

**AIR POLLUTION MODELLING AND FORECASTING USING
DATA-DRIVEN METHODS**

AIR POLLUTION MODELLING AND FORECASTING IN HAMILTON USING DATA-DRIVEN METHODS

By

TARANA A. SOLAIMAN

Bachelor of Science in Civil Engineering (Bangladesh Univ. of Engg. & Tech.)

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements

for the Degree

Master of Applied Science

McMaster University

Department of Civil Engineering

MASTER OF APPLIED SCIENCE (2007)
(Civil Engineering)

McMaster University
Hamilton, Ontario

TITLE: Air Pollution Modelling and Forecasting in Hamilton Using
Data-Driven Methods

AUTHOR: Tarana A. Solaiman
B.Sc. Engg. (Civil), (Bangladesh Univ. of Engg & Tech.)

SUPERVISORS: Dr. Paulin Coulibaly, Dr. Pavlos Kanaroglou

NUMBER OF PAGES: xv, 162

Abstract

The purpose of this research is to provide an extensive evaluation of neural network models for the prediction and the simulation of some key air pollutants in Hamilton, Ontario, Canada. Hamilton experiences one of Canada's highest air pollution exposures because of the dual problem associated with continuing industrial emission and gradual increase of traffic related emissions along with the transboundary air pollutions from heavily industrialized neighboring north-eastern and mid-western US cities. These factors combined with meteorology, cause large degradation of Hamilton's air quality. Hence an appropriate and robust method is of most importance in order to get an early notification of the future air quality situation. Data driven methods such as neural networks (NNs) are becoming very popular due to their inherent capability to capture the complex non-linear relationships between pollutants, climatic and other non-climatic variables such as traffic variables, emission factors, etc. This study investigates dynamic neural networks, namely time lagged feed-forward neural network (TLFN), Bayesian neural network (BNN) and recurrent neural network (RNN) for short term forecasting. The results are being compared with the benchmark static multilayer perceptron (MLP) models. The analysis shows that TLFN model with its time delay memory and RNN with its adaptive memory has outperformed the static MLP models in ground level ozone (O_3) forecasting for up to 12 hours ahead. Furthermore the model developed using the annual database is able to map the variations in the seasonal concentrations. On the other hand, MLP model was quite competitive for nitrogen dioxide (NO_2) prediction when compared to the dynamic NN based models.

The study further assesses the ability of the neural network models to generate pollutant concentrations at sites where sampling has not been done. Using these neural network models, data values were generated for total suspended particulate (TSP) and inhalable particulates (PM_{10}) concentrations. The obtained results show promising potential. Although there were under-predictions and over-predictions on some occasions, the neural network models, in general were able to generate the missing information and to obtain air quality situation in the study area.

Acknowledgements

First of all, I would like to gratefully acknowledge my supervisors Dr. P. Coulibaly and Dr. P. Kanaroglou, for their guidance and indispensable support in completing this research. I greatly admire them for their accessibility and patience, their professionalism and their scientific insight.

I also acknowledge Mr. Frank Dobroff, Air Quality Analyst, Ontario Ministry of Environment and Mr. Jim Cook, Ontario Climate Centre for providing the database used in this work.

The departmental staff has also been tremendous; my sincere thanks to Carol Robinson, Rebecca Woodworth and Tatiana Dobrovolska of Department of Civil Engineering and Pat DeLuca and Deane Maynard of School of Geography and Earth Sciences for answering all my questions and resolving issues of concern.

Thanks to my colleagues in the Water Resources and Hydrology Research Group: Ashok, Getnet, Sherry, Mike, Manu and my friends Shazia, Asif, Farhana and Samia for helping me time to time which made my life easier and more enjoyable.

Finally my grateful acknowledgement goes to my family members, my always inspiring mom (Professor Momena Akhtar Khodeja) and dad (Dr. Muhammad Solaiman), my sweet sisters (Dr. Tamanna Solaiman and Dr. Mahjabeen Solaiman), my sister-in-law (Dr. Javed Iqbal) and finally my loving husband, Ali Jayed for supporting through my endeavors. Without your love, support and understanding it would not have been possible for me to proceed with my studies.

Table of Contents

Abstract	iii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	x
List of Symbols	xii
Chapter 1: Introduction	1
1.1 Statement of the problem.....	1
1.2 Background.....	1
1.3 Research objectives.....	3
1.4 Structure of the thesis.....	5
Chapter 2: Study Area and Database	7
2.1 Study area: Hamilton census metropolitan area.....	7
2.1.1 Description of the study area.....	7
2.1.2 Meteorology of the study area.....	10
2.1.3 Air quality in Hamilton.....	10
2.2 Database.....	12
2.2.1 Air pollution database.....	12
2.2.2 Meteorological database.....	16
2.2.3 Land use database.....	19
2.2.4 Data filling technique.....	20
Chapter 3: Review of Literature	23
3.1 Basic air quality modeling approaches.....	23
3.1.1 Deterministic methods vs. Statistical methods.....	23
3.2 Neural network models.....	28
3.2.1 Mathematical aspects.....	29
3.2.2 Important aspects of NN modeling.....	29

3.2.2.1 Selection of input and output variables.....	30
3.2.2.2 Neural network architecture.....	31
3.2.2.3 Network training, cross-validation and model testing.....	32
3.2.2.4 Advantages and limitations.....	33
3.3 Application of NNs in air quality modeling.....	35
Chapter 4: Methodology.....	41
4.1 Temporal analysis of air quality.....	41
4.1.1 Multilayer perceptron	41
4.1.2 Time lagged feed-forward network	42
4.1.3 Recurrent neural network.....	44
4.1.4 Bayesian neural network.....	47
4.2 Spatial interpolation by kriging.....	51
Chapter 5: Network Design.....	53
5.1 Selection of best predictors.....	53
5.1.1 Correlation analysis.....	53
5.1.2 Sensitivity analysis.....	55
5.2 Model setup.....	61
5.3 Model performance evaluation.....	70
Chapter 6: Ozone and Nitrogen Oxide Forecasting Results.....	73
6.1 Ozone forecasting results.....	73
6.1.1 Model forecasting performances.....	73
6.1.2 Confidence interval with BNN.....	81
6.1.3 Seasonal variation.....	86
6.1.4 Annual vs. summer model.....	89
6.2 Ozone forecasting with land use variables.....	91
6.3 Nitrogen dioxide forecasting results.....	92
6.3.1 Model forecasting performances.....	92
6.3.2 Seasonal variation.....	93
6.4 Summary of ozone and NO ₂ forecasting results.....	98

Chapter 7: Simulation Results for Total Suspended Particulates and PM₁₀ Pollutants	100
7.1 Simulation of TSP.....	101
7.1.1 Comparison of different input variables.....	101
7.1.2 Model performance statistics.....	101
7.1.3 Seasonal variation.....	105
7.1.4 Confidence interval with BNN.....	108
7.1.5 Spatial interpolation.....	110
7.2 Simulation of PM ₁₀	113
7.2.1 Combination of input variables.....	113
7.2.2 .Model performance statistics.....	113
7.2.3 Seasonal variation.....	117
7.2.4 Confidence interval with BNN.....	118
7.3 Summary of TSP and PM ₁₀ results.....	120
Chapter 8: Conclusions and Recommendations	121
8.1 Conclusions.....	121
8.2 Recommendations for future work.....	125
References	126
Appendices	136

List of Tables

Table 2.1	Air pollutant stations in Hamilton
Table 2.2	Pollutant statistics (1994-2004)
Table 2.3	Meteorological variables used
Table 2.4	Land use within 200 m buffer of all stations
Table 3.1	Summary of statistical methods used in air pollution modeling (Schlink et al., 2001)
Table 5.1	Selected input variables for ozone based on sensitivity analysis
Table 5.2	Selected input variables for NO ₂ based on sensitivity analysis
Table 5.3	Selected input variables for TSP based on sensitivity analysis
Table 5.4	Selected input variables for PM ₁₀ based on sensitivity analysis
Table 5.5	Best model structure for ozone
Table 5.6	Best model structure for NO ₂
Table 5.7	Best model structure for TSP
Table 5.8	Best model structure for PM ₁₀
Table 6.1	Comparison of model performance at station 29000, 29114 and 29118
Table 6.2	Seasonal variation of model performance: Hamilton Downtown (29000)
Table 6.3	Annual model vs. summer model: Hamilton Downtown (29000)
Table 6.4	Comparison of model performance using land use variables: Hamilton Downtown (29000)
Table 6.5	Comparison of model performance of NO ₂ at station 29000 and 29118
Table 7.1	Combination of input variables for TSP: (a) TLFN, MLP, RNN models (b) BNN model performance
Table 7.2	Model performance statistics for TSP
Table 7.3	Comparison of seasonal bias at station 29000, 29025, 29102, 29113 and 29114
Table 7.4	Combination of input variables for PM ₁₀ : TLFN model performance
Table 7.5	Comparison of model performance for PM ₁₀
Table 7.6	Comparison of seasonal bias at station 300, 302 and 313

Table A.1	Sources and impact of air pollutants
Table A.2	Transfer functions in neural network model
Table A.3	Ground level ozone forecasting statistics: Hamilton Downtown (29000), Hamilton Mountain (29114) and Hamilton west (29118)
Table A.4 (a)	Seasonal variation of model performance for O ₃ : Hamilton Mountain (29114)
Table A.4 (b)	Seasonal variation of model performance for O ₃ : Hamilton West (29118)
Table A.5	Comparison of annual and summer models for O ₃ : Hamilton Downtown (29000)
Table A.6	Model forecasting performance with land use variable for O ₃ : Hamilton Downtown (29000)
Table A.7 (a)	Model forecasting performance for NO ₂ : Hamilton Downtown (29000)
Table A.7 (b)	Model forecasting performance for NO ₂ : Hamilton West (29118)
Table A.8	Seasonal variation of model performance: Hamilton Downtown (29000)

List of Figures

- Fig. 2.1** Location of study area in Ontario
- Fig. 2.2** Highways and industrial area around Hamilton (Source: Google map, 2007 / Imaginary Digital Globe, Earth Sat)
- Fig. 2.3 (a)** 10 year trend of ozone at some selected Ontario sites (*Data source, OME, 2004*)
- Fig. 2.3 (b)** 10 year trend of NO₂ at some selected Ontario sites (*Data source: MOE, 2004*)
- Fig. 2.4** Ontario MOE pollutant monitoring stations in Hamilton
- Fig. 3.1** Schematic diagram of biological neuron (*Hagan et al, 1996*)
- Fig. 3.2** Schematic diagram of a node (*Haykin, 1999*)
- Fig. 3.3** Configuration of a typical feed-forward neural network
- Fig. 4.1** Multilayer perceptron neural network architecture
- Fig. 4.2** Time lagged feed-forward neural network with time delay memory structure (*Coulibaly et al., 2001b*)
- Fig. 4.3** Fully recurrent neural network with feedback connection
- Fig. 4.4** Distribution of network outputs with error bars (*Bishop, 1995*): (a) width of the network outputs dominated by the noise of the data and (b) width of the network outputs dominated by the distribution of network weights
- Fig. 6.1 (a)** Model forecasting statistics for ozone: Hamilton Downtown (top) and Hamilton Mountain (bottom)
- Fig. 6.1(b)** Model forecasting statistics for ozone: Hamilton West
- Fig. 6.2 (a)** Scatter plots at 1 and 4 step ahead: Hamilton Downtown (29000)
- Fig. 6.2 (b)** Scatter plots at 6 step ahead: Hamilton Downtown (29000)
- Fig. 6.3** Comparative results of 2 (left) and 4 (right) hour ahead O₃ forecasting with 95% confidence interval: Hamilton Downtown (29000)

- Fig. 6.4** Comparative results of 2 (left) and 4 (right) hour ahead ozone forecasting with 95% confidence interval: Hamilton Mountain (29114)
- Fig. 6.5** Model forecasting statistics for NO₂: Hamilton Downtown (29000) (top) and Hamilton Mountain (29114) (bottom)
- Fig. 6.6 (a)** Scatter Plots of 2 hour ahead forecasting: NO₂ at site 29000
- Fig. 6.6 (b)** Scatter Plots of 4 hour ahead forecasting: NO₂ at site 29000
- Fig. 7.1** Scatter plots of observed vs. simulated values of TSP concentration at station 29000 (left) and 29025 (right)
- Fig. 7.2** Simulation result of TSP at station 29025 with 95% confidence interval
- Fig. 7.3** Comparison of observed and simulated values over space during spring 2003
- Fig. 7.4** Scatter plots for PM₁₀ concentration at station 300 (left) and 313 (right)
- Fig. 7.5** Simulation result of PM₁₀ at station 300 with 95% confidence interval
- Fig. A.1 (a)** Scatter plots of observed vs. simulated values of TSP values at station 29113 (left) and 29114 (right)
- Fig. A.1 (b)** Scatter plots of observed vs. simulated values of TSP values at station 29113 (left) and 29114 (right)
- Fig. A.1 (c)** Scatter plots of observed vs. simulated values of TSP concentration at station 29102
- Fig A.2 (a)** Simulation result of TSP at station 29000 with 95% confidence interval
- Fig. A.2 (b)** Simulation result of TSP at station 29113 with 95% confidence interval
- Fig. A.2 (c)** Simulation result of TSP at station 29114 with 95% confidence interval
- Fig. A.3 (a)** Comparison of observed and simulated values over space during fall 2003
- Fig. A.3 (b)** Comparison of observed and simulated values over space during fall 2003
- Fig. A.4** Scatter plots for PM₁₀ concentration at station 302
- Fig. A.5** Simulation result of PM₁₀ at station 302 with 95% confidence interval

List of Symbols

A_h	The Hessian matrix of the regularized error function
A	The weight matrix of the output layer neurons connected to the hidden neurons
b_j and b_o	Additional network parameters (biases) to be determined during training of the networks with observed input/output data sets
BNN	Bayesian neural network
C_o	Observed value
C_p	Predicted value
CMA	Census metropolitan area
CV	Cross-validation
DBD	Delta bar delta algorithm
E	Error
ED	Error function
E_w	Weight decay regularizer
$\nabla E(w)$	The gradient of error E with respect to the weights
$G()$	A logistic function characterizing the hidden nodes
g	The gradient of $y(x,w)$ with respect to the weights w evaluated at w_{MP}
k	Memory depth
m	The size of the hidden layer

MAE	Mean absolute error
MLP	Multilayer perceptron
N	The number of the measured value
n	The time step
$NMSE$	Normalized mean square error
NO_2	Nitrogen dioxide
O_3	Ground level ozone
$p(w D)$	The posterior probability distribution for the weights
$p(D)$	A normalizing factor
$p(D w)$	The dataset likelihood function and the denominator
$p(i x, w)$	The model for the distribution of noise on the target data for a fixed value of the weight vector w_{MP}
p	The size of sample
PEs	Processing elements
PM_{10}	Particulate matter with less than 10 micrometers in diameter
r	Correlation coefficient
R^2	Coefficient of determination
RB	Relative bias
$RMSE$	Root mean square error
RNN	Recurrent neural network
s_o	The prediction location

t	The target value
$TDNN$	Time delay neural network
$TLFN$	Time lagged feed-forward neural network
TSP	Total suspended particulates
W_{h_o}	The weight matrix of the hidden units connected to the output nodes
W	The total number of weights and biases in the network
W_h	The weights of the h hidden nodes that are connected to the context units
w_j	Weight vector for the connection between the hidden and output layers
w_{jl}	Weight matrix for the connection between the input and hidden layers
w_{MP}	The weight vector
x	The input vector
$X(n)$	Combined input to the processing elements at time step n
$x(n)$	The input pattern at time step n
$x'(t)$	The output of the hidden layer at time t given an input vector $x(t)$
$x_j(n)$	An individual input at the n^{th} time step
$y(x; w)$	The network output
y_j	The output of the RNN assuming a linear output node j ,

y_t	Observation at any time t
\bar{y}	The mean observation
$Z(s_i)$	The measured value at the i^{th} location
β	Noise hyperparameter
φ_1 and φ_2	Transfer functions at the output and hidden layers, respectively
$\varphi_j(k)$	The j^{th} autoregressive coefficient in an autoregressive mode
$\varphi_k(k)$	The partial autocorrelation function
λ_i	An unknown weight for the measured value at the i^{th} location
σ_t	The standard deviation

Chapter 1

Introduction

1.1 Statement of the problem

Atmospheric pollution has been a concern of the past and present societies. The gradual increase of industrialization and urbanization has a significant impact on present environment (Kuylenstierna et al, 2002). During 1980s air quality problems became more prevalent due to increase of motor vehicle emissions. In recent times, vehicular and industrial emissions have been recognized as two major sources of urban air pollution. Despite various regulations and acts for controlling industrial emissions, number of vehicles on roads continue to increase, thus traffic related emission has become the most significant source of present day urban air pollution (Blair, 2006). According to World Health Organization (WHO), vehicular and industrial emissions cause death of 3 million people world wide each year. Increase in green house gas emissions and temperature (climate warming) would further degrade global air quality and global warming as predicted by the IPCC fourth assessment report (IPCC, 2007).

1.2 Background

Human activities have always introduced many types of contaminants into the environment. Pollutants emitted from industry and power plants, exhaust emissions from transportation vehicles, radio nuclides from nuclear weapon tests and uranium mill tailings and pesticides, emitted into the atmosphere have been affecting the quality of air and has serious consequences for human health. Furthermore, the increasing demand of

energy and increase in world's population has driven the pollution level upward. The short and long-term effects of air pollution are varied and profound. In recent years the world has experienced acid rain, smog and depletion of the ozone layer as an outcome of air pollutants. A significant human health risk, causing serious respiratory and other illnesses is very common in areas with 'poor' air quality. Air pollution is thus, considered as a major concern and requires urgent attention.

The complexity of the nonlinear nature of pollutant formation is aggravated by the wide range of temporal and spatial variations of the meteorology and the chemical processes involved. Faster chemical reactions directly affect the local air quality which strongly depends on the local atmospheric conditions. Slower reactions on the other hand, have a large impact over wider regional or global spatial scale (Abdul-Wahab et al., 2005). The latter is particularly common in the province of Ontario, especially in southern Ontario because of the major emission contribution from neighboring U.S. states. The elevated level of ground level ozone and particulate matters are associated with distinct weather patterns which affect air quality in the lower Great Lakes Region. These weather conditions are strongly linked with slow-moving high pressure systems at south to the lower Great Lakes resulting in long-range transport of the smog pollutants from neighboring industrial and highly urbanized cities of the Mid-Western US and Ohio Valley during warm south to south-westerly wind flows. However, the adverse affects of this transboundary air pollution is more prevalent in western Ontario whereas the pollutants originating within Ontario have high impact in south central Ontario i.e. Greater Toronto Area (GTA) and other major population centers of Golden Horseshoe which accounts to the 61% of the total damages within the region. According to 2003 demographic statistics, Ontario, the largest province of Canada is burdened with almost \$9.6 billion in health and environmental damages each year because of smog pollution of which health damages comprises approximately 70% of the total economic damages costing around \$6.6 billion per year (Yap et al., 2005). A detailed list of the sources and impacts of some major air pollutants are shown in Table A.1 (Appendix).

As the sources and processes of environmental pollution are often originated from same source, a proper action can produce multiple improvements, also known as ‘co-benefits’ to the quality of air. For example, controlling nitrogen oxides emission alone can reduce smog, acid rain and green house gas warming. Therefore, an accurate modern and reliable air pollution forecast can play a significant role in managing the air quality system in a region. The complex relationship between meteorology and pollutant formation has been well documented by many authors (Derwent et al., 1998; O’Hare and Wilby, 1995; Abdul-Wahab et al., 2005) and attempts to develop satisfactory forecasting models have been numerous. But model selection has always been problematic because of the complex nonlinear relationship between the pollutant and the meteorological variables.

Neural networks are data-driven methods particularly suitable for the modeling of complex nonlinear functions. Because of its ability to forecast based on training from a wide range of historical databases, neural network based models have been shown to be powerful techniques for modeling air quality. Therefore, in addition to numerical methods available for air pollution modeling, simple-easy to use tools such as neural networks are may be used where human and financial resources are not available to allow complex numerical modeling. The proposed research aims to provide a comprehensive investigation of the potential of emergent neural network models to provide a robust air quality forecasting tool.

1.3 Research objectives

Although many researchers have undertaken several studies to model the air pollution problem in Hamilton, there is no well established nonlinear data-driven tool which can establish a good linkage between the historical and future pollutant situation based on meteorological information. The ultimate objective of this research is to explore the capability of the neural network models to capture this complex nonlinear meteorology-

pollutant relationship in Hamilton census metropolitan area and to develop a good forecasting tool for ground level ozone and nitrogen dioxide which are considered as two major problems in this region. The specific objectives of this study can be summarized as follows:

1. To investigate and characterize the complex nonlinear temporal and spatial variability of key air pollutants and their dependence on the meteorology in Hamilton region;
2. To assess the applicability of emergent neural network architecture, namely time lagged feedforward neural networks (TLFN), recurrent neural networks (RNN) and Bayesian neural networks for air quality modeling;
3. To develop an air quality forecasting tool based on the most robust neural network architecture;
4. To compare the optimal structures of these neural network architectures and configurations with the benchmark conventional multilayer perceptron (MLP) in order to determine the best possible predictions of ozone (O_3) and nitrogen dioxide (NO_2) concentrations in Hamilton;
5. To investigate the capability of the neural network model to simulate total suspended particulates (TSP) and inhalable particulate matter (PM_{10}) concentrations and to explore the methodologies for their improvements by introducing land use variables;
6. To explore the capability of the neural network models to simulate missing pollutant values at sampled sites based on nearby stations;
7. Predict pollutant concentrations at locations where sampling has not been done.

1.4 Structure of the thesis

This thesis is composed of 8 chapters, including an introduction in chapter 1. Chapter 2 presents the geographical characteristics, meteorology and the air quality situation of the study area of Hamilton. A brief description of the key air pollutants, meteorological and

land use database used in the study has also been presented. Finally, the data filling techniques applied to fill the missing gaps are described.

Chapter 3 covers literature reviewed for the purpose of this research. A comparison of deterministic and statistical models has been presented. A brief review of the neural network methodology, its architecture and configurations, advantages and limitations and application in air quality modeling has also been presented.

Chapter 4 explains various neural network methodologies along with a presentation of different model parameters and their significance in modeling. Finally a brief explanation of the spatial interpolation techniques has been described.

Chapter 5 emphasizes aspects of network design. Input variables selection criteria and the building of different neural network models have been presented. Lastly, a review of the statistics used during the evaluation of different model performances has been presented.

Chapter 6 and 7 summarizes the results obtained from the analysis for ground level ozone (O_3), nitrogen dioxide (NO_2), total suspended particulates (TSP) and inhalable particulate (PM_{10}) concentrations, respectively. A comparison of the forecasting performances of four neural network models specifically time lagged feed-forward neural network (TLFN), Bayesian neural network (BNN), recurrent neural network (RNN) and multilayer perceptron (MLP) models are presented for O_3 and NO_2 concentrations. In chapter 7 different cases of input variables are being tested using meteorological and land use variables in order to get best cases for each station. The simulated concentrations achieved with the best model structures are then interpolated spatially in order to obtain the concentrations at unsampled sites.

Finally concluding remarks and recommendations for future works have been presented in chapter 8.

Chapter 2

Study Area and Database

2.1 Study area: Hamilton census metropolitan area

2.1.1 Description of the study area

The area of the study (Fig. 2.1) is situated in the south-central Ontario which includes the city of Hamilton ($43^{\circ}15' N$, $79^{\circ}51' W$) and some part of the city of Burlington ($43^{\circ}19' N$, $79^{\circ}48' W$) located on the west shore of Lake Ontario approximately 100 km southwest from the city of Toronto.

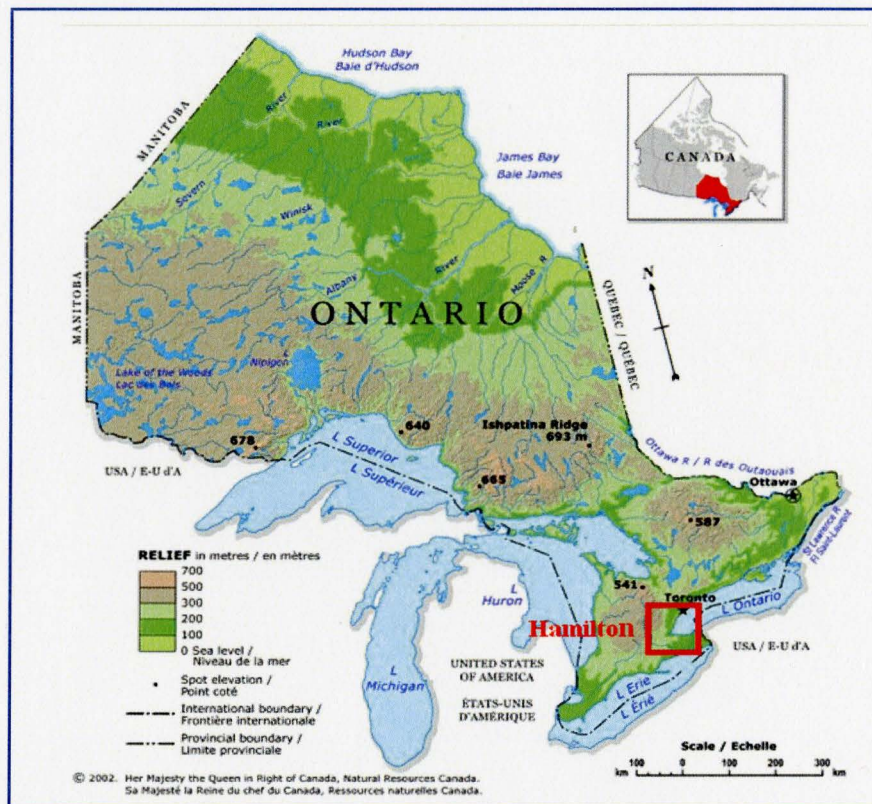


Fig. 2.1: Location of study area in Ontario

The area comprises of complex temporal and spatial climatic control because of its geographic location, topographic variations, urban morphology and land-water contrasts (Pouliou, 2005). It is located on the western end of the Niagara Peninsula and extended around the westernmost part of Lake Ontario. The two major physical features of the city are Hamilton harbor which marks the northern limit of the city and an approximately 100 m high Niagara escarpment which runs through the middle of the city thereby bisecting it into ‘upper’ and ‘lower’ levels.

In fact, the geography of Hamilton has played a key role in its development. During the 18th century Hamilton continued to boom below the mountain and started to expand towards the escarpment as we reached closer to 19th century. The outcome was the formation of a special urban form with distinct spatial patterns of socio-economic status within the city, with people of higher socio-economic status living in the south-west end and those with lower status tending to settle along the harbor and industrial areas in the north and northeast part of the city (Pouliou, 2005). The city is similar to many US cities in the Northeast, Midwest and Appalachia regions in terms of its high population density and decreasing heavy industry (Buzzelli & Jerrett, 2004). At present, Hamilton CMA is Canada’s 9th largest metropolitan area. It is considered to be one of the fastest growing metropolitan areas with 7% increase of population between 2001 and 2004 reaching from 0.49 million in 2001 to 0.71 million in 2004 (Statistics Canada, 2005).

Fig. 2.2 presents the highways and industrial areas around Hamilton region. A 200 km shoreline from Oshawa in north and extending up to Stoney Creek, at the east of Hamilton is continuously urbanized (Blair, 2006). Toronto-Hamilton region covering from Oshawa around the west end of lake Ontario to Niagara falls, also known as the ‘Golden Horseshoe’ is known as the most highly industrialized section of the country and Hamilton is the centre of it. Much of the development is occurring in the suburban areas at the south and south western part of the city which also aid in the increase of the traffic related emissions.

Two heavily used major highways Queen Elizabeth Way (QEW) and Chedoke Expressway (Highway 403) pass through the north-west and eastern part of the city. The QEW is one of the North America's oldest long-distance superhighways with over 200,000 average trips per day runs along the lakeshore from Toronto to the city of Buffalo, New York, USA. Highway 403 forms a loop from Highway 401 in Woodstock, passes through Hamilton and Burlington before terminating to the junction of Highway 401 and highway 410 in Mississauga.

