yao, Evaluation of PM ₁₀ forecasting based on the artificial neural network model and intake fraction in an urban area: A case study in Taiyuan City, China, J. Air Waste Manage. Assoc. 63 (2013) 755–763. URL: `http://www.tandfonline.com/doi/abs/10.1080/10962247.2012.755940`. doi:`10.1080/10962247.2012.755940`.

A. Azid, H. Juahir, M. E. Toriman, Prediction of the Level of Air Pollution Using Principal Component Analysis and Artificial Neural Network Techniques : a Case Study in Malaysia (2014). doi:`10.1007/s11270-014-2063-1`.

H.-d. He, W.-Z. 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. URL: `https://www.sciencedirect.com/science/article/pii/S0360132314001073`. doi:`10.1016/J.BUILDENV.2014.04.011`.

A. S. Luna, Prediction of ozone concentration in tropospheric levels using arti fi cial neural networks and support vector machine at Rio de, Atmos. Environ. 98 (2014) 98–104. doi:`10.1016/j.atmosenv.2014.08.060`.

U. Özdemir, S. Taner, Impacts of Meteorological Factors on PM 10 : Artificial Neural Networks (ANN) and Multiple Linear Regression (MLR) Approaches Contributed Articles Impacts of Meteorological Factors on PM 10 : Artificial Neural Networks (ANN) and Multiple Linear Regr, Environ. Forensics 5922 (2014). doi:`10.1080/15275922.2014.950774`.

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`.

X. Zhang, X. Rui, X. Xia, X. Bai, W. Yin, J. Dong, A hybrid model for short-term air pollutant concentration forecasting, 10th IEEE Int. Conf. Serv. Oper. Logist. Informatics, SOLI 2015 - conjunction with ICT4ALL 2015 (2015) 171–175. doi:`10.1109/SOLI.2015.7367614`.

Bibliography

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. URL: `http://dx.doi.org/10.5094/APR.2015.071`. doi:`10.5094/APR.2015.071`.

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. URL: `https://www.sciencedirect.com/science/article/pii/S1352231014009157`. doi:`10.1016/J.ATMOSENV.2014.11.050`.

G. Santos, I. Fernández-olmo, Estimation of PM 10 -Bound As , Cd , Ni and Pb Levels by Means of Statistical Modelling : PLSR and ANN Approaches, Water, Air, Soil Pollut. (2015). doi:`10.1007/s11270-015-2526-z`.

G. Zhu, P. Zhang, T. Tshukudu, J. Yin, G. Fan, X. Zheng, Forecasting traf fi c-related nitrogen oxides within a street canyon by combining a genetic algorithm-back propagation arti fi cial neural network and parametric models, Atmos. Pollut. Res. 6 (2015) 1087–1097. URL: `http://dx.doi.org/10.1016/j.apr.2015.06.006`. doi:`10.1016/j.apr.2015.06.006`.

B. Zou, M. Wang, N. Wan, J. G. Wilson, X. Fang, Y. Tang, Spatial modeling of PM 2 . 5 concentrations with a multifactoral radial basis function neural network (2015) 10395–10404. doi:`10.1007/s11356-015-4380-3`.

M. R. Chellali, H. Abderrahim, A. Hamou, A. Nebatti, J. Janovec, Artificial neural network models for prediction of daily fine particulate matter concentrations in Algiers, Environ. Sci. Pollut. Res. (2016) 14008–14017. URL: `http://dx.doi.org/10.1007/s11356-016-6565-9`. doi:`10.1007/s11356-016-6565-9`.

W. Ding, J. Zhang, Y. Leung, Prediction of air pollutant concentration based on sparse response back-propagation training feedforward neural networks, Environ. Sci. Pollut. Res. (2016) 19481–19494. URL: `http://dx.doi.org/10.1007/s11356-016-7149-4`. doi:`10.1007/s11356-016-7149-4`.

Bibliography

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.

J. He, Y. Yu, Y. Xie, H. Mao, L. Wu, N. Liu, S. Zhao, Numerical Model-Based Artificial Neural Network Model and Its Application for Quantifying Impact Factors of Urban Air Quality, Water, Air, Soil Pollut. 227 (2016) 235. URL: http://link.springer.com/10.1007/s11270-016-2930-z. doi:10.1007/s11270-016-2930-z.

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.

S. D. Lightstone, F. Moshary, B. Gross, Comparing CMAQ forecasts with a neural network forecast model for PM2.5 in New York, Atmosphere (Basel). 8 (2017). doi:10.3390/atmos8090161.

J. Zhang, W. Ding, Prediction of air pollutants concentration based on an extreme learning machine: The case of Hong Kong, Int. J. Environ. Res. Public Health 14 (2017) 1–19. doi:10.3390/ijerph14020114.

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. URL: https://doi.org/10.1016/j.apr.2017.11.004. 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. URL: http://dx.doi.org/10.1016/j.apr.2018.02.006. 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

Bibliography

(2018) 129–139. URL: `https://www.sciencedirect.com/science/article/pii/S1352231018301651`. doi:`10.1016/J.ATMOSENV.2018.03.027`.

J. Nidzgorska-Lencewicz, Application of artificial neural networks in the prediction of PM10levels in thewinter months: A case study in the Tricity Agglomeration, Poland, Atmosphere (Basel). 9 (2018). doi:`10.3390/atmos9060203`.

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). URL: `https://www.sciencedirect.com/science/article/pii/S1309104218304975`. doi:`10.1016/J.APR.2018.11.004`.

L. M. B. Ventura, &. Fellipe De Oliveira Pinto, &. Laiza, M. Soares, A. S. Luna, A. Gioda, Forecast of daily PM 2.5 concentrations applying artificial neural networks and Holt-Winters models, Air Qual. Atmos. Heal. (2019). URL: `https://doi.org/10.1007/s11869-018-00660-x`. doi:`10.1007/s11869-018-00660-x`.

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 Keywords: Air pollutant concentrations forecasting Empirical wavelet transform Multi-agent evolutionary genetic algorithm Nonlinear auto regressive models w, Sustain. cities Soc. (2019). URL: `https://doi.org/10.1016/j.scs.2019.101471`. doi:`10.1016/j.scs.2019.101471`.

Z. Qi, T. Wang, G. Song, W. Hu, X. Li, Zhongfei, Zhang, Deep Air Learning: Interpolation, Prediction, and Feature Analysis of Fine-grained Air Quality (2017). URL: `http://arxiv.org/abs/1711.00939`. arXiv:`1711.00939`.