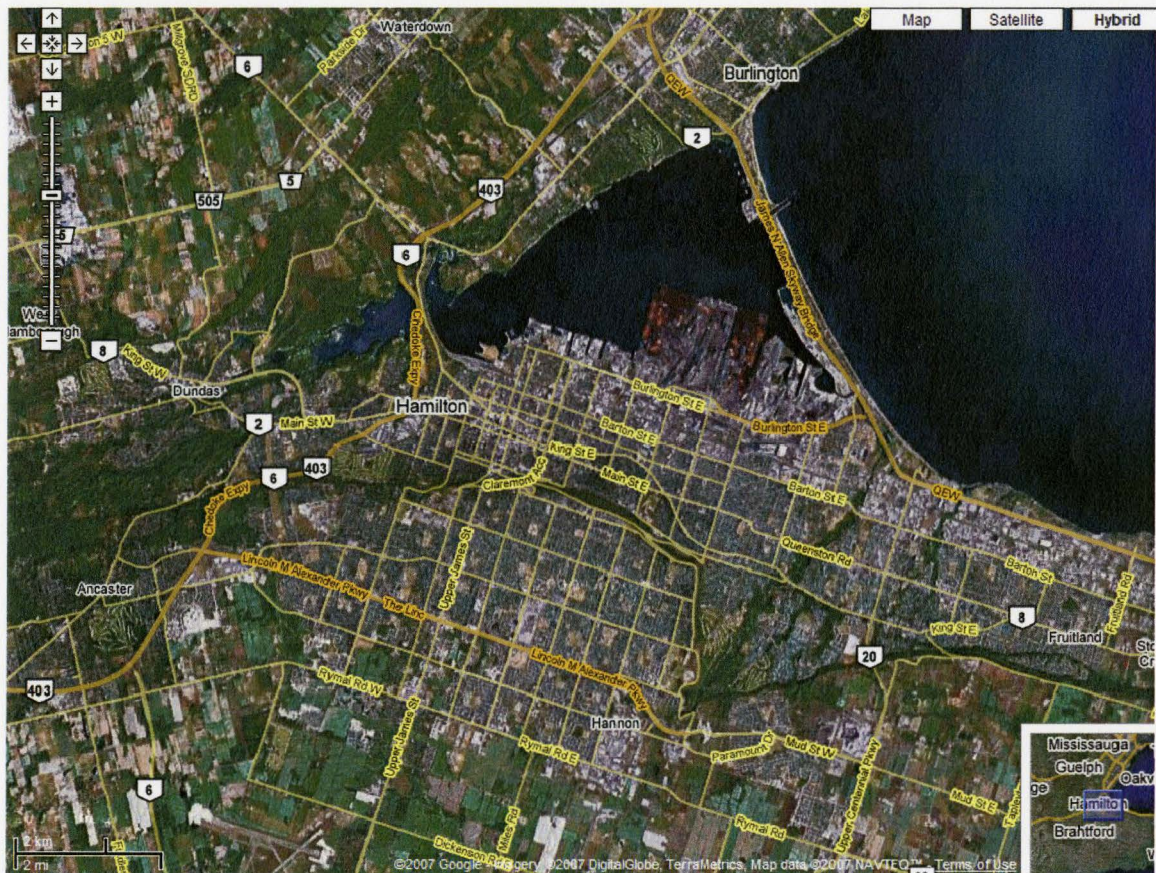


Fig. 2.2: Highways and industrial area around Hamilton (Source: Google map, 2007 / Imaginary Digital Globe, Earth Sat)

2.1.2 Meteorology of the study area

The climate of Hamilton according to the Koppen Climate Classification is Dfa type: humid continental, which means humid with severe winter and hot summer (KCCC, 2007). Lake Ontario, on the other hand tends to moderate the climate with cool summers and warmer winters (Blair, 2006).

The average temperature during January is -6.0°C which may fall to -28°C during extreme weather conditions. In July average temperature remains around 22.5°C which may rise up to 38°C during extreme conditions. Humidity remains higher during summer months; daytime highs in around 30°C with humidex values over 40°C during May through early October days. Heavy rainfall can occur during some summer days but weather rarely becomes extreme. Average annual precipitation recorded at Hamilton Airport is 899 mm. The average January snowfall is 113 cm which varied greatly from year to year. The escarpment has great effect on the climate in Hamilton; the weather of the lower part of the city is milder than on the Mountain which during winter is more prone to the lake effect snow carried by the wind. Again during summer temperature inversion occurs making the downtown warmer than the Mountain especially at night.

2.1.3 Air quality in Hamilton

Hamilton experiences one of Canada's highest air pollution exposures. For many years the dual problem associated from continuing industrial emissions and gradual increase of traffic emissions has been the main challenge to fight with. Like other cities in Southern Ontario, Hamilton's air is mainly affected by automobile, particularly truck traffic emission in and around the city, residential fuel use and local point source emissions from heavy duty industries within the city. The major industrial core situated in the northeast end comprises of one of the North America's largest steel making complexes producing spatially concentrated emissions in the northeast end of the city. For decades industrial emission was the major source of pollution in the region especially for TSP. The major expressways also run along the same section of the city and the

mobile sources are also contributing to the degrading air quality. The transportation sector is a leading sector of nitrogen oxides emission. Commuters and heavy transport trucks, in and around the city contribute much to the degrading quality of air. Transboundary pollutions from neighboring US states-Ohio, Illinois and Michigan are also key contributors of elevated ozone and particulate matters during summer season. Especially the pollutants emitted from the coal fired generating stations in the heavily industrialized Ohio River Valley, 1300 km southwest, travel hundreds of kilometers and pollutes Hamilton's air (Sahsivaroglu and Jerrett, 2003). The Nanticoke coal-fired generating station located on the northern shore of Lake Erie and 53 km south of the Hamilton city, also contributes to the local air pollution when wind blows from south. Moreover, the escarpment has a great effect on Hamilton's air, especially during summer. The north east lake breeze originated from Lake Ontario and the 100-120 m high escarpment above Lake Ontario, located 3 to 4 km from its shore combines together to produce advective temperature inversions. These inversions disperse pollutants from the steel core area along the lakeshore towards large and densely populated parts of the city. As a result, gradients of pollutants are created which run from the higher north east part of the city to the lower south, west and east end of the city (Buzzelli & Jerrett, 2004).

Recent air quality trend suggests that there have been significant reductions of benzene, benzopyrene, total reduced sulphur and sulphur dioxide (SO₂) concentrations from Hamilton's air due to actions taken to reduce emissions from industrial sectors and to smaller extent, the transportation sector. But less progress has been achieved in case of the criteria pollutants; levels of PM₁₀ and NO₂ remained unchanged and O₃ concentrations increased over the last decade. Fig. 2.3 (a) and (b) shows a comparative study of 10 year trends of O₃ and NO₂ concentrations at some highly polluted sites within Ontario. Ozone concentrations in Hamilton have increased compared to other Ontario sites with highest (20%) increase in Hamilton Mountain. Although NO₂ concentrations have decreased in 10 years, the rate of decrease in Hamilton is slower than other southern Ontario cities.

Between 2000 and 2004, 35 smog advisories were issued by Environment Canada for 95 days over Ontario (OME, 2004). In 2004, of eight total smog advisories released covering 20 days, seven were issued due to elevated ozone levels. The remaining one was issued due to higher levels of fine particulate matter with Air Quality Index (AQI) reading 103 and particularly confined to the Hamilton area for two days, resulting from the combination of local emissions and meteorological factors conducive to the build up of pollutants which included a frontal inversion created by a cold front across the city (OME, 2004). This clearly suggests that local impacts also contribute to the poor air quality in the city.

2.2 Database

2.2.1 Air pollution database

For this research, a pollutant database was obtained from the Ontario Ministry of Environment (MOE) s' Air Quality Information System (AQUIS) database. The ambient air quality network consists of 143 continuous monitoring instruments at 44 sites operated by the MOE's Environmental Monitoring and Reporting Branch (EMRB). In this study three O₃ and two NO₂, five TSP and three PM₁₀ monitoring sites within Hamilton CMA have been considered during 1994-2004 (Fig. 2.4). These sites are located in various parts of the city. The Hamilton Downtown site (29000 for O₃, NO₂, TSP; 300 for PM₁₀) collects O₃, NO₂, PM₁₀, TSP, CO₂ and is located in the Beasley Park which is a densely populated area close to Wilson Avenue, a heavily traveled one way street. The Wilson Avenue is believed to have direct influence, especially on O₃. There are no major industries near the site; the heavy industry is located approximately 2.5 km north-east of the site. Commercial district is located within 100 meters on the major arterial King and Cannon streets.

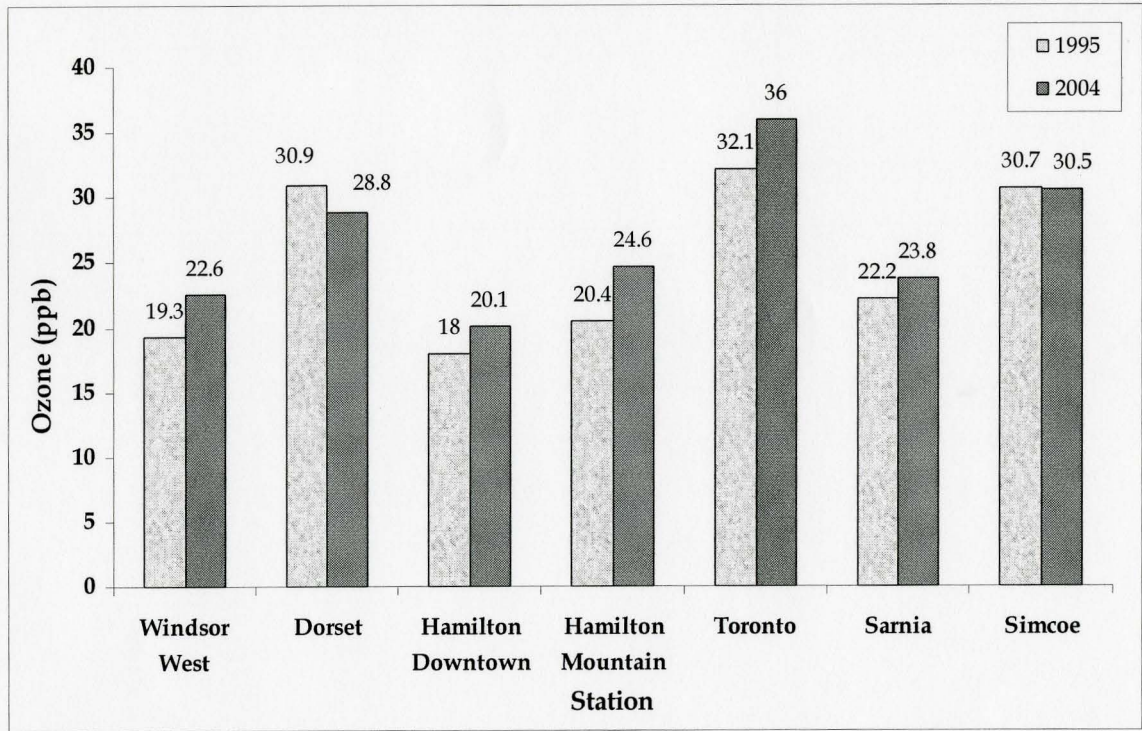


Fig. 2.3 (a): 10 year trend of ozone at some selected Ontario sites (Data source, OME, 2004)

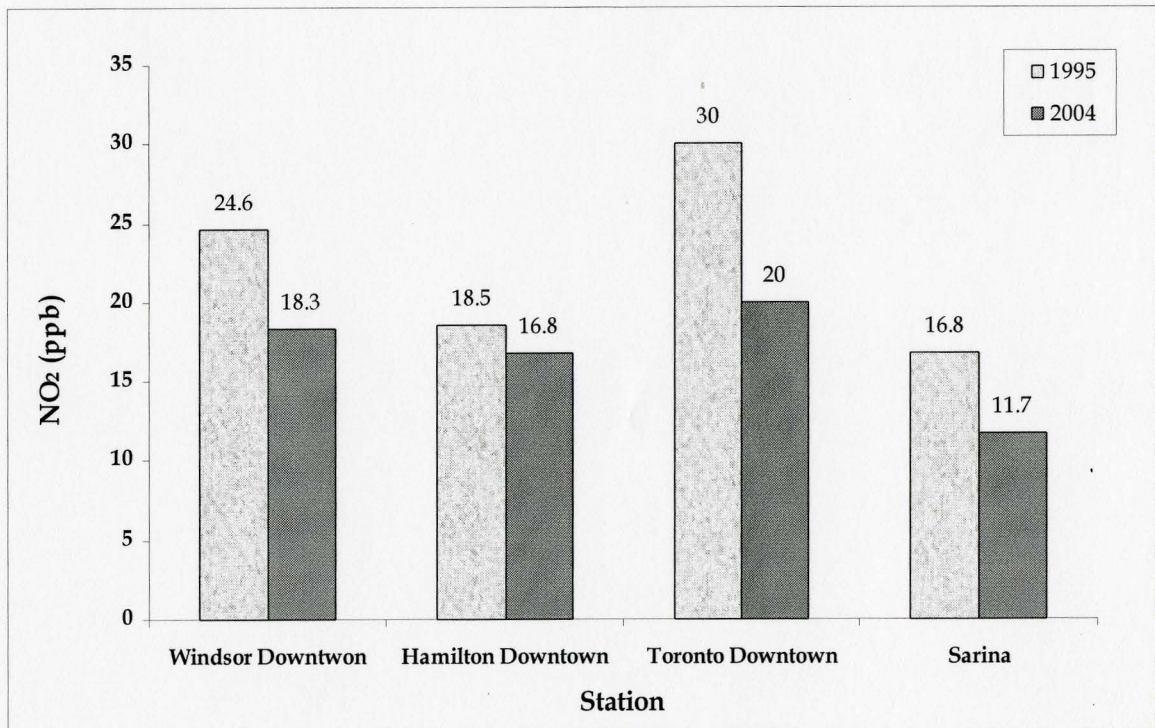


Fig. 2.3 (b): 10 year trend of NO₂ at some selected Ontario sites (Data source: OME, 2004)

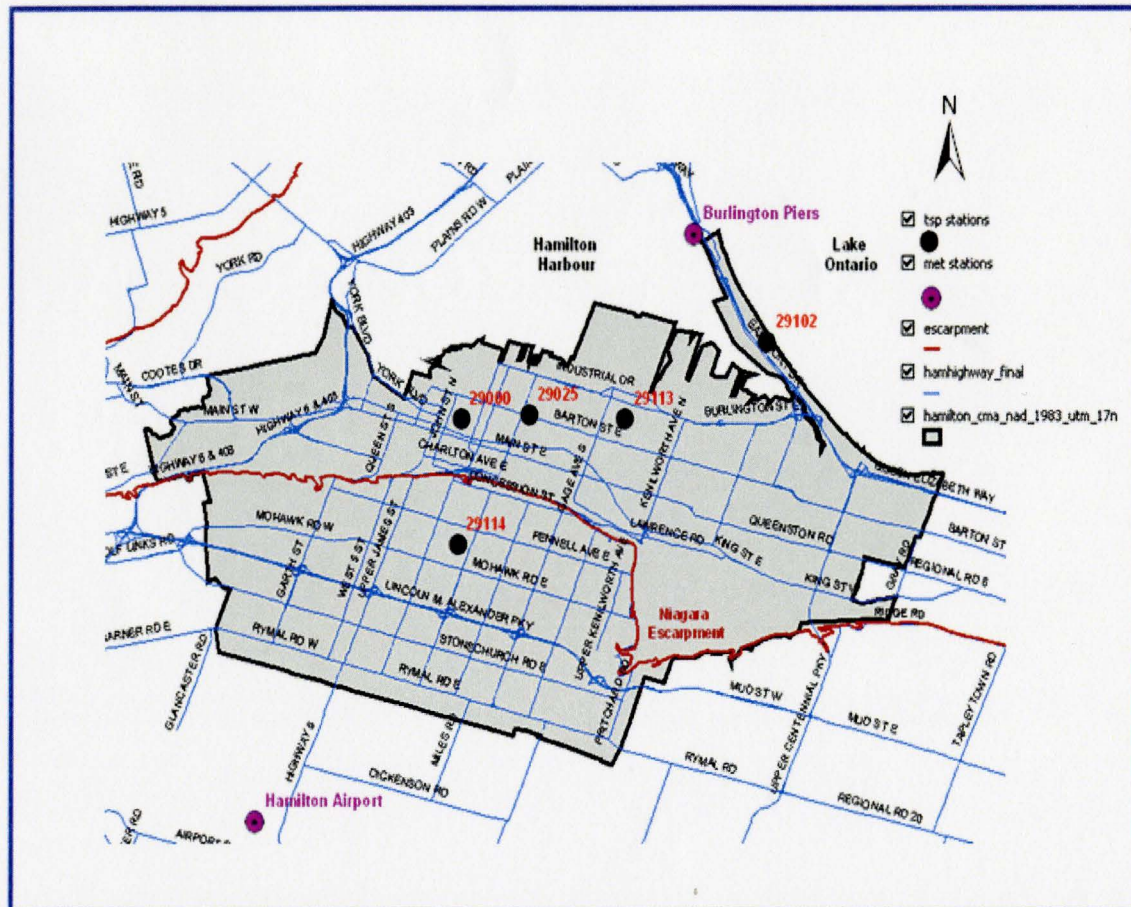


Fig. 2.4: Ontario MOE pollutant monitoring stations in Hamilton

The second site, on the mountain (29114) mainly collects O_3 and TSP and is situated on grounds of Linden Park School, at Vickers Avenue at East 18th Street. It is mainly residential except the school. There are no major arteries within 2 to 3 blocks; the nearest heavy industry is located at 5 km to the northeast below the escarpment. The third station 'Hamilton West' (29118) is established on grounds of hydro transformer Station on Main Street West, adjacent to the off-ramp from Highway 403. The area is generally residential but heavily traveled roadways dominate the immediate vicinity (i.e. HWY 403 and main street west). Heavy industries are situated 5 to 6 kilometers northeast. It collects O_3 and NO_2 .

O₃ and NO₂ measurements are being made using UV Photometry TE49C and Chemiluminescence TE42C analyzers developed by the Thermo Environmental Instruments Inc., MA, USA (OME, 2005). O₃ and NO₂ have been collected at hourly intervals which for this analysis were further reduced to bi-hour intervals in order to reduce the computational time.

TSP and PM₁₀ has been collected from five sites during the study period. Instruments were set up in Downtown, Mountain, Barton-Sanford, Gertude / Depew and Beach Boulevard sites. The Barton-Sanford and Gertrude / Depew sites are located in the industrialized zone of the city whereas Beach Boulevard site is situated on the north-west part near the lake Ontario. MOE collects particulate matter data on a six-day North American Synoptic Cycle (Jerrett et al., 2001); hence, both TSP and PM₁₀ observations are available in 6 day intervals for the study period of 1995-2004. Table 2.1 and 2.2 represents a comprehensive description of the study sites and the database considered in this research.

Table 2.1: Air pollutant stations in Hamilton

Station name	Station ID	Pollutant type	Coordinate		Period	Frequency
			Latitude (N)	Longitude (W)		
Hamilton Downtown (Elgin/Kelly)	29000	O ₃	43.26	-79.86	1994 -2004	2 hr
		NO ₂				2 hr
		TSP				6 day
		PM ₁₀				6 day
Hamilton Mountain (Vickers/East 18th)	29114	O ₃	43.23	-79.86	1994 -2004	2 hr
		TSP				6 day
Hamilton West (Main W. / Highway 403)	29118	O ₃	43.26	-79.9	1994 -2004	2 hr
		NO ₂				2 hr
Barton / Sanford	29025	TSP	43.26	-79.84	1994 -2004	6 day
Gertrude / Depew	29113	TSP	43.26	-79.82	1994 -2004	6 day
		PM ₁₀				6 day
Beach Blvd.	29102	TSP	43.28	-79.86	1994 -2004	6 day
		PM ₁₀				6 day

2.2.2 Meteorological database

Meteorological information was collected from the Ontario Climate Centre of Environment Canada. Three weather stations, located near the monitoring sites were considered initially. The stations are: Hamilton Airport (43°10'N, 79°55'W), Burlington Piers (43°18'N, 79°48'W) and Royal Botanical Garden (RBG) (43°17'N, 79°54'W). The data for the Hamilton Airport was available from 1990 and is still under full operation. The second station in Burlington had data available from 1994; however, the RBG site started operating on 2000 and its inclusion in the modeling did not significantly improve the result. Therefore the RBG station has been excluded from this analysis.

Table 2.2: Pollutant statistics (1994-2004)

Station ID	Station name	Pollutant	Missing values	Mean	Std. dev.	Max	Percentile		
							50	75	95
29000	Hamilton Downtown	O ₃	565 (0.15%)	18.93	15.71	114.50	16.00	27.00	49.50
		NO ₂	452 (1.2%)	21.33	11.01	100.00	20.00	28.00	41.50
		TSP	41 (6.96%)	61.36	31.70	204.00	53.00	77.00	125.60
		PM ₁₀	46 (7.74%)	25.00	14.00	91.00	22.00	31.00	52.00
29114	Hamilton Mountain	O ₃	1054 (2.49%)	23.97	16.93	120.50	22.00	33.00	56.00
		TSP	91 (15.45%)	48.65	23.94	168.00	44.00	63.00	93.00
29118	Hamilton West	O ₃	573 (1.36%)	19.12	15.72	107.00	17.00	28.00	49.00
		NO ₂	831 (2.28%)	20.22	12.00	142.50	18.00	27.50	42.50
29025	Barton/ Sanford	TSP	34 (5.77%)	72.24	33.45	224.00	67.00	88.00	137.00
29102	Beach Boulevard	TSP	51 (8.66%)	78.81	42.44	309.00	74.00	100.00	159.00
		PM ₁₀	47 (7.91%)	28.00	18.00	142.00	25.00	36.00	60.00
29113	Gertrude/ Depew	TSP	61 (10.36%)	89.93	50.66	348.00	77.00	111.00	187.00
		PM ₁₀	75 (12.63%)	36.00	20.00	121.00	30.00	46.00	79.00

Solar radiation, maximum temperature, wind speed, wind direction, relative humidity, dry bulb temperature, vapor pressure, etc. are the most important meteorological variables for O₃ and NO₂ (Gardner and Dorling, 2000; Agirre-Basurko et al., 2006; Schlink et al., 2006). However, successful use of these variables in air pollution modeling strongly depends on the availability of good quality data. For this specific work, it would have been ideal to include maximum temperature and solar radiation as input because they are very good indicators of the smog formation (Gardner and Dorling, 2000). Unfortunately these data were not available in any of the weather stations located near the three selected air quality monitoring network stations. The vapor pressure variable also could not be used because of low quality data with high number of missing values. An alternative way could be deriving the vapor pressure values from the relative humidity measurements using a known empirical formula. However, this approach was not considered as only the directly available database has been used for this study. Therefore, finally only four variables: wind speed (km/hr), wind direction (10s of degrees), dry bulb temperature (0.1⁰C) and relative humidity (%) data were used as input variables with the assumption that they might be able to capture the real chemistry of O₃ and NO₂ dispersion. Like O₃, the meteorological variables were also collected on an hourly basis during 1994-2004 which for this work was converted to bi-hourly values.

For the modeling of particulate matters more meteorological variables were used than O₃ and NO₂ model because of availability of daily values. Here total 10 meteorological variables namely wind speed, wind direction, relative humidity, maximum temperature, minimum temperature, mean temperature, precipitation, visibility and barometric pressure were considered and they were downloaded from the Environment Canada website.

Table 2.3: Meteorological variables used

Pollutant type	Meteorological variables	Notation	Unit
O ₃ and NO ₂	Wind Speed	WS	km/hr
	Wind Direction	WD	10s of degrees
	Relative Humidity	RH	%
	Dry Bulb Temperature	Temp	0.1 ⁰ C
TSP and PM ₁₀	Wind Speed	WS	km/hr
	Wind Direction	WD	10s of degrees
	Relative Humidity	RH	%
	Maximum temperature	Tmax	⁰ C
	Minimum temperature	Tmin	⁰ C
	Mean Temperature	Tmean	⁰ C
	Precipitation	Prec.	mm
	Barometric Pressure	BP	KPa
Visibility	Vis.	Km	

Measurements for the weather variables have been taken from 10 meter above the ground using standard measurement procedures. The temperature and relative humidity are measured with an MSC “dewcell” (Type E) thermometer which consists of a fiberglass sleeve saturated with a lithium chloride solution. The dewcell, accurate to 0.6⁰C above and 1.2⁰C below freezing is calibrated weekly using a mercury thermometer. The wind speed and wind direction measurements are taken using U2A anemometer. Average wind speed values are being measured at two minute period. The direction of wind is determined the by wind blowing with respect to the true or geographic north (360 degrees on the compass) and is expressed to the nearest 10 degrees. Daily barometric pressure and daily precipitation measurements were available only for Hamilton Airport site. Table 2.3 presents the meteorological variables used in the study.

2.2.3 Land use database

The land use database has been acquired from Desktop Mapping Technology Inc. (DMTI Spatial Inc.) through McMaster University Library. The DMTI classifies the entire urban area in Ontario into seven distinct types which include:

- (1) Commercial
- (2) Governmental and institutional
- (3) Open Area
- (4) Parks and recreational
- (5) Residential
- (6) Resource and industrial and
- (7) Water body

The land types within Hamilton were derived using spatial overlays of buffers with various radii around the monitoring sites. Percentage area of land use (in hectares) within buffers around the monitors with radii of 50, 100 and 200 m were tested and a 200m buffer appeared to provide better results. Table 2.4 presents the total percentages of various land uses within 200 m buffer area of all monitoring stations. The table shows that the Downtown station at Elgin/Kelly has mostly residential land (41.63%) followed by 25.95% resource and industrial area. The area near Barton/Sanford station serves multiple purposes with 32.75% residential land, 20.35% parks and recreational land, 18.06% governmental and institutional and 21.53% resource and industrial land area. The Beach Boulevard station is mainly residential with 84.98% residential development around it while the site located at Gertrude/Depew is entirely within industrial land area. The land area near the last station located on Mountain (Vickers/East 18th) also is multipurpose with 38.57% residential, 28.86% parks and recreational and 32.57% government and institutional area.

Table 2.4: Land use within 200 m buffer of all stations

Station name	Station ID	Land use type	Proportion of area (%)
Elgin/ Kelly	29000	Resource and industrial	25.95
		Residential	41.63
		Parks and recreational	8.45
		Government and institutional	21.27
		Commercial	2.67
Barton/Sanford	29025	Commercial	7.30
		Government and institutional	18.06
		Parks and recreational	20.35
		Residential	32.75
		Resource and industrial	21.53
Beach Blvd	29102	Water body	4.06
		Resource and industrial	3.75
		Residential	84.98
		Parks and recreational	7.21
Gertrude/ Depew	29113	Resource and industrial	100.00
Vickers/East 18th	29114	Residential	38.57
		Parks and recreational	28.86
		Government and institutional	32.57

2.2.4 Data filling technique

Instrument malfunctioning and maintenance cause some missing values both in the pollutant and the meteorological database; mostly up to four or five hours in a row. The following methods were applied for filling these gaps:

(1) For any specific time period, the missing values were filled using multivariate regression method (partial least square approach) based on nearest station values. The partial least square method is a generalized principal component analysis where a projected model is developed by predicting the Y values using the values of X . It can also be seen as a multiple regression method which has the ability to deal with multiple collinear X and Y variables. The model is expressed as:

ITEMS CHECKED OUT

THODE LIBRARY
11/12/08 03:33PM

Dates are in MONTH-DAY-YEAR format

RENEW YOUR MATERIALS ONLINE
Click on Library Catalogue at:
library.mcmaster.ca

9005014574721 DUE: 11/26/08
CALL NO: THESIS TA 07 031
TITLE: Air pollution modelling and forec

TOTAL: 1

$$Y = f(X) + E \tag{2.1}$$

f a polynomial and X refers to the predictor variables. The E corresponds to a fitted line, plane or hyper plane where both X and Y are treated as points in a multidimensional space in order to best fit the data. E can be further expressed as:

$$\begin{aligned} E &= 1 * \bar{X} + TP' + E \\ E &= 1 * \bar{Y} + UC' + F \\ E &= T + H \end{aligned} \tag{2.2}$$

T and H are principal components which summarize the X and Y variables respectively, U and V define the correlation between Y and T (X) and E , F and G are principal loadings W , expressed as weights are also present in the correlation between U and X and are used to calculate

For a specific time period, the temporal nearest neighbor (row interpolation) approach was considered to fill short gaps (up to 4 hours).

(3) For medium and larger gaps (4-8 hours), missing data were filled using the average of the previous year's value for that specific time period.

(4) For missing values of more than 10 days, that specific month was simply removed from the database. For both O₃ and NO₂, records are missing from Mid April 1997 to December 1997 and Mid March 2002 to Mid May 2002. So the values for 1997 and months March to May of 2002 have not been considered in this study.

(5) The missing values in wind direction data were filled by 1st order autoregressive or Markov model, denoted by AR (1). It can be expressed as:

$$x_t = \mu_x + \rho_x (1)(x_{t-1} - \mu_x) + \epsilon_t \tag{2.3}$$

Where, $\rho_x(1)$ is the first order serial correlation, μ_x is mean value, x_{t-1} is one day lagged observation and ε_t is the error.

Chapter 3

Review of Literature

3.1 Basic air quality modeling approaches

Studying the literature for all pollutants, a general point arises: no model can be regarded as universal; several classes of models are developed to solve particular pollution problems. Generally, air pollution modeling is carried out based on their features with respect to stationary or non stationary source conditions, meteorological conditions (stability, presence of thermal inversion, etc), type of emission sources (industrial, vehicular, continuous, etc), type of terrain (flat, mountainous or complex, etc), type of pollutant (inert or reactive) and time horizon of the simulation (Schlink, 2001). However, mathematical methods are widely used in air pollution modeling because of their higher capability to evaluate varying scenarios of different pollutants under different atmospheric conditions. Air pollution modeling is carried out mainly by two mathematical methods: *deterministic methods* and *statistical methods*.

3.1.1 Deterministic methods vs. Statistical methods

Deterministic models follow fundamental mathematical descriptions of atmospheric processes based on known scientific laws or relationships in which the output is represented by the air pollution concentration field and the inputs by emissions (Pagina, 2005). However, different strategies are followed in using these models. The Clean Air Act Amendments (CAAA), established by the Clean Air Act, USA established emissions reduction network to reduce risk to public health and to protect sensitive ecosystems. According to CAAA, “if a state is found with substantial violation of some of air pollution standard in the Clean Air Act, to bring it back with in the standard by proving the effectiveness of the plan, a comprehensive deterministic air pollution model

describing emissions, air transport, chemical transformation, and deposition of the pollutant and its precursors should be built. These models must be compared to the observed data to describe the situation well and to see how well the proposed controls are working”. And here lies the major draw back of the deterministic models. The act further extends “the deterministic models produce predictions for grid square over some temporal window. The data are obtained at individual points and often have a different temporal resolution from the model output. Consequently, it is not possible to compare the data to the model output directly. Hence, some manipulation of the data or the model output is recommended for comparability. But since the model output is already over the grid square, it seems inappropriate to smooth it further spatially in order to compare to non-smooth point measurements. Rather, we would be inclined to use the data to predict the model output, i.e. to predict the grid average values. However, this requires a rather data-rich situation, in which the prediction can be made with adequate precision”. Further more, deterministic assessments using deterministic models are considered subjective and limiting in their scope and do not objectively consider the possibility of deviation from the fixed values. Deterministic models, also known as cause/effect models are more suitable over spatially extensive areas like whole regions or large cities. They can be very important tool for practical applications since, if properly calibrated and used, they have the ability to provide unambiguous, deterministic source-receptor relationships. However, they require a large amount of data (e.g. emissions, gas temperature, wind data, air temperature, topography of the study area, etc). Hence in many occasions unavailability of sufficient data is one of the major causes of the uncertainty of deterministic models.

Statistical models, on the other hand, are based upon semi-empirical relationships between known past air pollutant concentrations and specific meteorology. Mathematical formulae are utilized to evaluate the influence of different meteorological conditions on contaminant concentrations which have already been dispersed into the atmosphere or will be formed photochemically. They are frequently used in air pollution studies especially for short term forecasting applied to the real time control of emissions or to air

quality assessment. Statistical methods, unlike deterministic models instead of describing the air pollution problem as a cause-effect problem, use air quality measurements to infer semi-empirical relationships.

Statistical Models can either be used as a '*black box mode*' where the time series of pollutant concentration are explained without any information in order to evaluate their intrinsic variations and without attempting any physical explanation, or they can be used as a '*grey box mode*' where within a statistical framework, deterministic relations are integrated using other conditions like meteorological phenomena (commonly available from air quality and meteorological monitoring networks) and emission patterns (comparatively less available, especially in real-time). These models are preferred for relevant measured concentration trends information rather than those obtained from deterministic analyses (Schlink, 2001). Moreover, the structures of statistical models are often simpler than the deterministic models and they can more easily be implemented and even can be used by non-experts.

Statistical models, also known as data-driven models, can be used effectively for short-term forecasting where it is assumed that there is statistical regularity in the data which can be captured by means of a mathematical function approximation technique. The captured regularity is then used to forecast pollution levels at a future point in time. This method is extremely powerful in designing an early warning system (Niranjan et al., 2001).

The complex relationship between the meteorology and pollutant concentrations has been well documented by many authors and attempts to develop a satisfactory statistical air quality forecasting model have been numerous (Derwent et al., 1998; O'Hare and Wilby, 1995; Abdul-Wahab et al., 2005). But model selection has always been problematic because of the need of a suitable model which can satisfactorily map the complex nonlinear relationship between the pollutant and the predictor meteorological

variables. Nowadays it is well recognized that non-linear models have higher capability to capture the composite relations related to pollutant formation. Table 3.1 presents a brief summary of some statistical models used in air pollution modeling.

Table 3.1: Summary of statistical methods used in air pollution modeling
(Schlink *et al.*, 2001)

Model	Type	Limitations/ Advantages
Classical Time Series, Box-Jenkins	Linear Stationary Parametric	Time consuming: low Amount of data needed: low Model readable: high Sensitivity to missing/noisy data: medium/medium General Fitting capabilities: medium Forecast capability exceedence: low
Component Models	Linear Stationary Parametric	Time consuming: medium Amount of data needed: medium Model readable: medium Sensitivity to missing/noisy data: medium/low General Fitting capabilities: medium Forecast capability exceedence: medium
Cyclostationary Models	Linear Non stationary Parametric	Time consuming: low Amount of data needed: low Model readable: high Sensitivity to missing/noisy data: low/medium General Fitting capabilities: medium Forecast capability exceedence: low
Dynamic Linear Regression	Linear Stationary Parametric	Time consuming: Medium Model readable: high Sensitivity to missing data: low General Fitting capabilities: medium Forecast capability exceedence: medium
Multivariate Regression	Linear Stationary Parametric	Model readable: high Sensitivity to missing data: low General Fitting capabilities: high Forecast capability exceedence: medium
Fuzzy Models	Non linear Stationary Parametric	Time consuming: high Amount of data needed: high Model readable: high Sensitivity to missing/noisy data: low/medium General Fitting capabilities: medium Forecast capability exceedence: medium

Generalized Additive Models	Non Linear Non stationary Non parametric	Time consuming: medium Amount of data needed: medium Model readable: high Sensitivity to missing/noisy data: low/medium General Fitting capabilities: high Forecast capability exceedence: medium
Hybrid Models	Non Linear Stationary Parametric	Time consuming: medium Amount of data needed: medium Model readable: medium Sensitivity to missing/noisy data: low/medium General Fitting capabilities: medium Forecast capability exceedence: medium
Neural Network Models	Non linear Stationary Parametric	Time consuming: high Amount of data needed: medium Model readable: low Sensitivity to missing/noisy data: low/low General Fitting capabilities: medium Forecast capability exceedence: high
Phase Space Embedding Models	Non linear Stationary Non parametric	Time consuming: medium Amount of data needed: medium Model readable: low Sensitivity to missing/noisy data: high/medium General Fitting capabilities: high Forecast capability exceedence: low
Wavelet Models	Non linear Stationary Parametric	Time consuming: high Amount of data needed: medium Model readable: low Sensitivity to missing/noisy data: low/low General Fitting capabilities: medium Forecast capability exceedence: high

3.2 Neural network model

Artificial neural network model, commonly known as neural network (NN) model can be defined as a massively parallel-distributed information processing system. This has certain performance characteristics that resemble to the biological neural networks of the human brain (Haykin, 1999; Rao, 2000a). It acquires knowledge through a learning process which involves finding an optimal set of weights for the connections and threshold values for the nodes (Rao, 2000b). Fig. 3.1 shows a conceptual representation of two biological neurons that inspired development of the model. Artificial neurons receive input from sensors or other artificial neurons, do calculations and pass outputs to other neurons. The key issue here is that information is processed by numerous neurons both parallel (by neurons belonging to the same layer) and linear (from neurons of one layer to another) (Bodri, 2000; Daliakopolous, 2004).

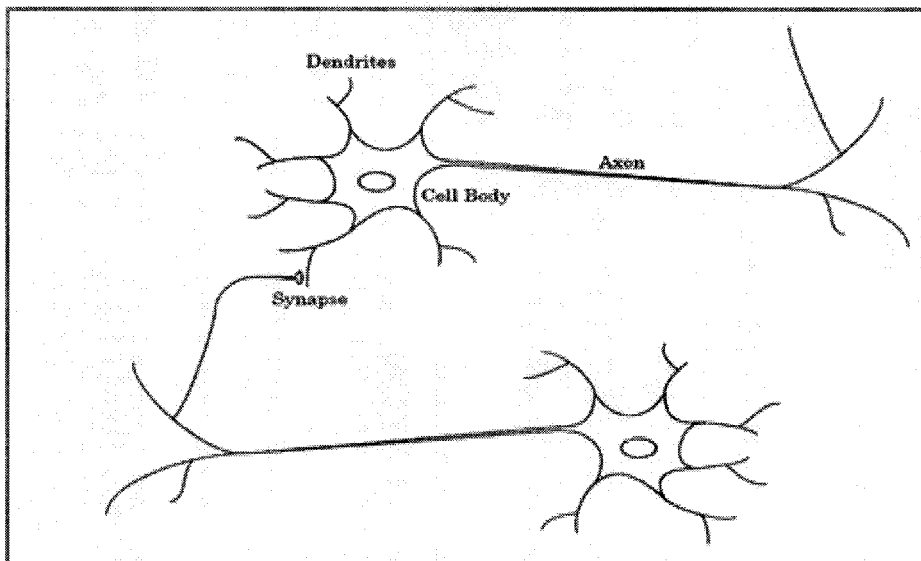


Fig. 3.1: Schematic diagram of biological neuron (Hagan et al, 1996)

3.2.1 Mathematical aspects

A schematic diagram of a neural network node has been presented in Fig. 3.2. Depending on the location of the node, the input variable may come from the system casual variables or the outputs of other nodes. The inputs form an input vector $\mathbf{X}=(x_1, x_2, \dots, x_m)$. Weights leading to the nodes form a weight vector $\mathbf{W}_j = (w_{k1}, w_{k2}, \dots, w_{km})$, where w_{km} denotes the connection weight from the k^{th} node in the preceding layer to this node (Rao, 2000a). The summing junction adds this weighted signals, hereby works as a linear combiner (Haykin, 1999). The output of the node k , y_k , can be computed from the function $\phi(.)$ with respect to the inner product of vector x and w_k-b_k , where b_k is the threshold value, known as bias, associated with this node. The bias increases or decreases the net output of the activation function depending on whether it is positive or negative, respectively.

Mathematically, a neuron can be described by the following terms (Haykin, 1999, Rao, 2000a):

$$y_j = \phi(X.W_j - b_j) \quad (3.1)$$

The function ϕ is called activation function and determines the response of a node to the total received input signal.

3.2.2 Important aspects of NN modeling

The development of a NN does not follow any specific rule and depends on previous successful applications in each field. Typically their development follows some general rules (Rao, 2000a):

- Information processing takes place at single elements called nodes or neurons;
- Signals are passed to adjacent nodes through the connection links;
- Each of these connection links has their own weights representing connection weight;
- Each node applies a nonlinear transformation called an activation function to its net output in order to determine the output signal.

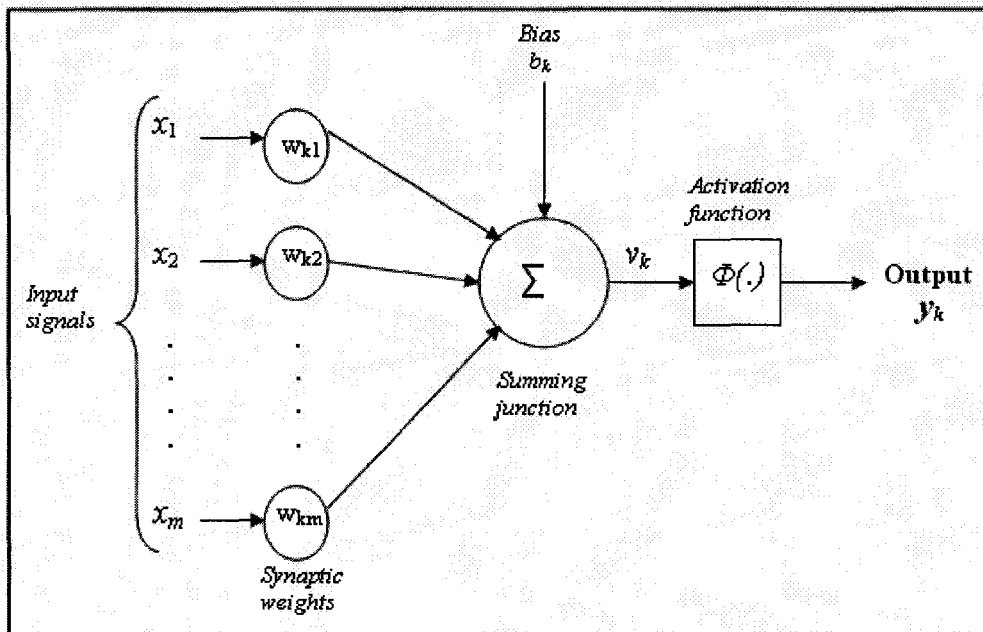


Fig. 3.2: Schematic diagram of a node (Haykin, 1999)

However there are certain issues that may become vital while developing the model. This section presents such issues that are of concern.

3.2.2.1 Selection of input and output variables

The major aim of an NN model is to generalize a relationship of the form (Rao, 2000a):

$$Y^m = f(X^n) \quad (3.2)$$

where X^n is an n-dimensional input vector constituted of variables $x_1, x_2, \dots, x_i, \dots, x_m$ and Y^m is a m-dimensional output vector consisting of resulting variables $y_1, y_2, \dots, \dots, y_m$. The term generalized implies that the functional form $f(\cdot)$ will not be revealed explicitly but rather be represented by the network parameter. In air pollution modeling variables x_i can be meteorological variables such as wind speed, wind direction, temperature, rainfall, relative humidity, solar radiation, cloud cover height; traffic variables such as traffic volumes, occupation percentage, velocity, etc; and emission sources such as

anthropogenic (combustion from coal fire plants, burned particles, traffic related emissions, waster deposition in landfills, oil refining, power plant operation and industrial activities, toxic gases, etc) and natural (dust, pollutants emitted from digestion of animals, smoke, volcanic eruptions) sources that are used to compute the values of y_i which normally are pollutant variables of interest (e.g. NO_2).

In NN based modeling, the selection of input variables is important to map successfully the undergoing process of pollutant formation as the influential variables are not always previously known. Hence a firm understanding of the pollutant forming system is an important prerequisite to apply NN successfully. For example, physical insight into the studied problems can lead to a better selection of the input variables. In this way loss of important variables can be prevented and also irrelevant variables, confusing the process can be removed. When a sufficient database is available, a sensitivity analysis can be used to measure the relevancy of the input variables with the output. Hence a more condensed or parsimonious network can be achieved by using only variables which have high sensitivity with the process.

3.2.2.2 Neural network architecture

A neural network can be characterized by its architecture, presented by the network topology and pattern of connections between nodes, its method of determining the connection weights and the activation functions that it employs. A typical neural network is composed of a series of nodes organized in parallel. They can either be classified according to the number of layers (e.g. single layer, bilayer or multilayer) or by the direction of information flow and processing within the network. A network consists of a hierarchy of processing units organized in a series of two or more mutually exclusive sets of neurons or layers. The information flow in the network is restricted to a flow, layer by layer, from the input to the output, hence also called feedforward network (Fig. 3.3). Thus the output of a node in each layer depends only on the inputs it receives from previous layers and the corresponding weights (Rao, 2000a).

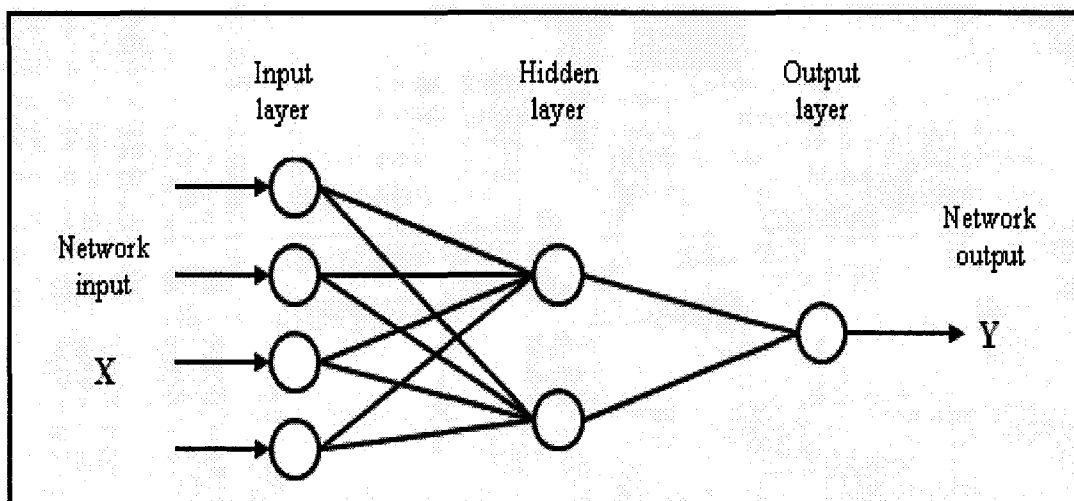


Fig. 3.3: Configuration of a typical feedforward neural network

On the other hand, in the recurrent network there is at least a feedback loop. Information flows through nodes in both directions- from input to outputs and vice versa. The presence of feedback loops has a profound impact on the learning capability of the network and on its performance. It involves the use of particular structures like unit-delay elements (or tap delay) which result in a non-linear dynamical behavior, assuming that the neural network contains nonlinear characteristics (e.g. nonlinear activation functions).

3.2.2.3 Network training, cross-validation and model testing

Generally the available database is divided into three parts: training or calibration, cross-validation and testing or validation. The major objective of training is to minimize the error function by searching for a set of connection weights and threshold values that allow the NN to produce outputs equal or close to the targets (Rao, 2000a). It can be of two types: supervised and unsupervised. A supervised training needs an external force to guide through the training process hence indicating that a sufficiently large number of inputs and outputs are required to map the underlying relationship between the input and output variables. It involves the iterative adjustments and optimization of connection

weights. The weights and threshold values are assigned randomly and are adjusted during training based on the difference between the model outputs and the target values. The process continues until a weight space resulting in lowest possible error is found. Hence overfitting or overtraining may take place resulting in a network depending more on the individual values rather than the overall trends in the dataset. This results in a very good training but with inferior prediction values.

To prevent such problems, cross- validation is recommended. The idea here is to stop training when the error starts to rise. Initially the error for both training and cross-validation goes down but after the optimal value is reached the errors in the training set continues to decrease while those in the cross- validation set starts to rise. It indicates that further training may overtrain the model. So training is stopped at this point assuming the current sets of weights and thresholds are optimal values. In the case of a small dataset, the easiest way to stop overfitting is to stop training when mean square error decreases significantly. The performance of a model output can be evaluated by subjecting it to a new pattern not seen during training. The performance of the model can be determined by computing the prediction error between the predicted and the desired values, plotting model output versus observed values and so on. It is also necessary to repeat the training and cross-validation to ensure satisfactory result (Rao, 2000a).

3.2.2.4 Advantages and limitations

There are various aspects which make neural network models an attractive tool. Firstly they are capable of recognizing the relationship between the input and output variables without explicit physical information. Hence there is no necessity to assume an underlying data distribution which is usually adopted by other statistical modeling techniques. NN models have the ability to model the nonlinearity of the underlying process without solving the complex differential equations. Moreover, they don't need any prior assumptions about the mathematical relations between the input and target variables unlike regression based models. They are less sensitive to error term

assumptions and can tolerate noise, chaotic components and heavy tails better than most of other methods (Slini et al., 2006). The distributed processing pattern of the network prevents large loss of information that can occur because of the noisy values in the input and output. Other advantages include greater fault tolerance, robustness and adaptability especially compared to expert systems, because of the large number of interconnected processing elements that can be trained to learn new patterns (Slini et al., 2006). To compensate the changing circumstances, they are capable to adopt new solutions over time. They possess other inherent information-processing characteristics. These characteristics, along with the non-linear nature of the activation function increase generalization capability of NN and make them desirable for larger classes of problems (Rao, 2000a).

However, NN models have their own limitations too. Like other data-driven models, success of a NN application depends largely on the quality and quantity of the database being considered. This most often makes the whole process complex and prevents its successful use. Even in a situation where a historical database is available, there is no certainty that the condition would be consistent over time. Hence a more stable and homogeneous database is desirable. The temporal variations can also be computed by considering past information of input/output variables. However, there is no clear indication on how far back should be considered which makes the whole process complicated. Another major limitation of any neural network model is they cannot be interpreted easily although several sensitivity tests and comparison of model performances may provide insight into the model. However, in argument it can be said that a good ‘black box’ model with good prediction accuracy is better than a poorly performed yet well understood physically based model. Another advantage of using NN models is there is no standard way of adopting a specific network for a given problem. The choice of the network architecture and learning algorithm has to be determined from the users’ past experience and preferences rather than the physical aspects of the problem (Rao, 2000a). It has been unable to reveal the cause-effect interactions of the phenomena

which, as suggested by Kolehmanien et al. (2003), is possible to overcome by introducing different types of neural networks together and by analyzing the characteristic behavior of the data prior to forecasting (Schlink et al., 2003).

3.3 Application of NN in air quality modeling

For the purpose of the thesis a large number of scientific papers were reviewed. They were mainly related to environmental modeling, especially air quality with an emphasis on neural networks. Priority was given to the papers that compared neural network model with other statistical models in terms of their performance. This section provides a brief review of some selected recent papers where NN has been successfully applied.

An elaborate literature review of different deterministic and statistical models used in modeling air quality have been presented in the project report funded by the European Community under the ‘Information Society Technology’ program (Schlink et al, 2001; Nunnari, 2001). In spite of its drawbacks of interpretation, a neural network model was included as one of the main tasks in that project because of its flexibility and capacity to model the non-linear behavior of complex atmospheric phenomena.

Benvenuto and Martini (2000) applied neural networks model for data quality control of environmental time series and reconstructing missing data. Their results confirmed NN to be an improved tool relative to classical models and depicted its utility in restoring time series methods. Several authors and researchers have compared neural network models with linear regression models (Yi and Prybutok, 1996; Comrie, 1997). Yi and Prybutok (1996) used nine input variables: the morning ozone concentration, the maximum daily temperature, levels of CO₂, NO, NO₂ and NO_x and wind speed and direction to predict the maximum daily surface ozone concentration in an industrial area and found the MLP model to be superior than the regression models. Similar results were

also obtained in the prediction of summer time daily maximum hourly ozone concentration in various urban areas by Comrie (1997).

Gardner and Dorling (2000) compared a regression-tree model with linear regression and MLP model to quantify the nonlinearity and interactions between predictor variables while modeling the hourly surface ozone concentrations. Their result clearly demonstrated the accuracy of MLP model and regression-tree model in capturing the underlying relationship between meteorological and temporal predictor variables and hourly ozone concentrations.

Bordignon et al. (2002) developed non-linear non-parametric models for short term forecasting of future maximum 1 hour and maximum 8 hour ground level ozone concentrations in Padova district in Northern Italy and later compared the results with additive model, regression tree models and MLP models to improve the developed model performance. Their result proved that the combination of boosting procedures (Freund and Schapire, 1997) and artificial neural networks has the capability to provide an improved short term forecast of ozone concentrations.

Zickus et al. (2002) compared four machine learning methods of different complexity: logistic regression, decision tree, multivariate adaptive regression splines and neural networks models in order to compare the variable selection and prediction performances of PM₁₀ concentrations in Helsinki, Finland and found superior forecasting performances of neural network and multivariate adaptive regression splines techniques.

Kukkonen et al. (2003) made an extensive evaluation of the neural network models to predict NO₂ and PM₁₀ concentrations and later compared it with the deterministic model for Helsinki area. Five neural network models, a linear statistical model and a deterministic model were compared and concluded that NN models can be useful and

fairly accurate tools of assessment in predicting NO_2 and PM_{10} concentrations in urban areas.

Chaloulakou et al. (2003) made a comparative assessment of neural networks and regression models for forecasting summertime ozone in the Athens basin at four representative stations showing different behavior. The performances indicated that the NNs provide better estimates of ozone concentrations at the monitoring sites while the more often used linear models were less efficient at accurately forecasting high ozone concentrations.

Chelani (2005) tested the concepts of chaotic systems theory to build feed-forward neural network model for predicting chaotic time series of inhalable particulate matter (PM_{10}) concentration in Delhi, India and found that neural networks model is capable of modeling the chaotic time series data. Chelani et al (2002) in their previous works compared a three layer neural network model with a hidden recurrent layer with the multivariate regression model to predict sulphur dioxide concentration at three sites in Delhi. Their results demonstrated that a neural network can be a better alternative to the multivariate regression model.

Hooyberghs et al. (2005) examined the feasibility of a statistical short-term forecasting tool for ambient PM_{10} concentrations in Belgium by developing a neural network model and found that day-to-day fluctuation of the PM_{10} concentration in Belgium is largely driven by meteorological conditions and to a lesser extent by changes in anthropogenic sources.

Ordiers et al. (2005) analyzed and bench marked a neural network model for short term $\text{PM}_{2.5}$ predictions in the central-south border region of the U.S. particularly in the area of El Paso, Texas and Ciudad Juarez in Chihuahua. They developed three different topologies of neural network: multilayer perceptron (MLP), radial basis function (RBF),

and square multilayer perceptron (SMLP) and compared the results with a persistence model and a linear regression model. Their analysis clearly demonstrated that the neural network approach not only outperformed the classical models but also showed fairly similar values among similar topologies. The results also reflected a more stable behavior than the so called classical models.

Recently Agirre-Basurko et al. (2006) presented a comparison between two MLP based models and one multiple linear regression model to forecast hourly ozone and nitrogen oxide level in Bilbao (Spain) using traffic variables, meteorological variables and ozone and NO₂ data as input variables. The performances of the model results were compared with persistence of levels and the observed values. The performance of MLP models, as expected, is found better than the multiple linear regression models.

Schlink et al. (2006) attempted to link two key aspects of ground level ozone problem: assessment of health effects and forecasting using 15 different statistical models in an inter-comparison study in 10 European regions. Their study (Schlink et al, 2006; Schlink et al, 2003) recommended that in operational air pollution forecasting, neural networks and generalized additive models have the capacity to handle the strong nonlinear associations between the atmospheric variables.

Athanasiadis et al. (2006) performed a comparative study of ozone forecasting for the Greater Athens Area (GAA) using conventional statistical methods (Linear regression, ARIMA and Principal component analysis) and data-driven classification algorithms such as NN, decision trees, conjunctive rules, support vector machines, decision tables and fuzzy lattice rules. Their study clearly showed that the performances of classification methods are far better than the so called conventional methods in terms of model performance and operational potential.

A neural network developed by Dutot et al. (2006) combined to form a neural classifier to forecast hourly maximum concentration of ozone in the central area of France has also shown great potential.

Elminir and Abdel-Galil (2006) applied NN trained with a back-propagation algorithm to predict daily PM_{10} , CO_2 , and NO_2 concentrations at 14 sites in Cairo, Egypt. Their results also showed that the NN with a single layer based on the standard BPT algorithm resulted in a very efficient model to forecast long term air pollutant concentrations in the study area with 96% prediction accuracy for PM_{10} .

Perez and Reyes (2006) developed an integrated NN model and later compared it to a linear and a persistent model to forecast the maxima of 24 hour average of PM_{10} concentrations 1 day in advance at 5 monitoring sites in Santiago, Chile. Their result proved the neural model to be more accurate than the linear models.

Slini et al. (2006) compared multilayer perceptron (MLP) model with classification and regression tree (CART), linear regression and principal component analysis to forecast daily PM_{10} concentration at Thessaloniki, Greece during 1994-2000. The results clearly indicated superior performances of CART and MLP model compared to the conventional models. The study further emphasized the importance of adequate and appropriate climatic data for accurate meteorology based forecasting.

Sousa et al. (2007) developed multiple linear regression and artificial neural network models based on principal components to predict ozone concentrations in Oporto, Portugal and later compared their performance with multiple linear regression and feed-forward neural networks based on the original data and also with principal component regression. Their result showed that the use of principal components as inputs improved both models prediction by reducing their complexity and eliminating data collinearity.

Hence it is well established that NN models, especially the static multilayer perceptron (MLP) model, are more capable of capturing the complex nonlinearity of the weather-pollutant relationships compared to other statistical models. However, in temporal problems, measurements from physical systems are no longer an independent set of input samples, but functions of time. To exploit the time series structure in the inputs, the neural network must have access to this time dimension. While the MLP models (also known as feed-forward neural networks) are popular in many application areas, they are not well suited for temporal sequence processing due to the lack of time delay and/or feedback connections necessary to provide a dynamic model. They can be used as pseudo-dynamic models only by using successively lagging multiple inputs based on correlation and mutual information analysis of the input data. There are however various types of neural networks that have internal memory structures which can store the past values of input variables through time and there are different ways of introducing ‘memory’ in a neural network in order to develop a temporal neural network. Time lagged feed-forward (TLFN) and recurrent networks (RNN) are two major groups of dynamic neural networks mostly used in time series forecasting (Coulibaly et al., 2001 a, b; Dibike and Coulibaly, 2006). In this study, three emergent dynamic neural network models are developed and their performances have been compared with widely used MLP model.

Chapter 4

Methodology

4.1 Temporal analysis of air quality

Artificial neural network models have proven to be very powerful and efficient methods for dealing with complex problems of associations, classification and prediction. This chapter deals with the methodologies of neural network architectures and interpolation techniques applied in the research.

4.1.1 Multilayer perceptron

Multilayer perceptrons (MLPs) constitute probably the most widely used network architecture and has been widely applied in atmospheric science (Gardner and Dorling, 1998; 1999; 2000; Ordieres et al., 2005). They are composed of a hierarchy of processing units organized in a series of two or more mutually exclusive sets of neurons or layers, as illustrated in Fig. 4.1, which is a model representing a nonlinear mapping between an input vector and an output vector. The nodes are connected by weights and output signals which are a function of the sum of the inputs to the node modified by a nonlinear transfer or activation function. The information flow in the network is restricted to a flow, layer by layer, from the input to the output, hence also called feed-forward network (Coulibaly et al, 2001b). The architecture of a multilayer perceptron although varies, but generally consists of several layers of neurons. The input layer only serves to pass the input to the network rather than performing any computation. The inputs and outputs of the multilayer perceptrons can be represented as single vectors. Such a network may have one or more hidden layers and finally an output layer. By selecting a suitable set of connection weights and transfer functions it has been shown that a multilayer perceptron

can approximate any smooth, measurable function between the input and output vectors (Gardner and Dorling, 1998).

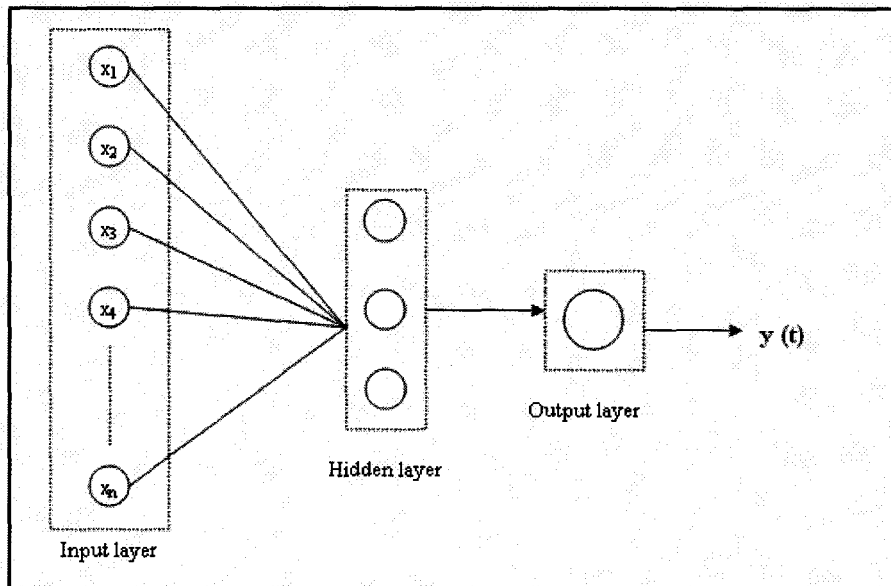


Fig.4.1: Multilayer perceptron neural network architecture

4.1.2 Time lagged feed-forward network

A time lagged feed-forward neural network (TLFN) is an extension of the standard MLP models which can be formulated by replacing the neurons in the input layer of an MLP with a memory structure, known as a tap delay line or a time delay line. The size of the memory structure (tap delay line) depends on the number of past samples that are needed to describe the input characteristics in time and it has to be determined on a case-by-case basis. TLFN uses delay-line processing elements, which implement memory by simply holding past samples of the input signal shown in Fig. 4.2.

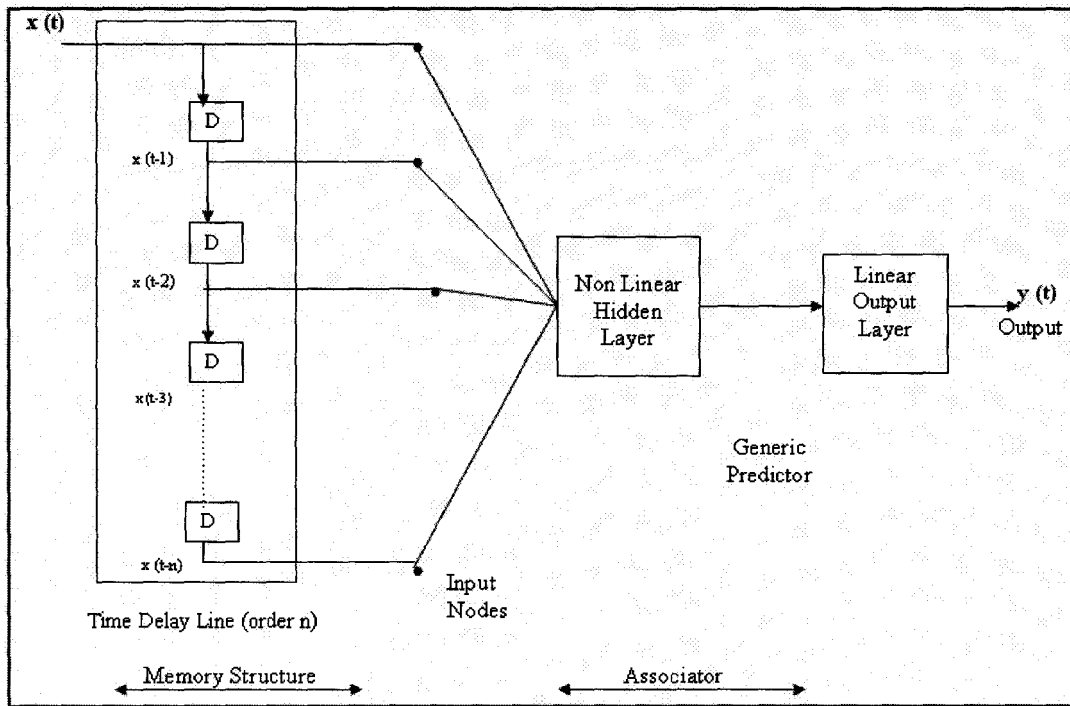


Fig. 4.2: Time lagged feed-forward neural network with tap delay memory structure (Coulibaly et al., 2001b)

The output of such a network with one hidden layer is given by (Dibike & Coulibaly, 2006):

$$y(n) = \phi_1 \left(\sum_{j=1}^m w_j y_j(n) + b_o \right) \quad (4.1)$$

$$= \phi_1 \left\{ \sum_{j=1}^m w_j \phi_2 \left[\sum_{l=0}^k w_{jl} x(n-l) + b_j \right] + b_o \right\} \quad (4.2)$$

where m is the size of the hidden layer, n is the time step, w_j is the weight vector for the connection between the hidden and output layers, w_{jl} is the weight matrix for the

connection between the input and hidden layers and φ_1 and φ_2 are transfer functions at the output and hidden layers, respectively, and b_j and b_o are additional network parameters (biases) to be determined during training of the networks with observed input/output data sets. For the case of multiple inputs (of size p), the tap delay line with memory depth k can be represented by:

$$X(n) = [x(n), x(n-1), \dots, x(n-k+1)] \quad (4.3)$$

$$X(n) = (x_1(n), x_2(n), \dots, x_p(n)), \quad (4.4)$$

where $x(n)$ represents the input pattern at time step n , $x_j(n)$ is an individual input at the n^{th} time step and $X(n)$ is the combined input to the processing elements at time step n . Such a delay line only ‘remembers’ k samples in the past. An interesting attribute of the TLFN is that the tap delay line at the input does not have any free parameters; therefore, the network can still be trained with the classical back propagation algorithm. The TLFN topology has been used effectively in nonlinear system identification and time series prediction (Coulibaly et al., 2001b).

4.1.3 Recurrent neural network

Depending on the architecture of feedback connections, the recurrent neural networks (RNN) can be of three types: the Jordan RNN (Jordan 1986) that has a feedback or recurrent connection from the output layer to its inputs; the locally RNN (Frasconi et al., 1992) which uses only local feedback, and the globally RNN (Elman, 1990) which has feedback connection from its hidden layer neurons back to the inputs. Regardless of the types, the important feature of RNN is that the feedback connections are applied through a context unit which consists of delay units. The RNN model used in this work is the basic Elman type RNN (Elman, 1990). Fig. 4.3 presents a typical diagram of the fully recurrent network. The network consists of four layers: the input layer, the hidden layer,

the context unit, each with n number of nodes and the output layer with one node. A common trial and error method is used to select the number of nodes because of the problem dependency of the network geometry. Each input unit is connected with every hidden unit, as is each context unit. Conversely, there are one-by-one downward connections between the hidden nodes and the context units leading to an equal number of hidden and context units. In fact, the downward connections allow the context units to store the outputs of the hidden nodes (i.e. internal states) at each time step; then the fully distributed upward links feed them back as additional inputs.

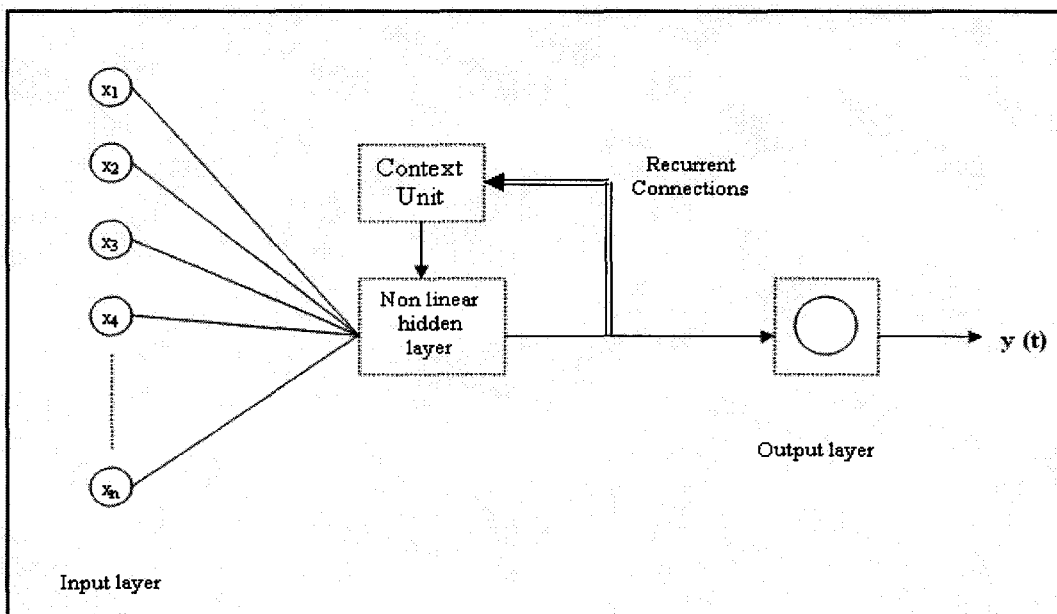


Fig.4.3: Fully recurrent neural network with feedback connection

Therefore the recurrent connections allow the hidden units to recycle the information over multiple time steps and thereby to discover temporal information contained in the sequential input and relevant to the target function (Coulibaly et al., 2001b). Thus the RNN has an inherent dynamic (or adaptive) memory provided by the context units in its recurrent connections. The output of the network depends not only on the connection weights and the current input signal but also on the previous states of the network, which can be shown by the following equations (Coulibaly et al., 2001b):

$$y_j = Ax'(t) \quad (4.5)$$

$$x'(t) = G[W_h x'(t-1) + W_{h_o} x(t-1)] \quad (4.6)$$

Where $x'(t)$ is the output of the hidden layer at time t given an input vector $x(t)$, $G(\cdot)$ denotes a logistic function characterizing the hidden nodes, the matrix W_h represents the weights of the h hidden nodes that are connected to the context units, W_{h_o} is the weight matrix of the hidden units connected to the input nodes, y_j is the output of the RNN assuming a linear output node j , and A represents the weight matrix of the output layer neurons connected to the hidden neurons. The Elman-style RNN is a state-space model since (4.6) performs the static estimation and (4.5) performs the evaluation. The major characteristics of the model are the interactions between the context units and the hidden nodes; One by one downward connection between the hidden nodes and the context units ultimately leads to an equal number of hidden and context units. Moreover, the upward connections between the context units and hidden nodes are distributed fully in a manner that each context unit stimulates all the hidden nodes (Coulibaly et al., 2001b). The context units receive the outputs of the hidden nodes through the downward connections and store them and the upward link feed them back again as additional input. In this way the information is recycled over multiple time steps and relevant information related to the predicted output are revealed. Hence the final output of the network depends both on the combination weights, current input signals and previous states of the network. Therefore a fully RNN can be suitable to the air pollution modeling where past information can also be vital as well as the current state.

According to Coulibaly et al. 2001b, a major difficulty, however, with RNN is the training complexity because the computation of $\nabla E(w)$, the gradient of the error E with respect to the weights, is not trivial since the error is not defined at a fixed point but rather is a function of the network temporal behavior.

4.1.4 Bayesian neural network

The Bayesian neural network model used in this work is developed by Khan and Coulibaly (2006). A Bayesian approach implements the conventional or standard learning process; but instead of a single set of weights it considers a probability distribution of weights. According to Khan and Coulibaly (2006), the process starts with a suitable prior distribution, $p(w)$, for the network parameters (weight and biases). Once the data D is observed, Bayes' theorem is used for deriving an expression of the posterior probability distribution for the weights, $p(w|D)$, as follows:

$$p(w|D) = \frac{p(D|w) p(w)}{p(D)} \quad (4.7)$$

where, $p(D|w)$ is the dataset likelihood function and the denominator, $p(D)$ is a normalizing factor, which can be obtained by integrating over the weight space as follows:

$$p(D) = \int p(D|w) p(w) dw \quad (4.8)$$

The left-hand side of (4.7) gives unity when integrated over all weight space. Once the posterior has been calculated, every type of inference is made by integrating over that distribution. Therefore, in implementing the Bayesian method, expressions for the posterior distribution, $p(w)$ and the likelihood function, $p(D|w)$ are needed. The prior distribution, $p(w)$, which is not related with data, can be expressed in terms of weight-

decay regularizer, $E_w = \frac{1}{2} \sum_{i=1}^w w_i^2$, where, W is the total number of weights and

biases in the network. A Gaussian prior is considered because it simplifies the total process.

Similarly, the likelihood function in Bayes' theorem (1), which depends on data, can be

expressed in terms of an error function, $E_D = \frac{1}{2} \sum_{n=1}^N (y^n(x^n; w) - t^n)^2$, where, x is

the input vector, t is the target value and $y(x; w)$ is the network output. Upon deriving the expressions for the prior and likelihood functions and using those expressions in (7), the posterior distribution of weights can be obtained. The posterior distribution over network weights provides a distribution about the outputs of the network. If a single-valued prediction is needed, the mean of the distribution is used and while the uncertainty about the prediction is needed the full predictive distribution is used to present the range of uncertainty about the prediction. The objective function in the Bayesian method corresponds to the inference of the posterior distribution of the network parameters. After defining the posterior distribution (objective function), the network is trained with a suitable optimization algorithm to maximize the posterior distribution $p(w|D)$. Thus the most probable value for the weight vector w_{MP} corresponds to the maximum of the posterior probability. Using the rules of conditional probability, the distribution of outputs, for a given input vector, x can be written in the form,

$$p(t|x, D) = \int p(t|x, w) p(w|D) dw \quad (4.9)$$

where $p(t|x, w)$ is simply the model for the distribution of noise on the target data for a fixed value of the weight vector w_{MP} , and $p(w|D)$ is the posterior distribution of weights. To make the integration analytically traceable in case of large datasets the posterior distribution $p(w|D)$ may be approximated to a Gaussian distribution (Walker, 1969; Khan and Coulibaly, 2006). So, equation 4.9 can be written as the following simplified form (Bishop 1995; Khan and Coulibaly, 2006):

$$p(t|x, D) = \frac{1}{(2\pi\sigma_t^2)^{1/2}} \exp\left(-\frac{(t - y(x; w_{MP}))^2}{2\sigma_t^2}\right) \quad (4.10)$$

The mean of this distribution can be given by $y(x; w_{MP})$ and the variance can be written as:

$$\sigma_t^2 = \frac{1}{\beta} + g^T A_h^{-1} g \quad (4.11)$$

Here β is a hyperparameter which is none other than the inverse variance of the noise model; g is the gradient of $y(x,w)$ with respect to the weights w evaluated at w_{MP} ; A_h is the Hessian matrix of the regularized error function. The standard deviation σ_t of the predictive distribution for the target t can be interpreted as an error bar on the mean value $y(x; w_{MP})$. This error bar represents the contributions from two sources, one is from the intrinsic noise on the target data represented by the first term of eqn. 4.11 and other one is from the width of the posterior distribution of the network weights that corresponds to the second term of eqn. 4.11. This can also be seen through Fig. 4.4. In Fig. 4.4 (a), the distribution of the network outputs in the Bayesian formalism is determined by the posterior distribution of the network weights $p(w|D)$ and the variance β^{-1} due to the intrinsic noise in the data. When the posterior distribution of the weights is very narrow in relation to the noise variance, the width of the distribution of network outputs is determined primarily by the noise. On the other hand if the posterior distribution of the network weight is broad comparing to the intrinsic noise in the data, the width of the network outputs is dominated by the distribution of network weights which is presented in Fig. 4.4 (b). In this way the Bayesian formalism allow to calculate the error bars. A more detailed description of BNN approach can be found in Khan and Coulibaly (2006).

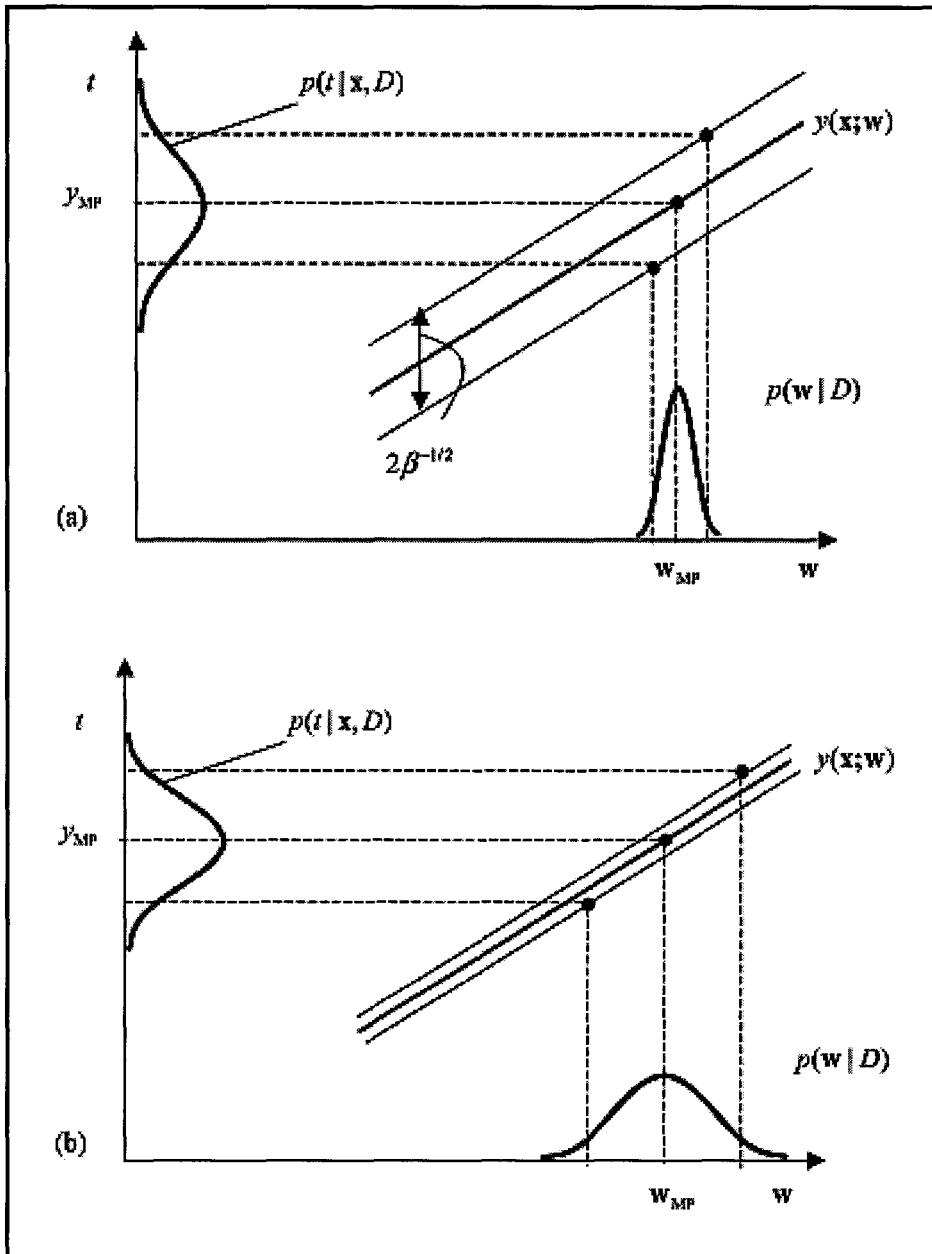


Fig. 4.4: Distribution of network outputs with error bars (Bishop, 1995): (a) width of the network outputs dominated by the noise of the data and (b) width of the network outputs dominated by the distribution of network weights

4.2 Spatial interpolation by kriging

Kriging is a class of statistical techniques for optimal spatial predictions based on statistical techniques such as autocorrelation, i.e. the statistical relationships among the measured points. It is called optimal because it is statistically unbiased i.e. on average the predicted and the observed value coincide and they minimize predicted mean-square error which is a measure of uncertainty or variability in the predicted values. So, these techniques not only have the capability of producing a prediction surface, but they can also provide some measure of the certainty or accuracy of the predictions. The basic objective of a kriging interpolation technique is to predict or interpolate the attribute values at points where sampling has not been done. The main statistical assumption of kriging is stationarity which means that statistical properties (means and variance) do not depend on the exact spatial locations, so the mean and variance of a variable at one location is equal to the mean and variance at another location. The correlation between any two locations depends only on the vector that separates them, not on their exact locations. Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. It fits a mathematical function to a certain number of points or all points within a specified radius, to determine the output value for each location.

Kriging weights the surrounding measured values to obtain predictions at unsampled locations. The general formula of kriging can be obtained from the weighted sum of the data:

$$\hat{Z}(s_o) = \sum_{i=1}^N \lambda_i Z(s_i) \quad (4.12)$$

where $Z(s_i)$ is the measured value at the i^{th} location, λ_i is an unknown weight for the measured value at the i^{th} location, s_o is the prediction location and N is the number of the measured values. The basic difference between kriging and the inverse distance

weighted (IDW) technique is in IDW, the weight λ_i depends only on the distance to the prediction location while in kriging, the weights are based not only on the distance between the measured points and the predicted location but also on the overall spatial arrangement (such as, auto correlation) of the measured points.

Kriging is a multi step process which includes exploratory statistical analysis of the data, variogram modeling, creating the surface and exploring a variance surface. The basic idea here is the weights created by the variograms, minimize the variance in the estimated values. Kriging uses the following steps to create a prediction surface:

1. It creates a variogram and covariate functions to estimate the statistical dependence (i.e. spatial autocorrelation) values that depend on the model of autocorrelation. This is also known as ‘model fitting’.
2. It predicts the unknown values i.e. prediction.

There are several kinds of kriging technique: universal kriging, ordinary kriging, co-kriging, indicator kriging, etc. The choice of a kriging interpolation technique depends on the characteristics of the data and the type of spatial model desired (Lefohn et al., 2006).

Ordinary kriging is the most widely used kriging model which assumes that the constant mean is unknown. On the other hand, universal kriging assumes that there is an overriding trend in the data which can be modeled by a deterministic function i.e. a polynomial. Universal kriging should be used only when there is a clear trend in data and a scientific reason is present to explain the reason of de-trending. However, the main issue related to ordinary kriging is whether the assumption of a constant mean is justified. It is a simple prediction model and has a good flexibility.

Chapter 5

Network Design

The neural network models used in the study were developed using the NeuroSolutions v4 (NeuroDimension Inc., Gainesville, Florida). In this particular work, ground level ozone (O_3), nitrogen dioxide (NO_2), total suspended particulates (TSP) and respirable particulates (PM_{10}) concentrations collected from 1994 to 2004 have been used to compare the performance of neural network models. Because of over 6 months of missing values, the year 1997 has not been included. From the 10 years of observed data from 1994 to 2004, the first five years (1994-1996, 1998-1999) are considered for constructing the models, one year data (2000) for cross-validation and the remaining four years (2001-2004) of dataset were used for testing the models.

5.1 Selection of best predictors

In case of both the neural network models and the Bayesian neural network model, selection of most important and relevant predictors is the most vital task in the modeling process. For this work the predictors were selected based on linear autocorrelation and partial autocorrelation analysis and nonlinear sensitivity analysis.

5.1.1 Correlation Analysis

A correlation plot can be defined as a graphical data analysis in order to determine the correlation between the lags for a single time series (autocorrelation) or lags for a single time series after removing the linear dependency of the intermediate lags (partial autocorrelation) or lags of two time series (cross correlation). The autocorrelation plots are a commonly used tool for checking the randomness in the dataset which can be determined by computing autocorrelations of the data value at varying time lags. In the

case of a random dataset the autocorrelation value would be zero for any or all previous time steps, while there should be significant non-zero autocorrelations for the non-random dataset (NIST/SEMATECH, 2006).

Autocorrelation in fact, is the correlation of a dataset with itself which is offset by n -values. In mathematical form it can be expressed as:

$$\hat{\rho}_k = \frac{\frac{1}{n} \sum_{t=k}^{N-(k+1)} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\frac{1}{n} \sum_{t=0}^{N-1} (y_t - \bar{y})^2} \quad (5.1)$$

where N is the number of observations, y_t is observation at any time t and \bar{y} is the mean observation.

Partial autocorrelation plots, on the other hand, determine the correlation without any dependency on previous lags. It is useful to identify the order of a model. If the sample autocorrelation plot indicates that there is good correlation between the past and present values of a variable then a partial autocorrelation plot is examined to help identify the order of fairly correlated time points. If $\phi_j(k)$ is the j^{th} autoregressive coefficient in an autoregressive model so that $\phi_k(k)$ is the last coefficient, the partial autocorrelation equation, given by Yule-Walker can be expressed as:

$$\begin{aligned} \rho_j &= \phi_1(k)\rho_{j-1} + \phi_2(k)\rho_{j-2} + \dots + \phi_k(k)\rho_{j-k} \\ j &= 1, \dots, k \end{aligned} \quad (5.2)$$

where $\phi_k(k)$ is the partial autocorrelation function.

For this study, at first, correlation between the historical values of each predictor and predictand pollutants was examined to get an initial idea about the important lags. Then the partial autocorrelation (PACF) analysis was performed for each of the input and output variables to identify range of significant lags. From the analysis it was found that while considering bi-hourly values, lags up to 14 were approximately important for input meteorological variables and pollutants such as O₃ and NO₂. However, in case of 6 day pollutant values, lags up to 8 steps were significant. So these corresponding lags were selected for sensitivity analysis using a TLFN model to finally identify the most significant input variables.

5.1.2 Sensitivity analysis

The sensitivity analysis is a measure of the relative importance among the predictors (inputs of the neural network) which calculates the variation of the output variables with the variation of inputs. The basic idea is, the changes in the outputs even with a slight change in input variables are calculated. Each input is varied $\pm n$ times its standard deviation while keeping others fixed about their mean and the network output is calculated for a specific number of steps above and below the mean. The neural network measures the relative sensitivity, which is the ratio between the standard deviation of the output and the standard deviation of the input, which as a result, gives the relative importance of each input. In this study, initially sensitivity analysis was performed using TLFN and RNN models. The results however, showed similar sensitivity results. Hence finally only TLFN model was considered for further analysis.

The important lags selected by sensitivity analysis for the pollutant monitoring stations are shown in Table 5.1 through 5.4. The Tables present the most relevant identified variables for each station. Table 5.1 gives the details of selected input variables for ozone in three ozone monitoring sites. A sensitivity value of 0.6 has been set as the lower limit to consider the predictors to be ‘significant’. Only predictors equal to or greater than sensitivity value of 0.6 were considered for O₃ and NO₂. From Table 5.1 it is clear that

variable ‘wind direction’ along with ‘wind speed’ at both Hamilton Airport and Burlington Piers sites have proven to be the two most dominant factors for pollutant dispersion. They show significant influence up to lag 14 for wind direction and lag 10 for wind speed which means wind speed and wind direction influence the pollutant dispersion up to 28 hours. Relative humidity, although not as significant as wind speed and direction, has some influence on the ozone. Average dry bulb temperature however, has not shown a higher sensitivity here. This was expected to some extent given that ground level ozone is typically more dependent on maximum temperature rather than the average temperature. The historical values of ozone itself shows high sensitivity up to lag 4. Similar results are shown for NO₂ (Table 5.2).

Table 5.1: Selected input variables for O₃ based on sensitivity analysis

Variables	Selected lags for ground level O ₃		
	Hamilton Downtown	Hamilton Mountain	Hamilton West
HA_WS	10	0, 1	6
HA_WD	2, 3, 4, 6, 11, 12, 13	1, 4, 5, 7, 11, 12, 13, 14	0, 2, 4, 5, 6, 9, 12, 13, 14
HA_RH	-	-	-
HA_Temp	-	-	-
BP_WS	0, 3, 4, 6, 7, 8, 11	1, 2, 3, 5, 7, 9, 10, 12	6, 12
BP_WD	1, 2, 3, 4, 5, 8, 9, 11, 12, 13, 14	2, 5, 6, 7, 9, 10, 11, 14	0, 1, 2, 3, 4, 5, 8, 9, 12, 13
BP_RH	6, 8	-	-
BP_Temp	-	-	-
O ₃	1	1	1, 4

Legend: HA: Hamilton Airport BP: Burlington Piers

Table 5.2: Selected input variables for NO₂ based on sensitivity analysis

Variables	Selected lags for NO ₂	
	Hamilton Downtown	Hamilton West
HA_WS	8, 10	1
HA_WD	2, 3, 4, 9, 10, 11, 12, 13	1, 2, 3, 7, 8, 12
HA_RH	-	-
HA_Temp	-	-
BP_WS	0, 1, 6, 8, 10, 11, 12, 14	9, 11
BP_WD	1, 6, 11, 13	1, 7, 10
BP_RH	11	-
BP_Temp	-	-
NO ₂	1	1

Legend: HA: Hamilton Airport BP: Burlington Piers

In order to get the best possible input variables for total suspended particulates (TSP) and PM₁₀ pollutants, a different approach was followed. The selection of the screened variables were performed based on three cases: case 1 deals with the meteorological variables only as input, case 2 includes the nearby stations' pollutant values as predictors and finally case 3 comprises of a combination of meteorology and nearby station pollutant values as input for the target station. For TSP, initially sensitivity analysis on three stations: 29000, 29025 and 29102 was performed using TLFN model and it was found that except station 29102, case 2 provided better results for the remaining two stations. The reason behind the different result among the stations may be because of the larger distance of 29102 than the other 4 stations, so when considering nearby stations as input, it did not give a good result. Due to the locations of remaining 4 stations, located within shorter distances between themselves, it is expected that their response to the meteorology and to themselves would be same. Hence the final screening was performed for 29113 and 29114 based on the results of 29000 and 29025. Table 5.3 represents the screened variables selected from sensitivity analysis for TSP. From the table it is seen that lagged concentrations from station 29114 were the most dominant variables for site 29000 and 29025. On the other hand, lags from all sites contributed for the pollutant observation at 29114. Where in case of site 29102, meteorological variables predominated; wind speed from Hamilton Airport, wind directions and mean temperature

from both Hamilton Airport and Burlington Piers were more relevant. Relative humidity and maximum temperature of these two stations are also found to influence the TSP levels at this station. The results for PM_{10} , presented in Table 5.4 were similar as TSP except for site 29102 where combined effect of meteorology and nearby sites (wind speed and direction from Hamilton Airport and Burlington Piers and maximum temperature of Hamilton Airport along with pollutant values of site 300) has been noticed.

Table 5.3: Selected input variables for TSP based on sensitivity analysis

Variables	Selected lags for TSP				
	29000	29025	29102	29113	29114
<i>Case-1: Meteorological variables</i>					
HA_WS	3	0	0, 1, 2, 5, 6	-	
HA_WD	0, 3	0, 4, 6	0, 2	-	
HA_RH	-		0	-	
HA_Tmax	-	0	-	-	
HA_Tmin	6		-	*	*
HA_Tmean	4	3, 6	1, 2	-	
HA_Prec	-	-	-	-	
BP_WS	3	2, 6	-	-	
BP_WD	6	0, 4, 6	0, 1, 3	-	
BP_RH	-		0	-	
BP_Tmax	-		5	-	
BP_Tmin	-	2, 3	-	-	
BP_Tmean	4	6	0, 2, 5	-	
<i>Case-2: Other stations</i>					
29000	-	0, 2	1	0, 7	0, 7, 8
29025	0, 7	-	1, 5	0	0, 6, 8
29102	0, 7	6	-	0	-
29113	0	-	0	1	-
29114	0, 5, 6, 8	0, 3, 4	0, 1, 6	5	3, 8
<i>Case-3: Meteorological variables and stations</i>					
HA_WS	2, 8	2	-		
HA_WD	1, 4, 5	1, 4	0		
HA_Tmax	-	-	4, 7		
HA_Tmin	-	0			
HA_Tmean	0		5		
BP_WS	-		1		
BP_WD	-	1, 6	8		
BP_Tmax	-	-	1		
BP_Tmin	-	7	1		
BP_Tmean	5	-	-		
29000	-	-	-		
29102	-	-	-		
29113	-	-	-		
29114	-	-	-		

Legend: HA: Hamilton Airport

BP: Burlington Piers

Table 5.4: Selected input variables for PM₁₀ based on sensitivity analysis

Variables	Selected lags for PM ₁₀		
	300	302	313
<i>Case-1: Meteorological variables</i>			
HA_WS	1, 2	0	
HA_WD	2, 4	0, 2, 4, 6	
HA_RH	-	-	
HA_Tmax	-	6	
HA_Tmin	2		*
HA_Tmean	-	-	
HA_Prec	5	-	
BP_WS	-	-	
BP_WD	4	0, 3	
BP_RH	-	-	
BP_Tmax	-	7	
BP_Tmin	-	-	
BP_Tmean	0	-	
<i>Case-2: Other stations</i>			
300	2, 4, 5, 6, 8	0, 3, 4, 5	0, 2, 6, 8
302	0	-	1, 6, 7
313	0, 2	0, 5, 6, 7	1, 6
<i>Case-3: Meteorological variables and stations</i>			
HA_WS	0, 1, 2, 6	0	
HA_WD	0, 5	0, 1	
HA_RH	-	-	
HA_Tmax	-	7	
HA_Tmin	5, 6	-	
HA_Tmean	-	-	
HA_Prec	-	-	*
BP_WS	-	2	
BP_WD	7, 8	0, 8	
BP_RH	-	-	
BP_Tmax	2	-	
BP_Tmin	-	-	
BP_Tmean	4	-	
300	-	0	
302	-	-	
313	0	-	

Legend: HA: Hamilton Airport

BP: Burlington Piers

5.2 Model setup

Selection of an appropriate architecture of any neural network model is a prerequisite behind its successful use since the structure directly influences the computational complexity and generalization capability of a model. A more complex than necessary model can over-train the model while a too-simple model with fewer numbers of nodes than needed may not be able to learn from the data successfully. Because of the absence of a standard methodology of selecting an appropriate network, a trial-and-error procedure has been applied to get the optimal model parameters.

For each station, same input variables identified from sensitivity analysis were used for all four models in order to compare model performance. The performances of the models were assessed based on the model performance statistics (root mean square error (RMSE) and correlation coefficient (r), etc.) generated directly by the model. A detailed description of the model performance statistics have been given in the following section. Except BNN, the comparison of MLP, TLFN and RNN model parameters gave similar results. In this study, trial-and-error approach was carried out with the screened variables by varying model parameters and the best ones were selected by comparing model performance until the optimum network was achieved. Table 5.5 through 5.8 presents a comparative result of the model parameters. Out of the 9 parameters tested in the NN models, processing element (PEs) or number of nodes in each layer is the most important one as the number of PEs directly affects the overall computing power of the network; hence, it should be chosen based on the complexity of the input-output data. It is important to choose a minimum number of processing elements for the dataset which will give the least possible error with least computing time and model complexity. Moreover, a good generalization capability of the NN model depends on choosing the appropriate minimum number of processing elements needed for the dataset. The second important parameter is epoch or number of iterations of the model. The ‘depth in sample’ parameter represents the number of taps or delays within the memory structure of the network. A

minimum number of epochs and depths are necessary to get optimum results while higher than necessary can increase the computational time particularly in case of large dataset. The learning rule, also known as gradient search is a term by which the correction term is specified. It is used to calculate weight updates. Once a learning rule is set, the rate at which the learning should be performed also has to be specified. An inadequate learning rate increases computational time while a higher rate makes weights unstable. Delta bar delta algorithm is an improved version of the back-propagation network. Unlike back propagation, delta bar delta algorithm uses a learning method where each weight has its own self-adapting coefficient. It does not use the momentum factor of the BP algorithm. The essence of the rule is to use past calculated error values for each weight to infer future calculated error values, hence by knowing the probable errors, the system takes intelligent steps in adjusting the weights. Furthermore, each connection weight has its individual learning rate which vary over time based on the current error information found with standard back-propagation; hence more degree of freedom is achieved which reduced the convergence time. Activation functions also referred to as transfer function describes the non-linearity of the hidden layers that give the neural network model an ability to learn difficult problem. A brief description of the activation functions is presented in Table A.2 (Appendix). In order to obtain predictions less sensitive to the initial conditions, 10 distinct runs are performed using the optimal parameters and the results from the best run achieved are taken as final result.

In BNN model, a 2 layer MLP network is used with the same set of input variables used in other NN models. The BNN network consists of one hidden layer, with tangent hyperbolic activation function and one output layer with linear processing unit (Khan and Coulibaly, 2006). The parameters of the BNN model, which runs in the MatLab environment, are quite different. Unlike other NN models, the initialization of parameters in BNN is performed using a distribution of parameters. The initial values of weights and biases can be achieved from a Gaussian prior distribution of zero mean and inverse variance α , also known as regularization coefficient or prior hyperparameter. Gaussian

prior has been selected in order to favor small values for the network weights because a network with large weights will usually give rise to a mapping with large curvature (Nabney, 2004; Khan and Coulibaly, 2006). Moreover, it is less complex in terms of computational simplicity. A single initial value has been chosen for both hidden and output layer weights for prior hyperparameter α . An error model for the data likelihood function is needed to define the objective function. So it is approximated that the target data is generated from a smooth function with additive zero-mean Gaussian noise. Like α , a single initial value for both hidden and output layer weights are also chosen for hyperparameter β . Once the prior and likelihood functions are defined, the objective function is set as posterior distribution of weights. Next the network training is performed by trial and error and the network weights are being optimized using scaled conjugate gradient optimization technique to get the most probable weights by maximizing the posterior distribution of weights $p(W|D)$. After the network has been trained, the predictions are performed using eqn. 4.10 where the posterior distribution is assumed Gaussian. Error bars are calculated using eqn. 4.11. The 95% confidence interval of the mean output $y(x;w_{MP})$ have been estimated by adding and subtracting 2σ from $y(x;w_{MP})$ (Khan and Coulibaly, 2006).

Both for O_3 and NO_2 , out of 11 years bi-hourly dataset, 6 years (1994-1999) data were used for model calibration and one year (2000) data was used for cross-validation. The remaining 4 years data (2001-2005) was used for testing. On the other hand, for TSP and PM_{10} , the length of dataset is quite small because the pollutant values are measured once a week at an interval of 6 days. In this case, 8 years (1994-2001) data were used for calibrating the model and one year data (2002) for cross-validation. The remaining 2 year (2003-2004) dataset were used for testing the model. The reason behind using cross-validation dataset is to prevent overtraining when the cross-validation error starts rising. In this way the best weights of the network are saved automatically at the point when the cross-validation error goes to the lowest point. The architectural description of the models selected for the O_3 and NO_2 monitoring stations are presented in Tables 5.5 and

5.6. From the Table 5.5 it can be seen that for BNN and RNN the 8 and 10 hidden nodes (PEs) are most appropriate, while for TLFN and MLP, optimum number of nodes are 14 and 17 respectively for site 29000. Similar results have been achieved for sites 29114 and 29118. It is seen that overall MLP requires higher number of nodes in its hidden layer than remaining three models. These optimum nodes are achieved by setting different PEs ranging from 2 to 40 keeping other parameters the same. This is one of the distinct and important differences between the models. Higher number of processing elements clearly indicates that the computational cost of BNN, TLFN and RNN model is higher than the MLP model. For the TLFN and RNN models, different lengths of input delays varying from 1 to 10 have been tested with the optimum number of nodes and it is found that in most cases an input delay of 4 was appropriate. Even though it is seen that the number of hidden nodes and input delays in RNN are slightly lower than the TLFN model, still it needs higher computing time than TLFN because of the recurrent connections. Number of iterations or epochs that has been tested includes 1000, 2000 and 3000 separately and it is seen that MLP models need less number of epoch to generate optimum result than other 3 models where in all cases, 2000 epoch was most appropriate. In all cases delta bar delta algorithm proved most appropriate. The NO₂ model performance results have been presented in Table 5.6. It has been found that with slightly higher number of hidden nodes than RNN and MLP, TLFN needs less number of iterations and input delays to produce its optimum results. The number of nodes required is higher in BNN than other three models. In case of TSP and PM₁₀ results, the number of hidden nodes in RNN is generally higher than other models; number of iterations is almost same in all models except for site 29025 where TLFN needed 3000 iterations to produce best results. Similar to O₃ results, PM₁₀ results presented in Table 5.8 show that MLP needed more hidden nodes than other models except for site 302 where RNN came up to be more complex.

Table 5.5: Best model structure for O₃

Stations	Model parameters	NN models		
		TLFN	RNN	MLP
29000	Processing element	14	10	17
	Epochs	2000	2000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	Gamma	-
	Depth in samples	4	4	-
	HL transfer function	Sigmoid	Sigmoid	Tan hyperbolic
	OL transfer function	Linear	Axon	Bias
	Stopping criteria	Cross validation	Cross validation	Cross validation
29114	Processing element	7	8	12
	Epochs	2000	2000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	Gamma	-
	Depth in samples	10	4	-
	HL transfer function	Sigmoid	Sigmoid	Tan hyperbolic
	OL transfer function	Axon	Axon	Bias
	Stopping criteria	Cross validation	Cross validation	Cross validation
29118	Processing element	13	13	17
	Epochs	2500	2000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	Gamma	-
	Depth in samples	10	4	-
	HL transfer function	Sigmoid	Sigmoid	Tan hyperbolic
	OL transfer function	Linear	Axon	Axon
	Stopping criteria	Cross validation	Cross validation	Cross validation

Legends: TLFN: Time lagged feed-forward network

RNN: Recurrent neural network

MLP: Multilayer perceptron

TDNN: Time delay neural network

Station	BNN model parameters				
	Nhidden	Option 14	Nouter	Alpha	Beta
29000	8	2000	15	0.01	50
29114	14	2000	15	0.01	40
29118	12	1000	10	0.015	50

Legends: Nhidden: No. of processing units

Nouter: No. of loops

Option 14: No. of iteration in each loop

Alpha: Initial prior hyperparameter

Beta: Initial noise parameter

Table 5.6: Best model structure for NO₂

Stations	Model Parameters	NN Models		
		TLFN	RNN	MLP
29000	Processing element	11	7	5
	Epochs	1000	1000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	Gamma	-
	Depth in samples	2	4	-
	HL transfer function	Sigmoid	Sigmoid	Sigmoid
	OL transfer function	Bias	Axon	Linear
	Stopping criteria	Cross validation	Cross validation	Cross validation
29118	PE	5	5	5
	Epochs	1000	2000	2000
	Learning Rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input Transfer Function	TDNN	Gamma	-
	Depth in Samples	4	4	-
	HL Transfer Function	Sigmoid	Sigmoid	Sigmoid
	OL Transfer Function	Linear	Axon	Linear
	Stopping Criteria	Cross validation	Cross validation	Cross validation

Legends: TLFN: Time lagged feed-forward network RNN: Recurrent neural network
MLP: Multilayer Perceptron TDNN: Time delay neural network

Station	BNN model parameters				
	Nhidden	Option 14	Nouter	Alpha	Beta
29000	16	2000	15	0.01	50
29118	16	1000	15	0.01	50

Legend: Nhidden: No. of processing units Nouter: No. of loops
Option 14: No. of iteration in each loop Alpha: Initial prior hyperparameter
Beta: Initial noise parameter

Table 5.7: Best model structure for TSP

Stations	Model parameters	NN models		
		TLFN	RNN	MLP
29000	Processing elements	5	20	5
	Epochs	2000	2000	2000
	Learning rule	Delta bar delta	Delta bar delta	Conjugate gradient
	Input transfer function	TDNN	TDNN	-
	Depth in samples	4	10	-
	HL transfer function	Tan hyperbolic	Sigmoid	Tan hyperbolic
	OL transfer function	Linear	Linear	Linear
29025	Processing elements	15	8	17
	Epochs	3000	2000	2000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	Gamma	TDNN	-
	Depth in samples	2	4	-
	HL transfer function	Tan hyperbolic	Sigmoid	Tan hyperbolic
	OL transfer function	Axon	Axon	Bias
29113	Processing elements	34	15	8
	Epochs	1000	1000	1000
	Learning rule	Conjugate gradient	Conjugate gradient	Momentum
	Input transfer function	TDNN	TDNN	-
	Depth in samples	10	4	-
	HL transfer function	Sigmoid	Sigmoid	Tan hyperbolic
	OL transfer function	Linear	Linear	Linear
29114	Processing elements	10	18	6
	Epochs	2000	2000	1000
	Learning rule	Momentum	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	TDNN	-
	Depth in samples	4	10	-
	HL transfer function	Sigmoid	Sigmoid	Sigmoid
	OL transfer function	Bias	Linear	Linear
29102	Processing elements	8	28	6
	Epochs	1000	1000	2000
	Learning rule	Delta bar delta	Conjugate gradient	Conjugate gradient
	Input transfer function	TDNN	Gamma	-
	Depth in samples	4	2	-
	HL transfer function	Sigmoid	Sigmoid	Sigmoid
	OL transfer function	Bias	Linear	Axon

Legends: TLFN: Time lagged feed-forward network
MLP: Multilayer perceptron

RNN: Recurrent neural network
TDNN: Time delay neural network

Station	BNN model parameters				
	Nhidden	Option 14	Nouter	Alpha	Beta
29000	7	2000	15	0.01	50
29025	10	2000	15	0.01	50
29113	10	2000	15	0.01	50
29114	15	1000	15	0.01	50
29102	20	2000	15	0.01	50

Legend: Nhidden: No. of processing units
Option14:No.ofiterationineach loop
Alpha: Initial prior hyperparameter

Nouter: No. of loops
Beta: Initial noise parameter

Table 5.8: Best model structure for PM₁₀

Stations	Model parameters	NN models		
		TLFN	RNN	MLP
300	Processing elements	10	8	30
	Epochs	1000	1000	1000
	Learning rule	Delta bar delta	Delta bar delta	Momentum
	Input transfer function	TDNN	Gamma	-
	Depth in samples	10	4	-
	HL transfer function	Tan hyperbolic	Tan hyperbolic	Sigmoid
	OL transfer function	Linear	Linear	Linear
313	Processing elements	6	10	18
	Epochs	1000	2000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	TDNN	-
	Depth in samples	4	4	-
	HL transfer function	Tan hyperbolic	Sigmoid	Sigmoid
	OL transfer function	Linear	Linear	Linear
302	Processing elements	8	26	4
	Epochs	2000	2000	1000
	Learning rule	Delta bar delta	Delta bar delta	Delta bar delta
	Input transfer function	TDNN	TDNN	-
	Depth in samples	4	4	-
	HL transfer function	Sigmoid	Sigmoid	Sigmoid
	OL transfer function	Linear	Linear	Linear

Legends: TLFN: Time lagged feed-forward network
MLP: Multilayer Perceptron
RNN: Recurrent neural network
TDNN: Time delay neural network

Station	BNN model parameters				
	Nhidden	Option 14	Nouter	Alpha	Beta
300	7	1000	15	0.005	50
313	8	2000	15	0.01	50
302	8	2000	15	0.005	50

Legend: Nhidden: No. of processing units
Option14: No. of iteration in each loop
Alpha: Initial prior hyperparameter

Nouter: No. of loops
Beta: Initial noise parameter

5.3 Model performance evaluation

Many model performance statistics are available in order to assess the accuracy of the estimates. For this particular work, the model performance and forecasting results were compared by a set of five statistics. A brief description of these statistics is given below:

The **Root mean square error (RMSE)** which is the square root of the differences between the observations C_o and predicted values C_p :

$$RMSE = \left(\frac{1}{N} \sum_{i=1}^N (C_{P_i} - C_{O_i})^2 \right)^{1/2}$$

where N is the number of observations, C_{O_i} and C_{P_i} are observed and predicted values respectively. The mean square errors provide a general illustration of the relevancy of the simulated values by giving a global goodness to fit by including errors and biases in the calculation. The lower the RMSE value, the better the model.

RMSE, however, doesn't necessarily reflect whether the two sets of data move in the same direction. For instance, by simply scaling the network output, we can change the MSE without changing the directionality of the data. This limitation can be overcome by introducing a second index, correlation coefficient, r .

The **correlation coefficient (r)** between an observed value C_{O_i} and a desired model output C_{P_i} is defined by:

$$r = \frac{\frac{1}{N} \sum_{i=1}^N [(C_{P_i} - \bar{C}_{P_i})(C_{O_i} - \bar{C}_{O_i})]}{\sqrt{\frac{1}{N} \sum_{i=1}^N (C_{P_i} - \bar{C}_{P_i})^2} \sqrt{\frac{1}{N} \sum_{i=1}^N (C_{O_i} - \bar{C}_{O_i})^2}}$$

Where N is the number of observations and \bar{C}_{O_i} and \bar{C}_{P_i} are the mean observed and predicted values respectively. This statistic provides a measure of the prediction ability of a model and it is an important tool for comparing two models as it is independent of the scale of data. The r value can range from -1 (perfect negative correlation) to 1 (perfect positive correlation) through 0 where 0 means no correlation. An r value of 0.9 and above is very satisfactory, 0.8 to 0.9 presents a fairly good model but below 0.7 is considered unsatisfactory.

The ***coefficient of determination (R^2)*** which is simply the square of the coefficient of correlation, assesses the strength of an association between two variables. It is also a measure of the ability of a model to predict the concentrations, which are different from mean. Moreover, it provides a useful comparison between the models since it is independent of the scale of data. It lies between zero and unity; the closer to unity, the greater the explanatory power.

The ***normalized mean squared error (NMSE)*** is another version of the mean square error which is normalized with the object of establishing comparisons among different models (Agirre-Basurko et al, 2006).

$$NMSE = \frac{MSE}{Var(Co)}$$

$$= \frac{\frac{1}{N} \sum_{i=1}^N (C_{P_i} - C_{O_i})^2}{\frac{1}{N} \sum_{i=1}^N (C_{O_i} - \bar{C}_{O_i})^2}$$

The ***mean absolute error (MAE)*** is a linear score which means that all the individual differences are weighted equally in the average. In short, it measures the average

magnitude of the errors in predicted dataset without considering their direction. It can be expressed as:

$$MAE = \frac{1}{N} \sum_{i=1}^N | C_{O_i} - C_{P_i} |$$

For a perfect fit, $C_{O_i} = C_{P_i}$ and hence MAE becomes zero. So the MAE ranges from 0 to infinity where 0 corresponds to the ideal condition which in particular permits to compare the appropriateness of using the models.

The *relative bias (RB)* provides a measure of the magnitude of bias between the observed and target data. It can be expressed as:

$$RB = \frac{\frac{1}{N} \sum_{i=1}^N (C_{P_i} - C_{O_i})}{\bar{C}_O}$$

Chapter 6

Ozone and Nitrogen Dioxide Forecasting Results

As described in chapter 5, first 6 years (1994-1999) of the observed meteorological data (as predictors) and the historical O₃ and NO₂ data (as predictand) has been used for model calibration at each monitoring stations. In order to prevent overtraining of the models, 1 year (2000) data has been selected for cross-validation. The remaining 4 years (2001-2004) of data are being used to test the model. After good statistical agreement between the observed and simulated values has been achieved with the training data, the models are then used to perform 12 steps i.e. 24 hour ahead forecasting of O₃ and NO₂ level at the monitoring stations.

Firstly, the performances of the ozone models are compared using standard statistical model performance measures. Scatter plots are then drawn in order to assess the relationship or association between the observed and predicted concentrations. A comparison of the seasonal performances is investigated and further analysis is done using seasonal model to improve their performance. Further analysis has been performed by including land use type around the stations as logical inputs to improve model results. The analysis of the NO₂ models also follow the same steps but using only the best procedure. The discussion emphasized the testing results only as they provide real evaluation information about model performance owing to the use of independent datasets from calibration.

6.1 Ozone forecasting results

6.1.1 Model forecasting performances

After modeling has been done for the three ozone monitoring stations in Hamilton, the performances of each model were compared. Table 6.1 presents the overall

performances of the models to forecast ozone at three ozone monitoring stations during testing period. A detailed description of the forecasting performances of the models is provided in Table A.3 (Appendix) for a comparative assessment of the models. The most efficient model should have least root mean square error (RMSE), mean absolute error (MAE), normalized mean square error (NMSE) and highest correlation coefficient (r) and coefficient of determination (R^2). It has been seen that the performances of all the models deteriorated with time as expected. Overall performances indicate that all four models performed satisfactorily up to 6 steps (12 hour) ahead with generally slight variations in between themselves. It appeared that TLFN model outperformed other 3 models in terms of its RMSE and r values up to 4 hour ahead followed by RNN model. RNN model then degraded at a slower rate than the TLFN and finally has shown better results than TLFN from 6 to 12 hour ahead forecasts.

Firstly, the results of 2 hour and 4 hour ahead forecasting show similar forecasting skill for all three stations. The TLFN model resulted in a lower $RMSE$, r , R^2 and MAE values than the rest three models. The performance of RNN, BNN and MLP model is quite similar in terms of r and R^2 values while the $RMSE$ and MAE results showed that RNN is slightly superior to the static MLP model. The $NMSE$ values obtained shows a bit different results with lowest $NMSE$ value for BNN model. Overall, all models have showed similar performances; TLFN and RNN had *coefficient of correlation* r (0.91-0.93) values slightly higher than 0.90 compared to BNN and MLP (0.89-0.92) models for one step ahead forecasting and R^2 values higher than 0.80 which clearly demonstrates the efficiency of the models for that forecasting period. The r value for the forecasting period $t+2$ is also satisfactory. Further analysis has been done to investigate the percent of improvement of the models in terms of RMSE values which has been presented in Table 6.1. When moving from 2 hour to 4 hour ahead forecasting, large (around 50%) drops in model performance has been observed in terms of RMSE values except RNN model performance at site 29118 where the RMSE value increased only 2%. This means for 2

step (4 hour) ahead forecasting at site 29118, RNN has higher performance than remaining three models.

Unlike the forecasting results of first two periods, the 3 to 6 step ahead forecasting results demonstrate slightly better performances of RNN model over TLFN and other models. The lowest RMSE, NMSE and MAE values were obtained by the RNN model for downtown station and mountain station. In case of Hamilton Mountain, TLFN experienced lower errors than other three models. The greatest values of *coefficient of correlation* r were also obtained for RNN model which clearly shows that RNN model worked better than other models during this time frame. Similar to two other stations, all models showed variation in their performances up to 6 step (12 hour) ahead forecasting; the values beyond this point, however remained same. The lower change (below 10%) in the RMSE values while moving from 4 (8 hour) to 6 (12 hour) step which also supports this explanation. The graphs in Fig. 6.1 clearly indicate that the peak performances of these four models can be obtained up to 6 steps ahead i.e. 12 hours. The graphs also reveal the better performances of TLFN and RNN models over static MLP and BNN models. These results further suggest that the inclusion of time delay and/or adaptive memory (context unit) in MLP have the capacity to improve the results obtained from conventional static neural network (MLP in this case). These performances, however indicate that RNN model has the best generalization performance and suggest that the relationship between ozone and the meteorology can better be represented using these predictor variables still with room for improvement at three ozone monitoring sites in Hamilton.

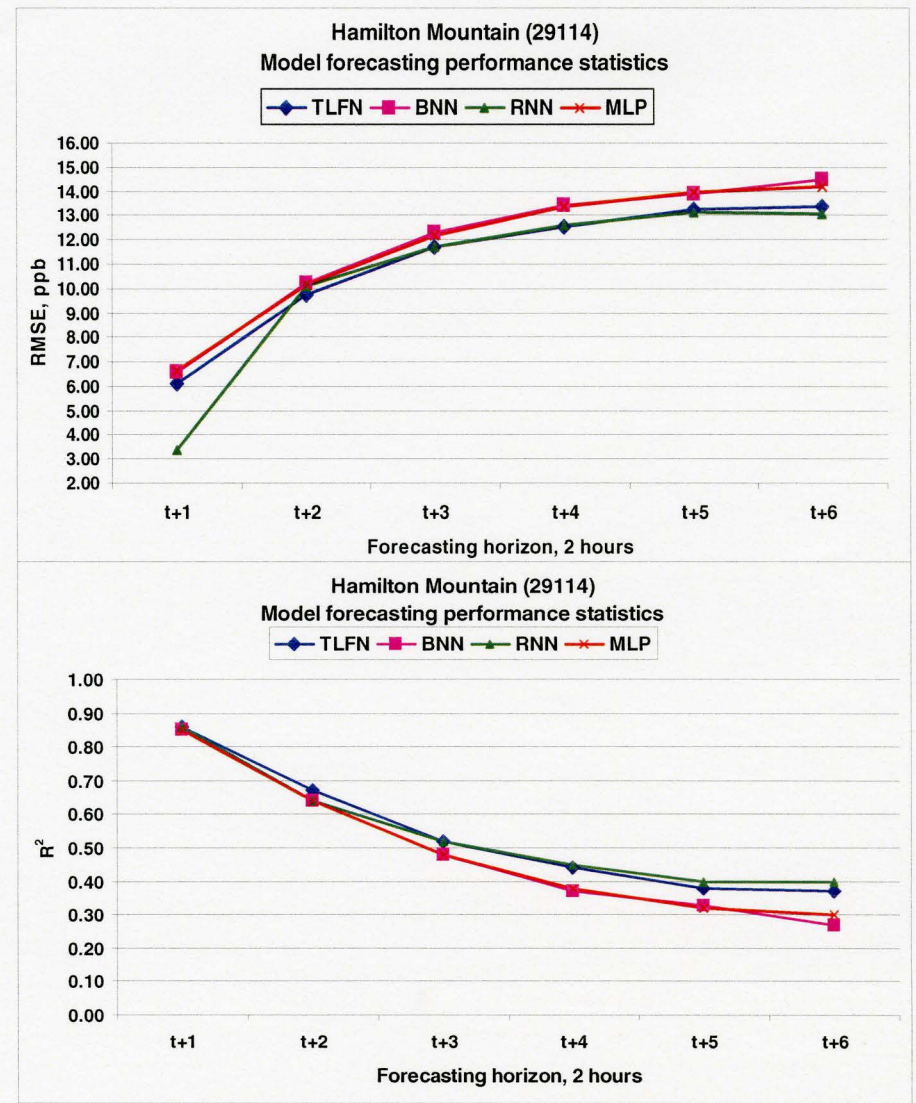
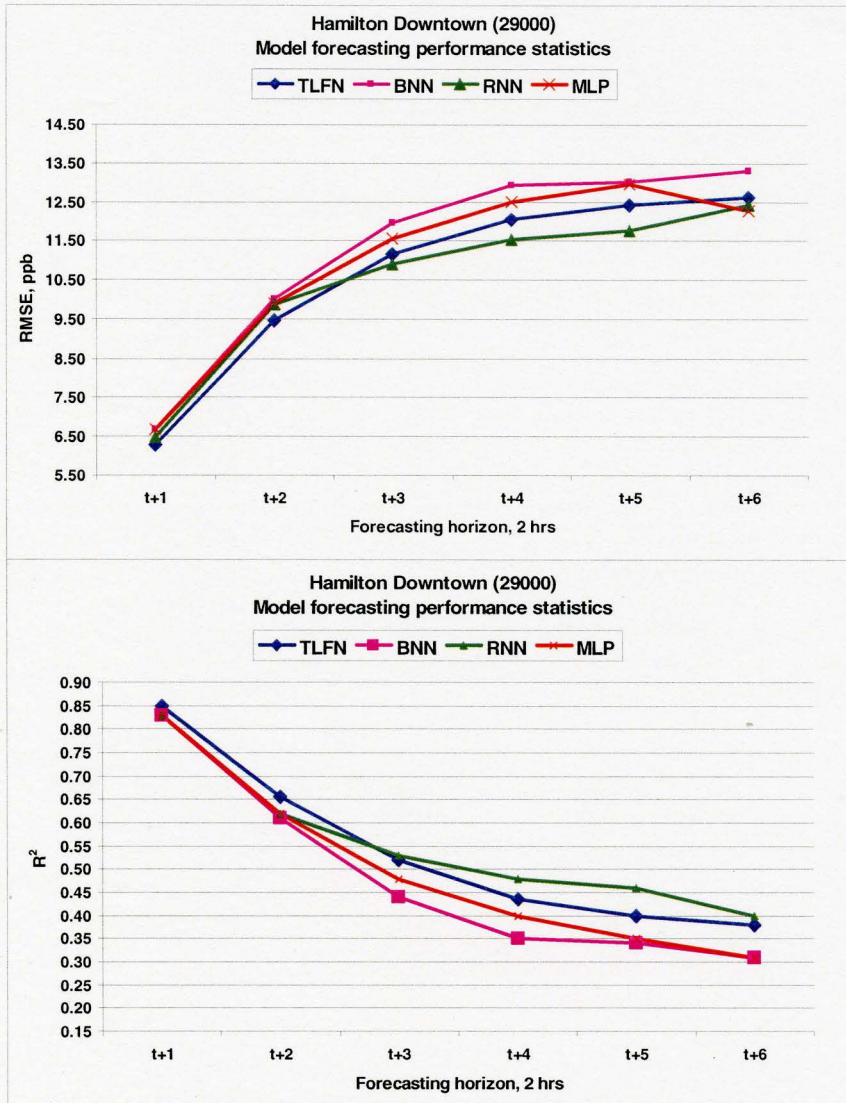


Fig. 6.1 (a): Model forecasting statistics for ozone: Hamilton Downtown (left) and Hamilton Mountain (right)

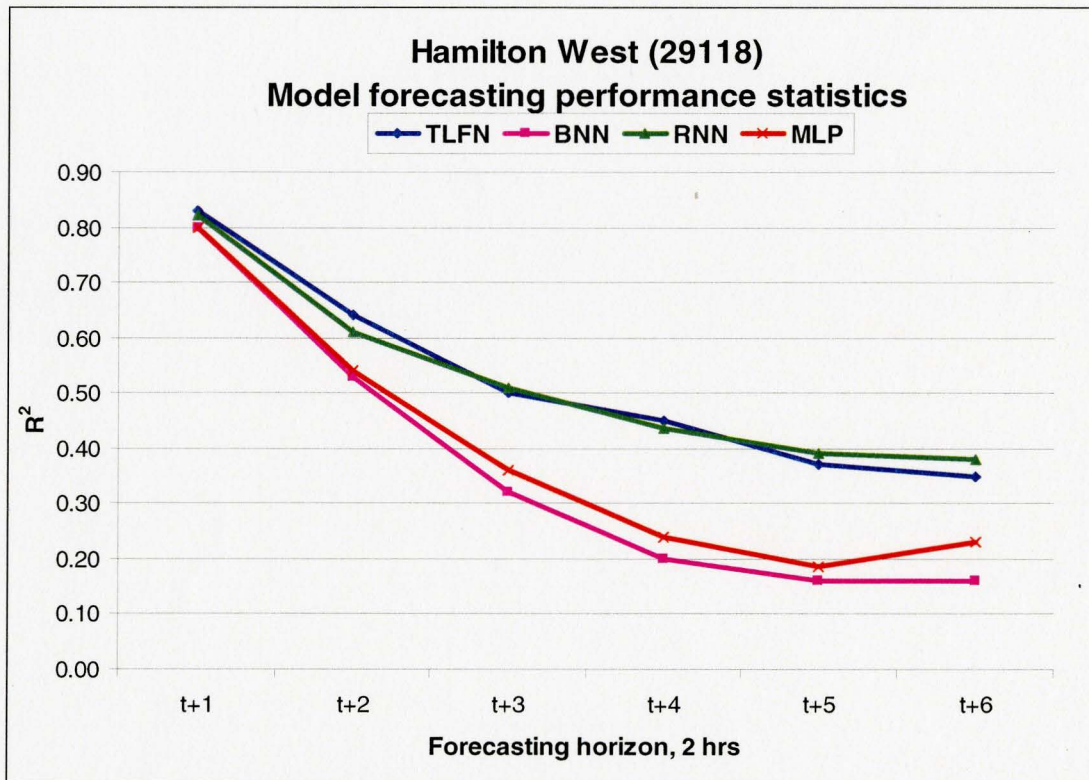
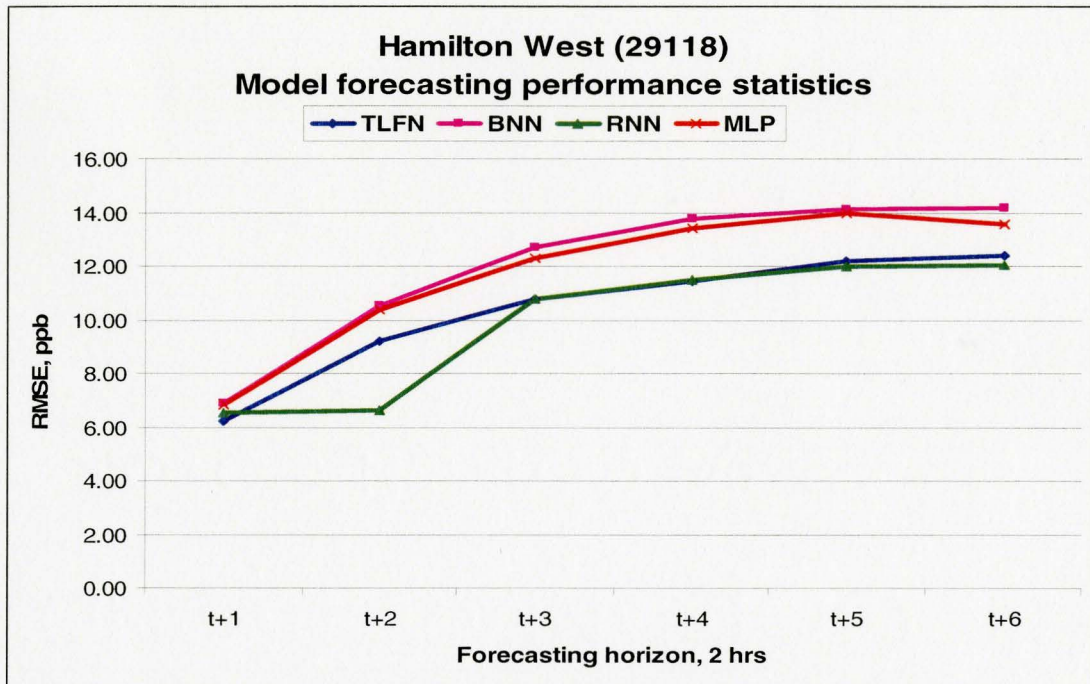


Fig. 6.1(b): Model forecasting statistics for ozone: Hamilton West

Table 6.1: Comparison of model performance at station 29000, 29114 and 29118

Model	Forecasting period, 2 hrs	29000		29114		29118	
		RMSE	% of improvement	RMSE	% of improvement	RMSE	% of improvement
TLFN	1	6.28		6.13		6.22	
	2	9.46	-50.64	9.73	-58.73	9.21	-48.07
	4	12.05	-27.38	12.52	-28.67	11.43	-24.10
	6	12.62	-4.73	13.39	-6.95	12.39	-8.40
BNN	1	6.68		6.57		6.89	
	2	10.01	-49.85	10.24	-55.86	10.55	-53.12
	4	12.94	-29.27	13.46	-31.45	13.76	-30.43
	6	13.32	-2.94	14.52	-7.88	14.17	-2.98
RNN	1	6.48		6.36		6.51	
	2	9.87	-52.31	10.12	-59.12	6.64	-2.00
	4	11.53	-16.82	12.58	-24.31	11.51	-73.34
	6	12.42	-7.72	13.06	-3.82	12.07	-4.86
MLP	1	6.68		6.62		6.83	
	2	9.89	-48.05	10.17	-53.63	10.40	-52.27
	4	12.28	-26.49	13.35	-31.27	13.44	-29.18
	6	12.51	-1.84	14.24	-6.67	13.57	-1.00

Legends: TLFN: Time lagged feed-forward network
MLP: Multilayer perceptron

RNN: Recurrent neural network
TDNN: Time delay neural network

To further assess the model performance in general, scatter plots between the observed and the predicted concentrations were plotted. The best fit line through the observed and predicted concentrations provides another approximation to test the model performance (Gardner and Dorling, 2000). The idea here is: most accurate results will have intercept tending to 0 and gradients tending to 1. Fig. 6.2 (a), (b) and (c) show the scatter plots of t+1, t+4 and t+6 forecasting period in Hamilton Downtown area. In case of 1 step ahead forecasting, the MLP and BNN model diverge significantly from the 45° line and tended to shift towards right. TLFN and RNN models during this time period also shifted but still remain closer to the ideal line. These patterns clearly indicate that both RNN and TLFN model performed more accurately than conventional MLP model. Thus adding an input delay memory or a context unit to the static MLP can be a good alternative to improve the forecasting accuracy. The forecasting results of t+4 (8 hour ahead) and t+6

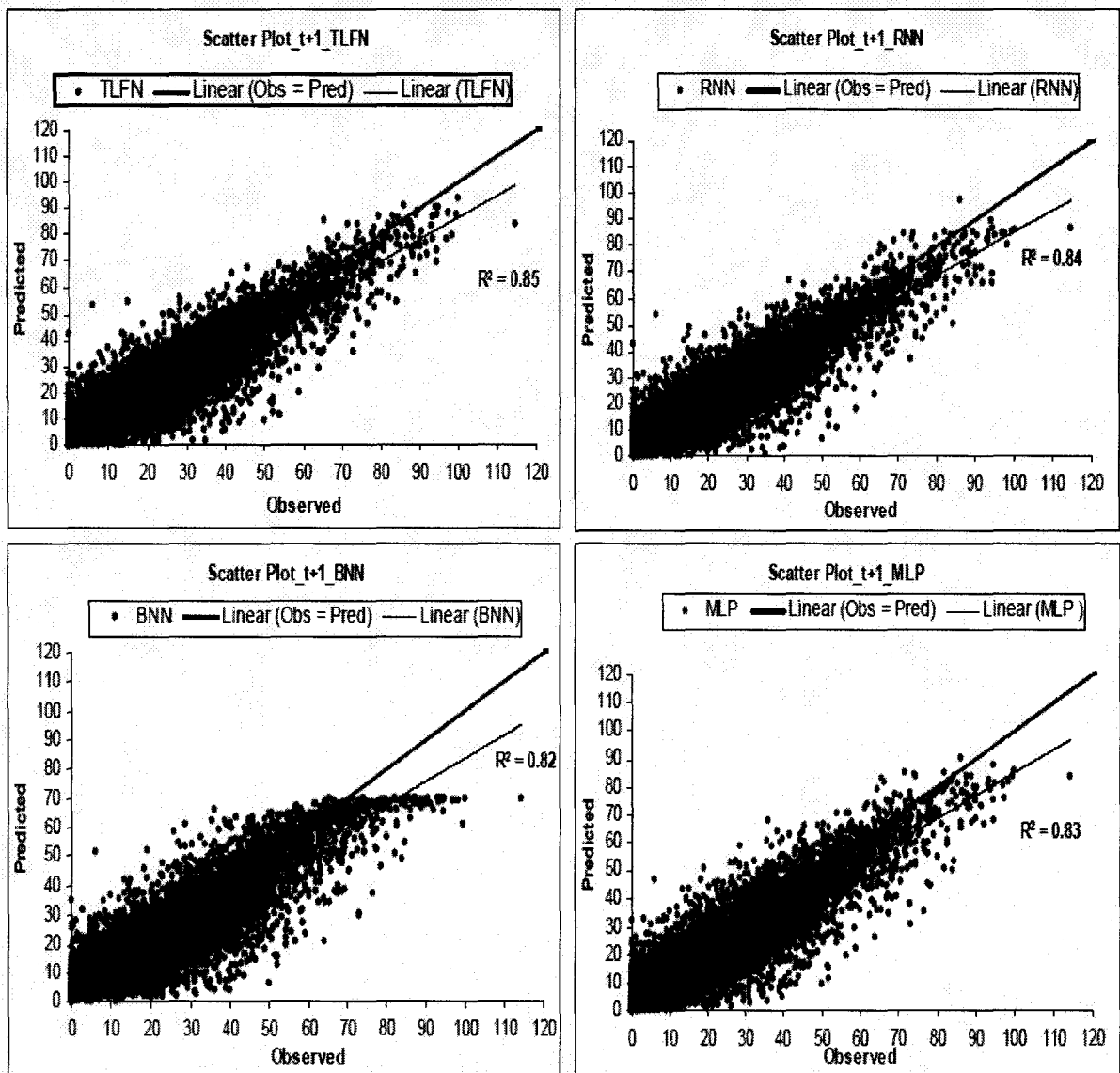


Fig. 6.2 (a): Scatter plots at 1 step (2 hour) ahead: Hamilton Downtown (29000)

In case of low concentrations, predicted values by RNN are relatively better with MLP performing inferior to other two models. Even though the extremely higher and lower values are due to extreme conditions, RNN appears to be more capable of capturing those underlying extreme phenomena. The temporal representation capability of global RNN model is better than the static MLP model and slightly better than the TLFN model. Thus, adding an input delay or an adaptive memory to the conventional MLP can be a good alternative for improving forecasting efficiency.

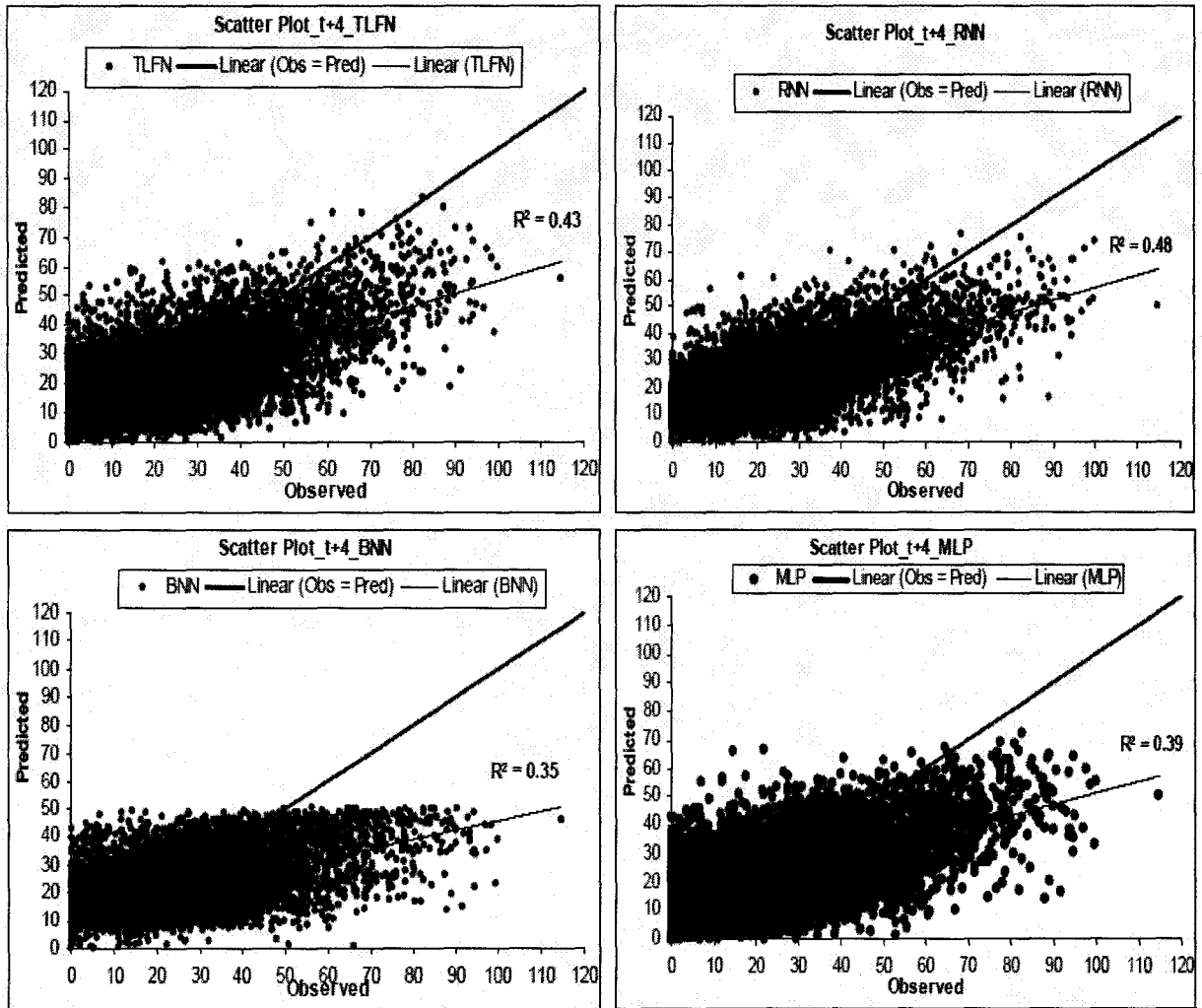


Fig. 6.2 (b): Scatter plots at 4 step (8 hour) ahead: Hamilton Downtown (29000)

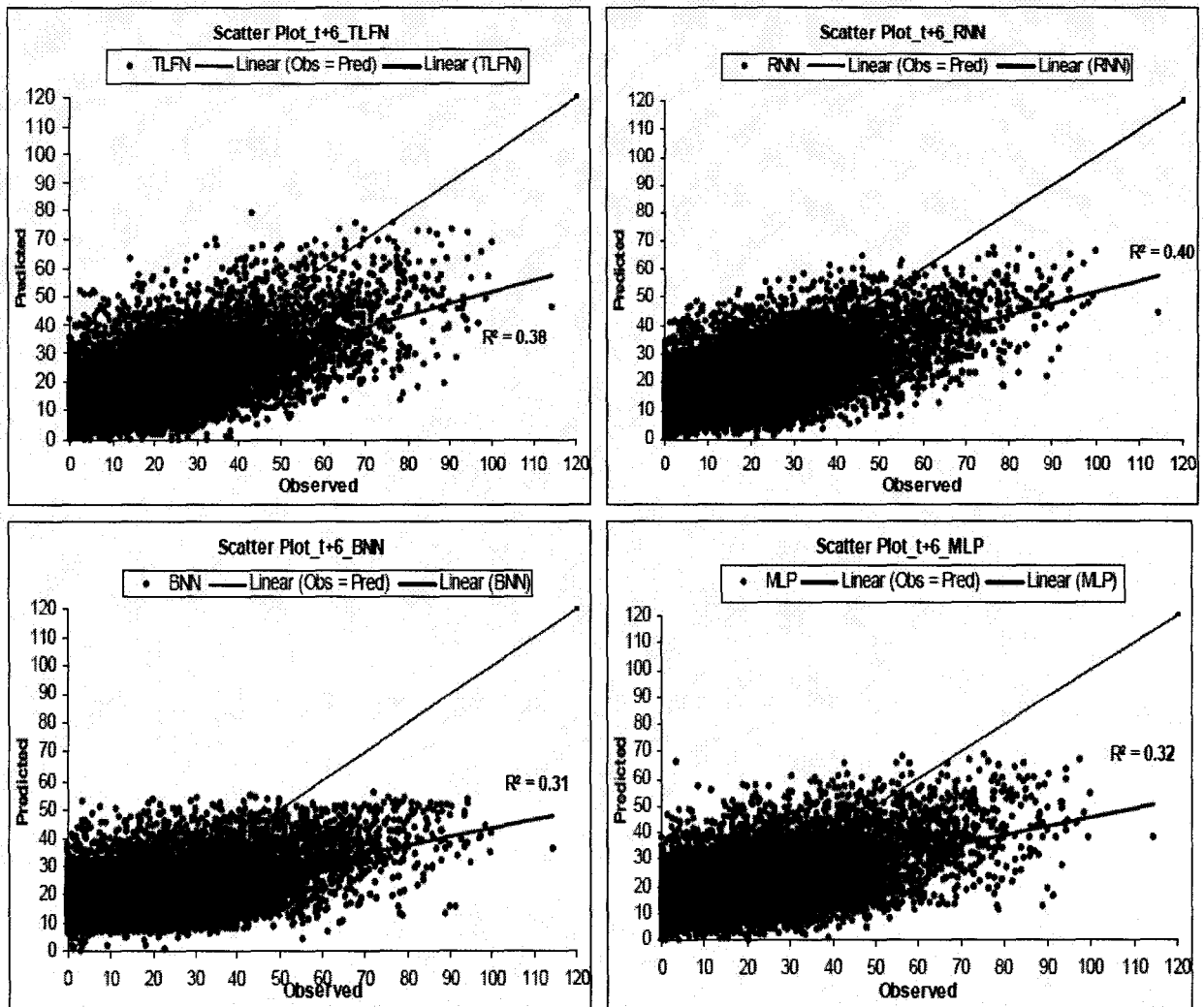


Fig. 6.2 (c): Scatter plots at 6 step ahead: Hamilton Downtown (29000)

6.1.2 Confidence interval with BNN

The performance of the BNN model, although not superior to other neural network model, indicates that they can be a good alternative for short term forecasting (up to 4 hour). Moreover, the BNN model is simpler than the other NN models in terms of the number of neurons. The reason behind this simple-yet-better performance of BNN model may be due to the consideration of parameter uncertainty in the form of probability distributions of weights and biases, and finding the outputs of the networks by integrating over the weight space of posterior probability distribution instead of using single ‘best

set' of weights as in the case of conventional MLP model. This parameter uncertainty consideration and the high computational capability of the nonlinear processing unit increase the capacity of BNN model to outperform the widely used MLP model. The BNN model reproduced concentrations well along with high and low values. Summertime values of the ozone concentrations produced by the models are calculated further with 95% confidence interval using the BNN model for 2 and 4 hour ahead forecasting period. Representation of the confidence intervals about mean estimates is the additional advantage of BNN model which the conventional neural network models cannot provide (Khan and Coulibaly, 2006). Fig. 6.3 (a) and (b) and 6.4 present the confidence interval plots for 2 hour (1 step) and 4 hour (2 step) hour ahead forecasts for mid July-mid Aug, 2004 at sites 29000 and 29114. The uncertainty bands created by the 2 hour ahead BNN model hold both the observed and other modeled values quite well. Hence the performances of BNN and other NN based models are quite competitive; both models performed well in predicting concentrations including high and low concentrations.

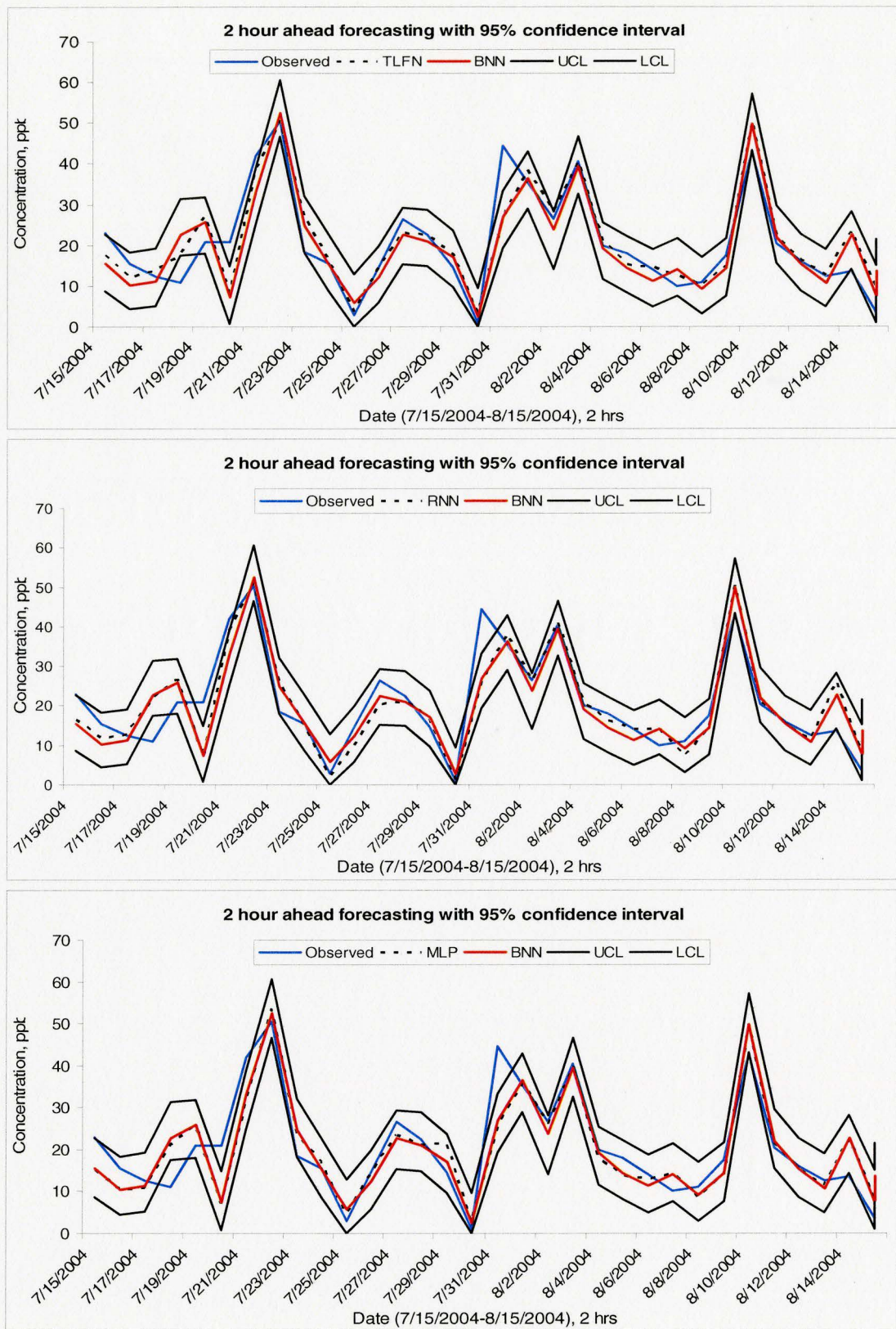


Fig. 6.3 (a): Comparative results of 2 hour ahead ozone forecasting with 95% confidence interval: Hamilton Downtown (29000)

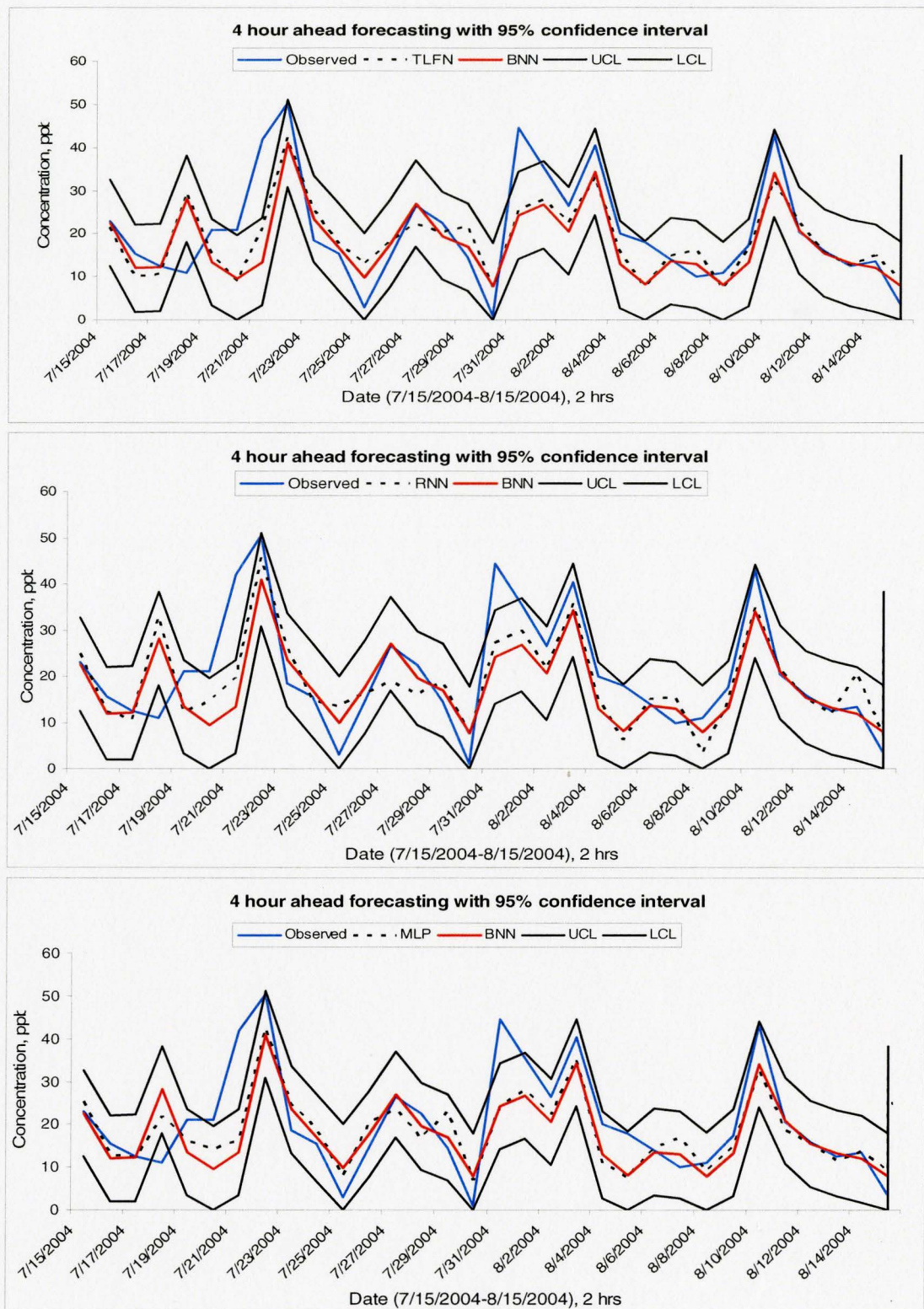


Fig. 6.3 (b): Comparative results of 4 hour ahead ozone forecasting with 95% confidence interval: Hamilton Downtown (29000)

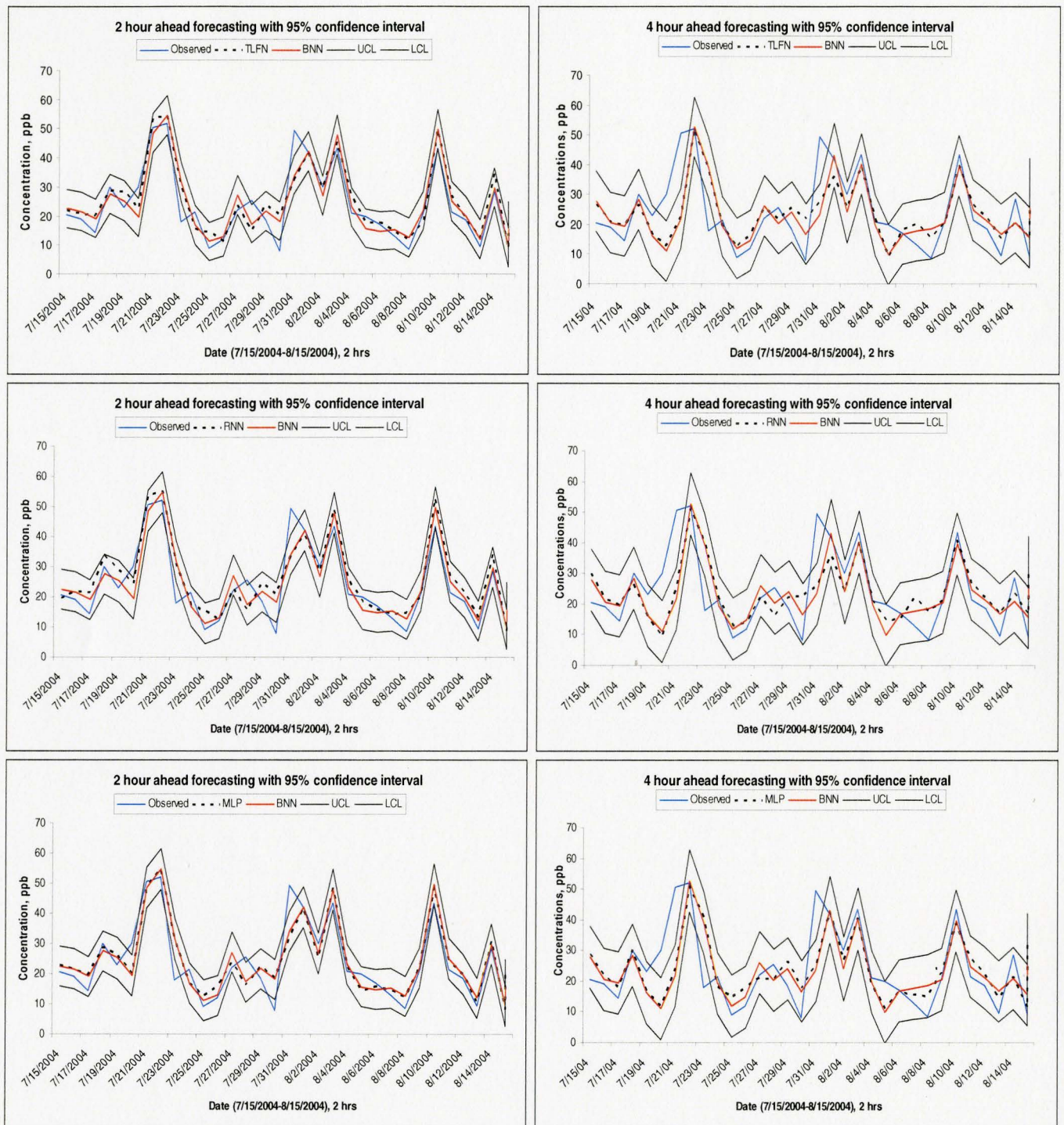


Fig. 6.4: Comparative results of 2 (left) and 4 (right) hour ahead ozone forecasting with 95% confidence interval: Hamilton Mountain (29114)

Except a few deviations, the BNN model band for 4 hour ahead forecasting has also been able to hold both the observed and predicted values. But the gradually increasing uncertainty band indicates the degradation of model performance compared to the previous time step. TLFN, RNN and MLP modeled values fall within the prediction band of BNN, indicating that it can be a reliable tool where uncertainty estimate is of particular concern. These deviated modeled values from the observed concentration are not the limitation of BNN but rather the problem associated with the limitations of the models themselves.

6.1.3 Seasonal variation

It is noteworthy that the ground level ozone concentrations remain higher during May to September because of the photochemical reactions with Isoprene emission which is the principal hydrocarbon precursor of ozone over Southern Ontario and North-eastern United States during summer (McKean et al., 1991). Especially summer season observes frequent ozone episodes during June-August. High temperature and solar radiations act as elevating factor behind its formation. The higher concentrations are also associated with surface winds from south-west indicating the advection of regional pollution from heavily industrialized cities situated in north-eastern United States (Detroit, Chicago, Indianapolis, etc.) with varying additions of urban plumes from New York City and Connecticut. Additional factor like hydroxi-radical (HO_x) providing ozone formation has significant influence on the seasonal variation. Jacob et al. (1995) found evidence for a seasonal transition from NO_x – to hydrocarbon limited O_3 production over southern Ontario and eastern US cities in September, reflecting a decline of the HO_x supply down to levels that can be titrated by NO_x emission. This transition in the photochemical regime results in a large drop of the ozone production efficiency (Hirsch et al., 1996). Hence, the performances of each model have been further analyzed by season. Therefore observed and forecasted data were segregated in the following seasons:

Winter: December, January, February

Spring: March, April, May

Summer: June, July, August

Autumn: September, October, November

The performance of each model was then calculated separately and an inter-comparison of the model performances was performed for each season. Table 6.2 summarizes the forecasting statistics for 1, 2, 4 and 6 step ahead time for site 29000. The ozone trend in each season has been clearly noticed in the model performance; higher errors found during summer times followed by slightly better during spring. Winter has been the clearest season. A new statistical (i. e. seasonal) bias has been introduced to estimate the % of under-prediction or over-prediction per season. Interestingly, both spring and summer values are being under-predicted by the models and over prediction have occurred for winter and fall. This visualization again reveals the model's limitation to predict higher concentration episodes. A general observation is that overall whatever the forecast step, the MLP model has produced a higher bias as compared to TLFN and RNN models.

Table 6.2: Seasonal variation of model performance for O₃: Hamilton Downtown (29000)

Forecasting period	Season	Model	Model performance statistics					
			RMSE	r	R ²	NMSE	MAE	SB
t+1	Fall	TLFN	5.42	0.92	0.85	0.15	3.95	0.019
		BNN	6.03	0.90	0.81	0.19	4.39	0.015
		RNN	5.69	0.91	0.83	0.17	4.17	0.029
		MLP	5.88	0.91	0.82	0.18	4.29	0.011
	Summer	TLFN	7.86	0.91	0.84	0.17	5.77	-0.030
		BNN	8.63	0.90	0.81	0.20	6.33	-0.047
		RNN	8.06	0.91	0.83	0.17	5.91	-0.037
		MLP	8.40	0.90	0.82	0.19	6.21	-0.040
	Spring	TLFN	7.00	0.87	0.76	0.25	5.07	-0.034
		BNN	7.44	0.85	0.72	0.28	5.46	-0.036
		RNN	7.15	0.86	0.74	0.26	5.22	-0.026
		MLP	7.31	0.86	0.74	0.27	5.35	-0.037
	Winter	TLFN	4.44	0.87	0.76	0.24	3.33	0.017

		BNN	4.68	0.86	0.74	0.27	3.58	0.038
		RNN	4.60	0.86	0.75	0.26	3.50	0.043
		MLP	4.66	0.86	0.74	0.26	3.52	0.007
t+2	Fall	TLFN	8.36	0.80	0.65	0.36	6.33	0.05
		BNN	8.87	0.77	0.60	0.40	6.68	0.02
		RNN	8.90	0.77	0.60	0.41	6.74	0.06
		MLP	8.80	0.78	0.60	0.40	6.59	0.02
	Summer	TLFN	11.96	0.79	0.62	0.38	9.12	-0.06
		BNN	12.77	0.76	0.58	0.44	9.60	-0.08
		RNN	12.49	0.77	0.59	0.42	9.40	-0.07
		MLP	12.51	0.77	0.59	0.42	9.36	-0.09
	Spring	TLFN	10.19	0.70	0.49	0.53	7.74	-0.05
		BNN	10.74	0.67	0.45	0.59	8.16	-0.08
		RNN	10.41	0.68	0.46	0.55	7.99	-0.04
		MLP	10.63	0.68	0.46	0.57	8.13	-0.08
	Winter	TLFN	6.60	0.69	0.48	0.53	5.13	0.04
		BNN	6.80	0.67	0.44	0.56	5.41	0.04
		RNN	6.90	0.66	0.44	0.58	5.48	0.08
		MLP	6.87	0.66	0.44	0.57	5.40	0.03
t+6	Fall	TLFN	11.31	0.59	0.35	0.66	8.94	0.05
		BNN	11.86	0.53	0.29	0.72	9.43	0.08
		RNN	10.92	0.63	0.39	0.61	8.65	0.06
		MLP	11.88	0.53	0.28	0.72	9.40	0.07
	Summer	TLFN	16.25	0.57	0.32	0.71	12.61	-0.10
		BNN	17.33	0.50	0.25	0.80	13.45	-0.15
		RNN	16.14	0.56	0.31	0.70	12.78	-0.05
		MLP	17.23	0.51	0.26	0.79	13.37	-0.14
	Spring	TLFN	12.75	0.47	0.22	0.82	10.23	-0.08
		BNN	13.40	0.40	0.16	0.91	10.78	-0.12
		RNN	12.74	0.45	0.21	0.82	10.25	-0.07
		MLP	13.40	0.42	0.17	0.91	10.84	-0.12
	Winter	TLFN	9.02	0.32	0.10	0.99	7.21	0.00
		BNN	9.36	0.22	0.05	1.06	7.70	0.09
		RNN	8.65	0.36	0.13	0.91	7.01	-0.03
		MLP	9.32	0.26	0.07	1.05	7.54	0.07

Considering 1 step ahead forecasting of summer season, the bias generated by all models are relatively low. It ranged from 0.03% to 0.04% with highest from MLP. For 4

step ahead forecasting, summer season bias remained in the range of 0.1% to 0.15% again with higher bias generated by MLP. For 6 step ahead, the BNN model performance started degrading with slightly higher bias (0.15%) along with the MLP model (0.14%). During spring, the bias values are even higher in some cases; MLP again has produced highest bias. So overall, TLFN and RNN models have shown better performances for predicting seasonal values during summer and spring. During fall and winter, a different scenario has been captured. Here RNN has performed worse than other 3 models in terms of bias. Again BNN performance degraded later rapidly which indicated BNN may not be a good alternative for the further forecasting time after 4 step. During these seasons, TLFN has shown superior performances. Hence it can be said that considering all seasons TLFN has performed quite well comparing to its contemporary models. Similarly for Site 29114 and 29118, summer and spring was worse in terms of forecasting performances which has been shown in Table A. 4 (a) and (b) respectively (See Appendix).

6.1.4 Annual vs. summer model

It is now well established that summertime ground level ozone has been a serious problem for several decades in many metropolitan areas of the world. Therefore developing, maintaining and improving the ozone forecasting model is an important task for the environmental and health authorities (Chaloulakou, et al., 2003). In this study an attempt has been taken to develop a model especially for summer and hence is named as 'summer model'. The 'annual model' performance is then compared with the summer models. Table 6.3 and Table A. 5 (Appendix) presents a comparison of the forecasting statistics of annual and summer models. From the Table 6.3 it is seen that in terms of RMSE, the percentage of improvement is very less; in most cases summer models have performed slightly worse than the annual model. Only BNN model at 1 step ahead forecasting and RNN at 3 and 4 step ahead have shown some improvement. In most cases the summer model has shown an inferior performance which is most visible in case of TLFN and MLP models where it had more than 6% root mean square error than the

annual model. This scenario clearly demonstrates that the annual model is entirely capable of capturing the complex non-linear relationships of ozone formation regardless of the season in this case.

Table 6.3: Annual model vs. summer model: Hamilton Downtown (29000)

Forecasting period	Models	RMSE from annual model	RMSE from summer model	% improvement
t+1	TLFN	7.86	7.88	-0.25
	BNN	8.63	8.43	2.30
	RNN	8.06	8.08	-0.29
	MLP	8.40	8.37	0.34
t+2	TLFN	11.96	12.38	-3.51
	BNN	12.77	13.09	-2.51
	RNN	12.49	12.73	-1.92
	MLP	12.51	12.72	-1.68
t+3	TLFN	14.24	15.11	-6.11
	BNN	15.40	15.74	-2.21
	RNN	13.99	13.72	1.93
	MLP	14.78	15.49	-4.83
t+4	TLFN	15.55	16.07	-3.34
	BNN	16.86	17.20	-2.02
	RNN	14.92	14.71	1.38
	MLP	16.04	16.69	-4.05
t+5	TLFN	15.95	16.80	-5.33
	BNN	16.98	18.05	-6.30
	RNN	15.04	15.28	-1.60
	MLP	16.65	17.32	-4.02
t+6	TLFN	16.25	17.02	-4.77
	BNN	17.33	18.33	-5.77
	RNN	16.14	15.80	2.11
	MLP	17.23	17.66	-2.50

6.2 Ozone forecasting with land use variables

Because ozone concentration depends on the location and the type of the area which highly depends on the geography, traffic and population surrounding that location, therefore the annual model is further modified using the land use types around 200 m buffer of the pollutant monitoring sites. These variables consists of several land uses such as residential, commercial, institutional types described in section 2.2.3. Table 6.4 and Table A. 6 presents a comparative assessment of the model forecasting statistics with and

Table 6.4: Comparison of model performance using land use variables: Hamilton Downtown (29000)

Forecasting period	Model	RMSE		
		Without land use	With land use	% reduction in RMSE
t+1	TLFN	6.28	7.89	-25.62
	BNN	6.68	8.36	-25.10
	RNN	6.48	8.61	-32.89
	MLP	6.68	8.41	-25.85
t+2	TLFN	9.46	12.34	-30.40
	BNN	10.01	13.28	-32.64
	RNN	9.87	12.94	-31.10
	MLP	9.89	12.76	-29.02
t+3	TLFN	11.16	14.36	-28.68
	BNN	11.96	15.02	-25.59
	RNN	10.90	15.35	-40.87
	MLP	11.58	15.08	-30.20
t+4	TLFN	12.05	15.20	-26.13
	BNN	12.94	16.43	-26.95
	RNN	11.53	15.91	-37.95
	MLP	12.51	15.96	-27.55
t+5	TLFN	12.42	15.71	-26.49
	BNN	13.03	17.12	-31.41
	RNN	11.78	16.46	-39.73
	MLP	12.97	16.81	-29.60
t+6	TLFN	12.62	15.91	-26.08
	BNN	13.32	17.97	-34.93
	RNN	12.42	16.38	-31.89
	MLP	12.28	17.10	-39.29

without considering land use types. It appears that the performance of the basic annual model without considering land use types performed better than the later one. The performance of RNN was even worse; in most cases the inferiority of RMSE value exceeded 35% than the basic annual model. This type of results indicate that in temporal problems, the inclusion of land use types in a form of logical input may not be the right form to capture the impacts of different land use types rather using emission factors from different sources may be a good alternative to count the influences of different sources of the pollutants.

6.3 Nitrogen Dioxide Forecasting Results

This section describes the performances of NO₂ forecasting models at two NO₂ monitoring stations in Hamilton.

6.3.1 Model forecasting performances

The performances of the forecasting models have been compared in Table 6.5 and Table A. 7 (a) and (b) and Fig. 6.5. Similar to ozone model results, the performances of all models for NO₂ forecasting are very close. One important observation here is compared to TLFN, BNN and RNN models, the performances of MLP model are also competitive. All the models deteriorated with time up to 7 step ahead. The performances of the models can be further analyzed based on their RMSE values as presented in Table 6.5. It is seen that for 2 step ahead the RMSE value dropped largely (around 40% of the previous step) with RNN model as highest as 42.46% for site 29000 and 41.17% for site 29118. Here MLP performed slightly well with a drop of around 38% for both sites. MLP performed better up to 3 step ahead which got worse later compared to other three models (Table A. 7 in Appendix). TLFN, although was slightly inferior to the rest three models, started acting well from 4 step ahead. So overall performances of the models indicate that each of them has performed quite competitively and can be applied up to 6 step ahead where the r value remains equal to or greater than 0.40.

The performances of the forecasted models were further analyzed by plotting scatter plots for Hamilton downtown site. Fig. 6.6 (a) and (b) show the scatter plots of 2 and 6 hour ahead forecasting period in Hamilton Downtown area. In case of 2 hour ahead forecasting, all 4 models performed well, shifted slightly from the 45^o line and having coefficient of determination value more than 0.70. While for 6 hour, a large degradation appears, with concentrations shifting widely from the 45^o line. Like ozone models, the 4 hour ahead scatter plots also indicate that these models have limitation in capturing higher and lower concentrations. The performance of BNN, in particular is worse than rest three models in this case.

Table 6.5: Comparison of model performance of NO₂ at station 29000 and 29118

Model	Forecasting period, 2 hrs	29000		29118	
		RMSE	% improvement	RMSE	% improvement
TLFN	1	5.76		5.77	
	2	8.07	-40.10	8.04	-39.34
	4	9.27	-14.87	9.83	-22.26
	7	10.00	-7.87	10.74	-9.26
BNN	1	5.79		5.99	
	2	8.02	-38.51	8.39	-40.07
	4	9.41	-17.33	10.15	-20.98
	7	9.98	-6.06	10.89	-7.29
RNN	1	5.77		5.83	
	2	8.22	-42.46	8.23	-41.17
	4	9.30	-13.14	9.86	-19.81
	7	9.87	-6.13	10.60	-7.51
MLP	1	5.85		6.04	
	2	8.09	-38.29	8.39	-38.91
	4	9.35	-15.57	10.00	-19.19
	7	9.96	-6.52	10.93	-9.30

6.3.2 Seasonal variation

Similar to the ozone forecasting models, the performances of the NO₂ models were analyzed based on their seasonal performances. Table A. 8 represents a comparative study of the TLFN, BNN, RNN and MLP models at 1, 2 and 6 step ahead of time at site

29000. The results of three forecasting periods show similar trend; higher errors during spring and summer and comparatively low error during winter and fall. Particularly during 1 step ahead forecasting during spring, all models had root mean square error ranging from 7.15 ppb to 7.25 ppb which is nearly 30% higher than the RMSE values during winter. But in case of similarly for the 2 and 6 step ahead forecasting the RMSE values during spring were approximately 20% -25% higher than the winter. The r values were also similar. There were slight under-predictions (0.02-0.10%) during spring and winter seasons, which is due to these model's limitations to underestimate the higher and lower concentrations. The performances of these models, however, are very close to each other and hence have not been analyzed separately.

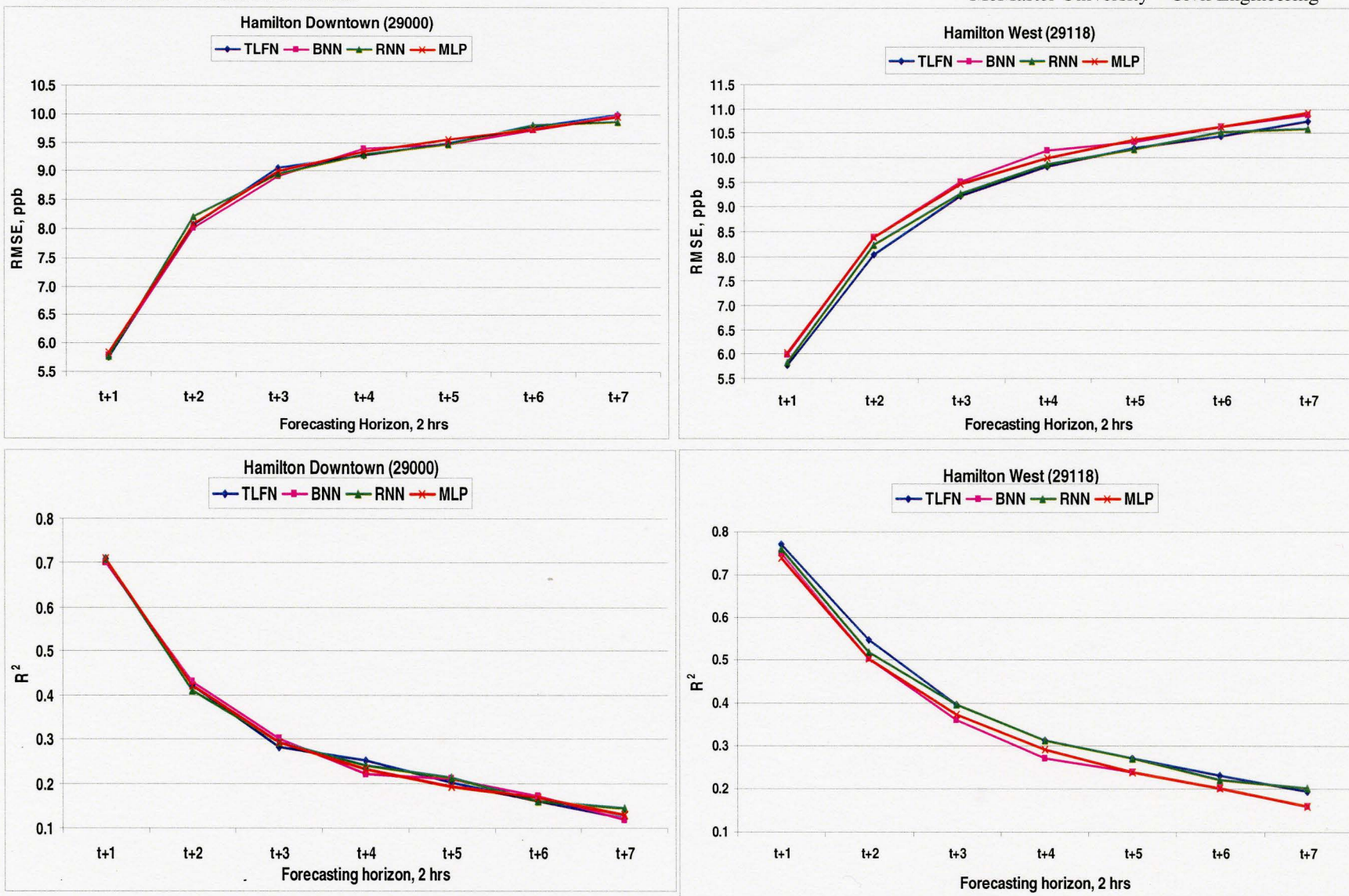


Fig. 6.5: Model forecasting statistics for NO₂: Hamilton Downtown (29000) (top) and Hamilton Mountain (29114) (bottom)

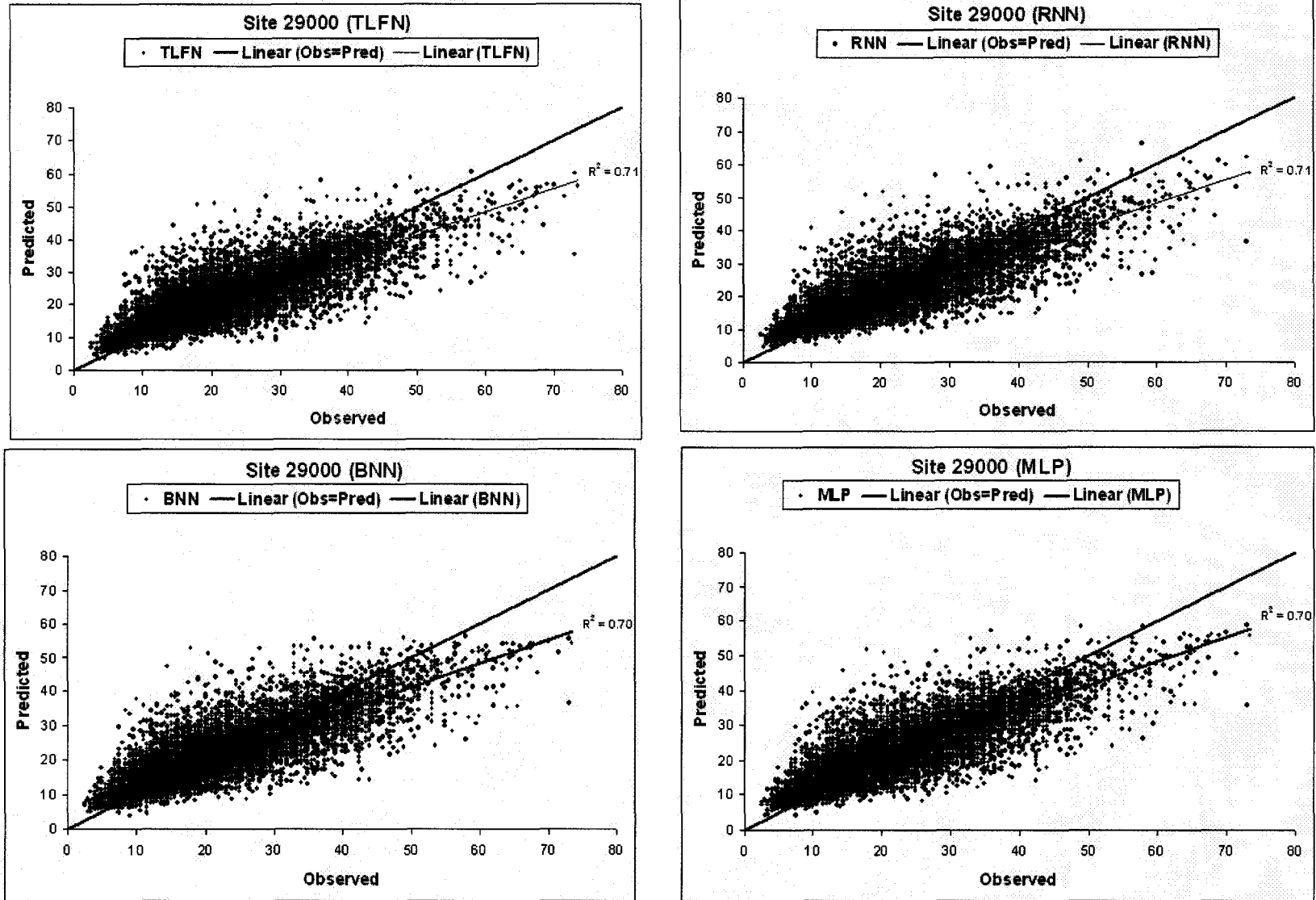


Fig. 6.6 (a): Scatter Plots of 2 hour ahead forecasting: NO₂ at site 29000

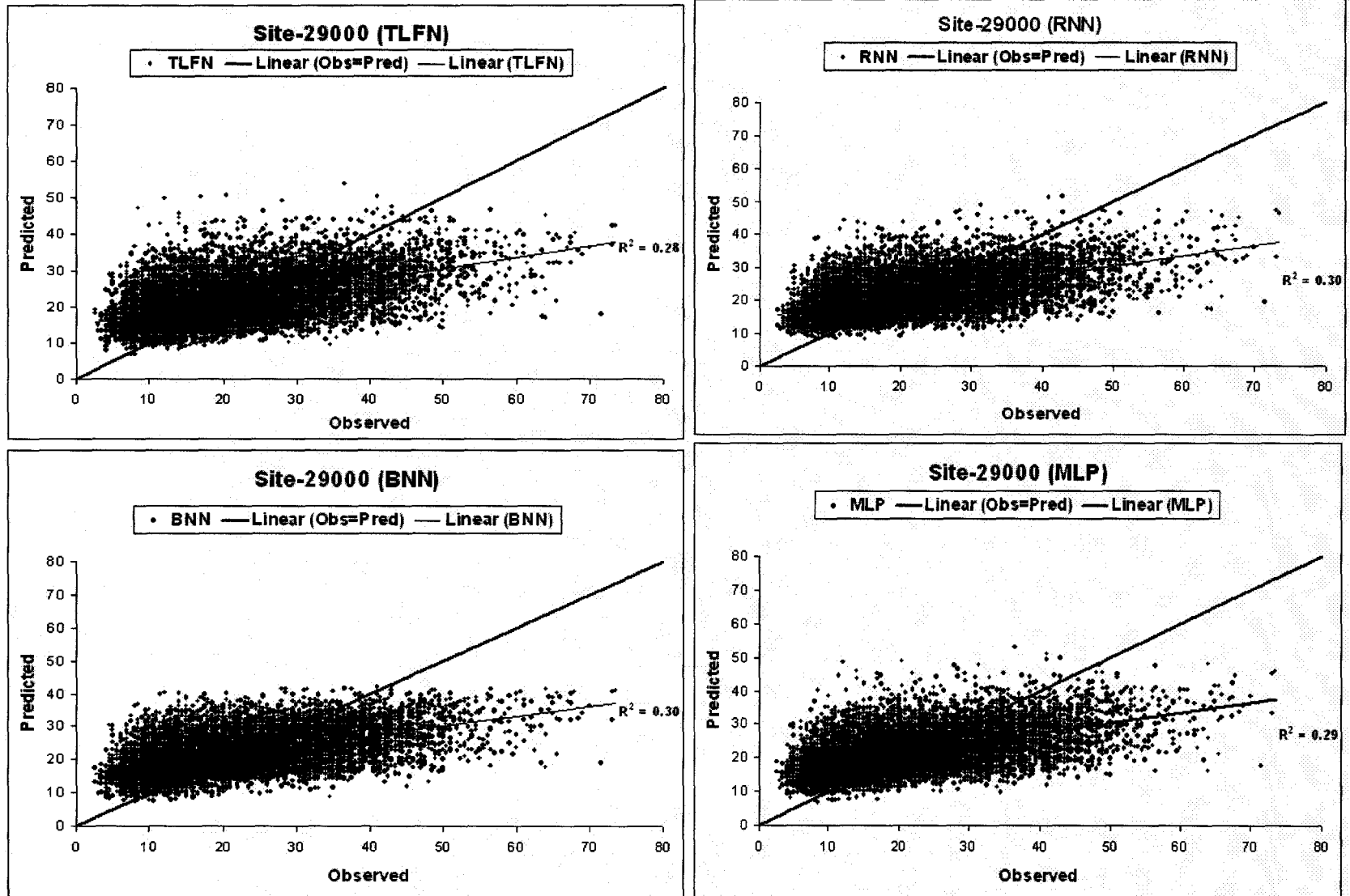


Fig. 6.6 (b): Scatter Plots of 6 hour ahead forecasting: NO₂ at site 29000

6.4 Summary of O₃ and NO₂ forecasting results

Through the analysis of the O₃ and NO₂ forecasting results for the monitoring sites within Hamilton using TLFN, BNN, RNN and MLP models, several conclusions can be drawn.

Considering the overall performances of O₃ forecasting models, the performances of the 4 neural network models are similar. All the models have shown variation up to 6 steps ahead with degrading with time. TLFN outperformed other models in first two steps followed closely by the RNN model. RNN slightly outperformed TLFN during 3 to 6 step ahead which clearly shows both of these models' superiority over BNN and static MLP models. When compared by season, the summer was the worst season with low performances. It has also been seen that the annual model is enough to capture the nonlinear relationships of the meteorology and pollutants in this case. However, the reason behind similar performances of annual and summer model may be due to the absence of more appropriate variables such as solar radiation and maximum temperature which are believed to be the two most influential parameters behind the high O₃ concentrations during summer period. The inclusion of land use variables with meteorology has not been able to improve the model performances rather it increases the computational time and cost. Hence the meteorology itself is able to forecast with significant efficiency in this study. This demonstrates that in temporal problems, the inclusion of land use types in a form of logical input may not be the right form to capture the impacts of different land use types rather using emission factors from different sources may be a good alternative to count the influences of different sources of the pollutants. The results also revealed superior performances of TLFN and RNN models over static MLP and BNN models. These results further suggest that the inclusion of time delay and/or adaptive memory (context unit) in MLP have the capacity to improve the results obtained from conventional static neural network (MLP in this case). These performances, however indicate that RNN model has the best generalization performance

and suggest that the relationship between ozone and the meteorology can better be represented using these predictor variables still with room of improvement at three ozone monitoring sites in Hamilton.

Model forecasting results for NO₂ show very close performances. In this case, the models showed varying performances up to 7 step (14 hour) ahead. While performing by season, spring was the worst with highest errors and winter came to be the best season with least errors. One interesting observation here is that the static multilayer perceptron model has competed quite well with the dynamic NN models. Hence applying MLP with relevant variables can be enough to get better forecasting result for NO₂.

Chapter 7

Simulation Results for Total Suspended Particulates and PM₁₀ Pollutants

This chapter deals with the TSP and PM₁₀ simulation results obtained from time lagged feed-forward neural network (TLFN), Bayesian neural network (BNN), recurrent neural network (RNN) and multilayer perceptron (MLP) models and interpolates the results obtained from five TSP monitoring sites in Hamilton region. Due to unavailability of a suitable dataset, forecasting could not be performed. Rather this chapter deals with “simulation experiments” where the best combinations of the input variables were identified by comparing different combinations of the input variables by trial-and error approach and then by identifying the optimal model performances in simulating TSP and PM₁₀ values at the monitoring locations. As described in chapter 2, a total of 10 years of database (1995-2004) has been considered. First 7 years (1995-2001) of the observed meteorological data (as predictors) and the historical TSP and PM₁₀ data (as predictand) has been used for model calibration at each monitoring stations. In order to prevent overtraining of the models 1 year (2002) data has been selected for cross- validation. The remaining 2 years (2003-2004) of data are being used to test the models. After good statistical agreement between the observed and simulated values has been achieved, the simulated model results were then interpolated spatially over the region to investigate their capability to generate values at unsampled sites.

Like the previous chapter, the analysis and discussion emphasize the test results only as they provide real evaluation information about model performance owing to the use of independent datasets from calibration.

7.1 Simulation of TSP

7.1.1 Comparison of different input variables

In order to get optimum results, here a different approach than the one used in O₃ and NO₂ pollutants forecasting is considered. This section discusses a comparative performance of the combinations of the input variables considered for simulating TSP. With a view to reduce computational time and complexity only two stations were tested with all four models. The comparative results of the performances of the models while considering different combinations have been summarized in Table 7.1. Here four combinations were considered: case 1 deals with considering only nearby sampled sites to simulate target station's concentrations while in case 2 meteorological variables from Hamilton Airport and Burlington Piers stations were used as predictors to model predictand stations. Case 3 includes percentages of different land use types within 200 m buffer of the predictor and predictand stations and finally case 4 uses meteorological variables from the two meteorological stations, historical TSP concentrations from nearby sampled sites and the percentage of land use types around the stations together. In case of site 29000, case 1 generates better results. Including land use with the nearby sites (i.e. case 3) has improved the results only at a small extent but it is rather complex and costly in terms of time. That's why case 1 was considered for further analysis. On the other hand, in site 29102, the best performance was achieved when the meteorological variables were used as predictor variables (Case 2). During the analysis, it is assumed that because remaining 3 stations (sites 29025, 29113 and 29114) are located near to site 29000, the best combinations would be same as site 29000. Hence they were analyzed using case 1.

7.1.2 Model performance statistics

The comparative analysis of neural network models has provided very interesting results. The performances of the model simulation results for five monitoring stations in Hamilton have been compared in Table 7.2. The performances of the models have been

assessed using RMSE, r , R^2 , NMSE and MAE. The results obtained from sites 29000 (Hamilton Downtown) and 29025 (Barton/Sanford) which are mostly residential areas, showed greater response to the MLP model at site 29000 with an *RMSE* value of 14.68 $\mu\text{gm}/\text{m}^3$ and R^2 value 0.76. RNN model performed second best with slightly higher *RMSE* value than the MLP model. Other statistics were also very competitive. The third station at Gertrude/Depew is situated totally in an industrial area. TLFN, RNN and MLP models performed better than BNN model. BNN was the inferior model here with an *RMSE* 17% less than the RNN model. The r value was also less than 0.75 with greater than 0.75 for other three models. In case of a multi-purpose area like site 29114, TLFN model performed better than rest three models with an *RMSE* value of 13.6 ppb and r value 0.83; the performances of MLP and RNN models were also competitive with *RMSE* of 13.91 ppb and 13.75 ppb and r value of 0.83 and 0.8 respectively. Finally none of the models worked well for site 29102. The reason behind this inferior performance can be the absence of nearby sampling sites and lack of weather data around the station as the dispersion TSP is more driven by the meteorological variables such as wind direction and wind speed which carries the pollutants from nearby highly polluted sites. In addition due to its location near Lake Ontario, the pollutant formation is most probably more dependent on the lake breeze effects, cloud cover height, etc. which has not been considered here.

The performances of each model have been compared by visual assessment of the observed and predicted concentrations. The observed versus simulated values for the test period i.e. 2003-2004 are plotted in Fig. 7.1 and Fig. A. 1 (a), (b) and (c). In case of site 29000 and 29025, it appears that the MLP model provided the most accurate predictions of the TSP concentration levels as the observed and simulated values tend to be in greater accordance compared to other three models. On the other hand, although TLFN and BNN models showed greater agreement between observed and simulated values, their performances are not satisfactory in simulating extreme values. The results generated for sites 29113, 29114 and 29102 are also similar. Interestingly, the inclusion of input delays

in the MLP network (i.e. TLFN) and RNN model with an adaptive memory requires 10% and 25% additional time than the static model without much improvement.

Table 7.1: Combination of input variables for TSP
(a) TLFN, MLP, RNN models

Target station	Input Variables	TLFN					RNN					MLP				
		RMSE	r	R ²	MAE	NMSE	RMSE	r	R ²	MAE	NMSE	RMSE	r	R ²	MAE	NMSE
29000	Case 1: Stations together	15.32	0.85	0.72	10.02	0.28	14.68	0.87	0.76	10.35	0.26	15	0.86	0.74	10.19	0.27
	Case 2: Met. Variables	26.79	0.44	0.19	22.22	0.83	26.42	0.44	0.19	20.98	0.81	26.6	0.44	0.19	21.5	0.77
	Case 3: Land use +stations	15	0.87	0.76	10.23	0.27	14.31	0.87	0.76	9.86	0.24	14.98	0.85	0.72	9.98	0.29
	Case 4: Met variables + land use + stations	24.59	0.55	0.30	19.57	0.72	23.21	0.6	0.36	19.04	0.64	24	0.56	0.31	19.65	0.85
29102	Case 1: Stations together	44.93	0.53	0.28	34.18	0.76	45.12	0.52	0.27	34.29	0.77	44.8	0.51	0.26	34.04	0.76
	Case 2:Met. Variables	39.09	0.55	0.30	30.22	0.7	39.17	0.54	0.29	29.43	0.7	39.15	0.53	0.35	29.6	0.72
	Case 3: Land use +stations	44.7	0.56	0.31	33.9	0.75	46.1	0.52	0.27	34.3	0.8	45.3	0.51	0.26	33	0.77
	Case 4: Met variables + land use + stations	44.49	0.51	0.26	31.5	0.75	43.02	0.55	0.28	30.2	0.71	44	0.52	0.27	30.5	0.76

Legends: TLFN: Time lagged feed-forward network
MLP: Multilayer Perceptron

RNN: Recurrent neural network
TDNN: Time delay neural network

(b) BNN model performance

Target station	Input variables	Model evaluation parameters				
		RMSE	r	R ²	MAE	NMSE
29000	Stations together	15.01	0.85	0.72	10.38	0.29
	Met. variables	27.02	0.41	0.17	21.06	0.89
	Case 3: Land use +stations	14.89	0.86	0.79	10	0.27
	Met variable + land use + stations	23.59	0.56	0.31	19.65	0.85
29102	Stations together	44.78	0.52	0.27	34.1	0.77
	Met. variable	39.09	0.54	0.29	29.6	0.72
	Land use + stations	45.3	0.51	0.26	33.14	0.79
	Met. Variable + land use + stations	44.06	0.53	0.28	30.61	0.78

Legend: Nhidden: No. of processing units Nouter: No. of loops Alpha: Initial prior hyperparameter
Option 14: No. of iteration in each loop Beta: Initial noise parameter

Table 7.2: Model performance statistics for TSP

Station	Models	Model performance statistics				
		RMSE	R	R ²	MAE	NMSE
29000	TLFN	15.32	0.85	0.72	10.02	0.28
	BNN	15.05	0.86	0.73	7.53	0.27
	RNN	14.68	0.87	0.76	10.35	0.26
	MLP	14.52	0.87	0.76	9.63	0.25
29025	TLFN	18.6	0.85	0.72	12.22	0.3
	BNN	19.68	0.83	0.68	13.1	0.34
	RNN	20.17	0.82	0.67	14.57	0.36
	MLP	18.01	0.85	0.72	11.39	0.29
29113	TLFN	28.51	0.81	0.66	21.49	0.64
	BNN	34.3	0.72	0.52	25.48	0.92
	RNN	28.3	0.75	0.56	21.51	0.63
	MLP	28.91	0.77	0.59	20.85	0.66
29114	TLFN	13.6	0.83	0.69	10.29	0.37
	BNN	15.38	0.8	0.64	11.41	0.47
	RNN	13.75	0.81	0.66	9.84	0.38
	MLP	13.91	0.83	0.69	10.19	0.39
29102	TLFN	39.09	0.55	0.3	30.22	0.7
	BNN	40.73	0.54	0.29	29.6	0.72
	RNN	39.17	0.54	0.29	29.43	0.7
	MLP	39.97	0.53	0.35	29.6	0.72

7.1.3 Seasonal variation

Owing to the variation of TSP concentrations in particular seasons, the performances of the models were also compared with their seasonal bias (%). Table 7.3 and Fig. A. 3 (Appendix) show a comparative assessment of the seasonal biases for TLFN, BNN, RNN and MLP models. For site 29000, the seasonal bias of winter TSP levels generated by the models range from an under-prediction of -7.69 % to an over-prediction of 3.2%. During spring when TSP concentration remains high in air, all the models tend to under-predict. BNN outperformed other three models during spring and summer with underestimation of -2.04% and overestimation of 0.18% respectively. Important observation here is except

MLP model, all models tend to over-predict during summer. On the other hand, autumn was the worst season compared to other seasons where over-prediction was 5%.

In case of site 20025, BNN performed well during winter and RNN performed well during spring and autumn. Important to note here that the performance of RNN model during autumn was more than 5% better compared to other models. Its performance during spring is also better than the rest three models (under-prediction of 3.31%). The bias generated by all the models during winter is pretty higher at site 29102.

All models largely under-predicted which ranges from 15 to 30%. Summer is the only season where the model performances are superior to the other seasons with underestimation of less than 5% by each model. Unlike the above mentioned sites, the biases generated at 29113 showed larger variations; BNN was the only model to produce less than 15% of over-prediction. MLP performed well during spring and autumn with RNN during summer. Interestingly, all models at site 29114 over-predicted in all seasons and the comparison showed that RNN largely outperformed other three models in each season.

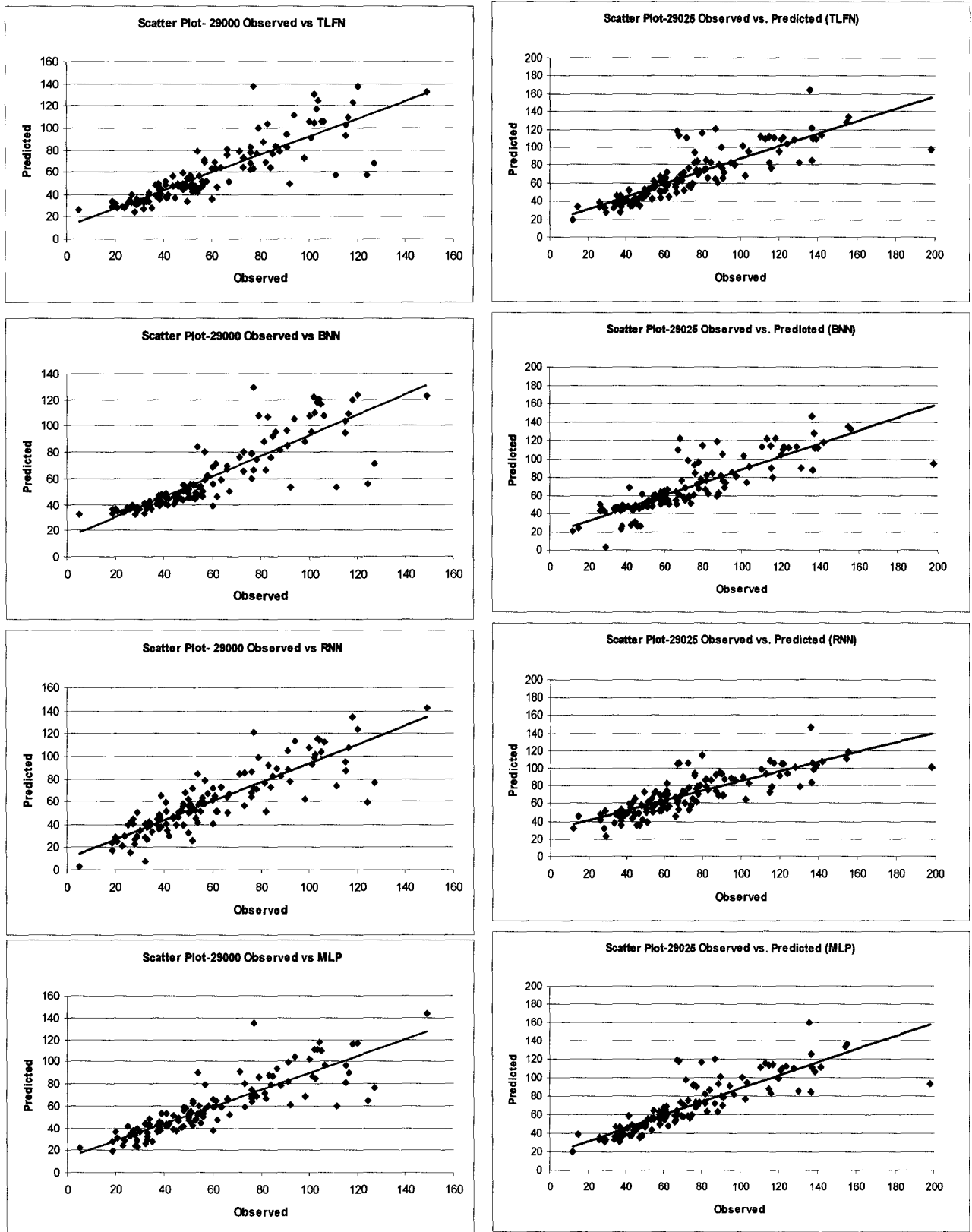


Fig. 7.1: Scatter plots of observed vs. simulated values of TSP concentration at station 29000 (left) and 29025 (right)

Table 7.3: Comparison of seasonal bias at station 29000, 29025, 29102, 29113 and 29114

Station	Model	Seasonal bias (%)			
		Winter	Spring	Summer	Autumn
29000	TLFN	-3.87	-3.72	1.2	6.5
	BNN	4.98	-2.04	0.18	10.51
	RNN	-7.69	-3.55	2.78	14.29
	MLP	3.2	-7.39	-4.58	7.85
29025	TLFN	-11.63	-5.87	0.54	-7.17
	BNN	-9.83	-5.19	3.83	-7.42
	RNN	-12.97	-3.31	7.31	-1.47
	MLP	-10.2	-4.13	3.05	-6.61
29102	TLFN	-18.73	26	-2.78	3.98
	BNN	-27.64	16.12	-5.66	3.86
	RNN	-20.09	29.72	-1.89	4.61
	MLP	-17.88	30.43	-1.27	6.11
29113	TLFN	48.43	32.3	45.28	31.56
	BNN	14.67	33.74	22.04	18.27
	RNN	31.32	22.96	15.03	13.82
	MLP	32.72	22.72	22.35	12.53
29114	TLFN	4.79	19.3	13.05	9.61
	BNN	6.88	28.31	17.06	11.31
	RNN	0.44	18.15	4.62	4.65
	MLP	3.23	24.64	13.27	9.72

7.1.4 Confidence interval with BNN

A comparison of the simulated concentrations during 2004 with the 95% confidence level of BNN models has been presented in Fig. 7.4 and Fig. A. 2 (a) through (c). Fig. 7.4 and appendix A. 2 (a) compare TLFN, RNN and MLP models with the BNN model at sites 29000 and 29025. The uncertainty bands created by the BNN models have been able to capture the observed and simulated values created by the other NN models. Representation of the confidence intervals about mean estimates is the additional advantage of BNN model which the conventional neural network models cannot provide (Khan and Coulibaly, 2006). On the other hand, at site 29113, TLFN model went outside the upper band and had suffered severe over-prediction bias. At site 29114, the observed and predicted values are within the band in 98% cases. But the wider uncertainty band indicates the degradation of the BNN model performance as well.

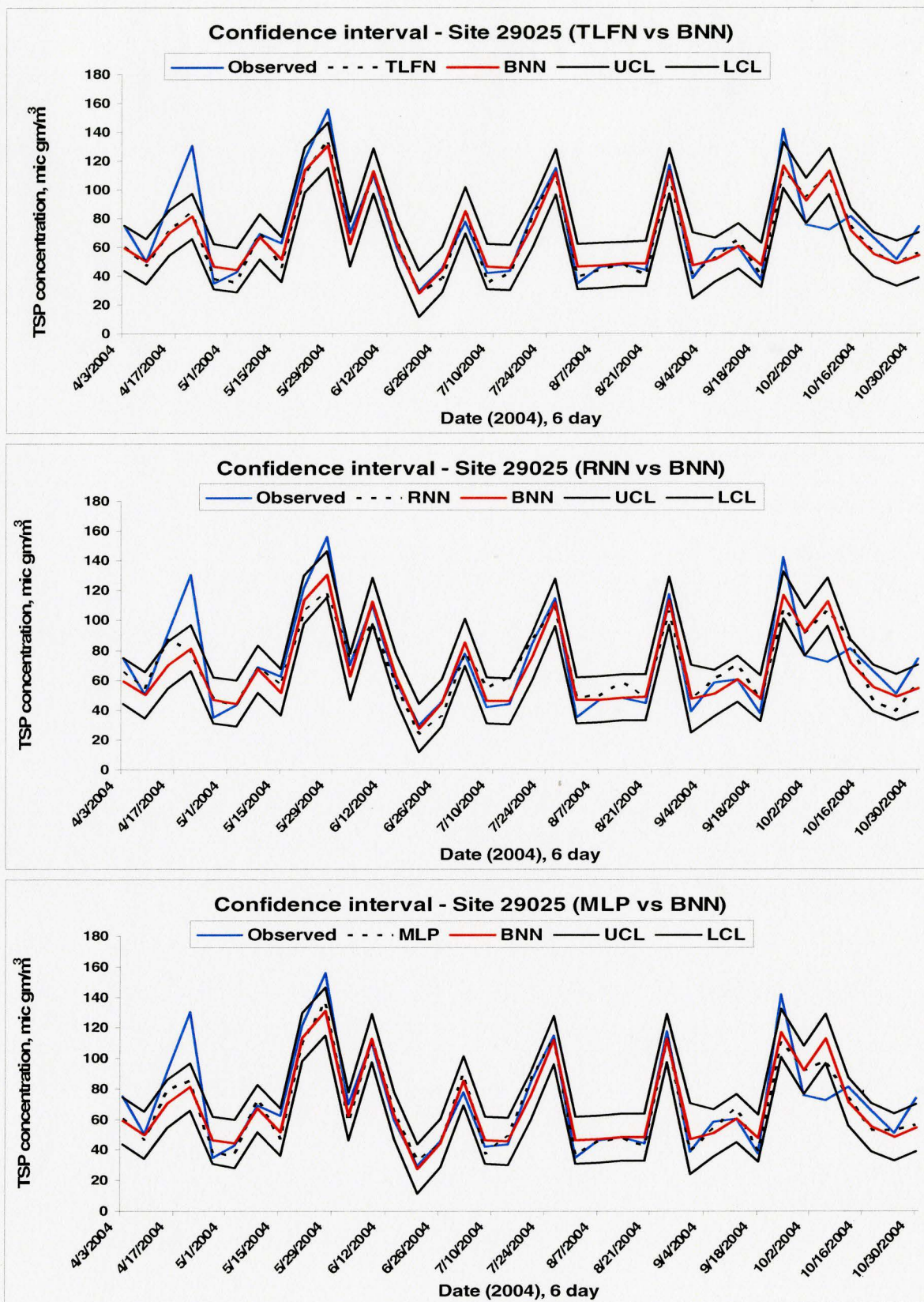


Fig. 7.2: Simulation result of TSP at station 29025 with 95% confidence interval

7.1.5 Spatial interpolation

In this study ordinary kriging has been applied because it is the most widely used kriging technique. Universal kriging is not appropriate in this case because no trend in data has been found. The reason of not using indicator kriging is that indicator kriging gives an estimation of a distribution of values within an area rather than the mean value of the area. A comparison of the krigged surface generated by the observed and simulated models has been presented in Fig. 7.5 and Fig. A. 3 (Appendix). Fig. 7.5 presents a comparison of the observed and simulated concentrations over space during spring 2003. The spatial distribution of the 5 sampled location indicates that the locations near site 29025 had higher concentrations of TSP ranging from 90-100 $\mu\text{ gm/ m}^3$ while the surfaces created by TLFN, BNN, RNN and MLP models has shifted the high concentration zone towards right around site 29113. The reason behind this shifting is that each model produced higher positive bias while simulating spring values; this over-prediction however has been largely reflected in the spatially interpolated area. On the other hand, the surface created around site 29114 shows relatively low concentration area and has been properly mapped using TLFN and MLP generated values. In case of fall 2003 (Fig. A 3 (a) and (b)), the performance of TLFN model is quite poor compare to other models, MLP model in this case performed better than other three models. Recall that no forecasting, rather a simulation has been done and a better performance has been achieved using MLP in this case.

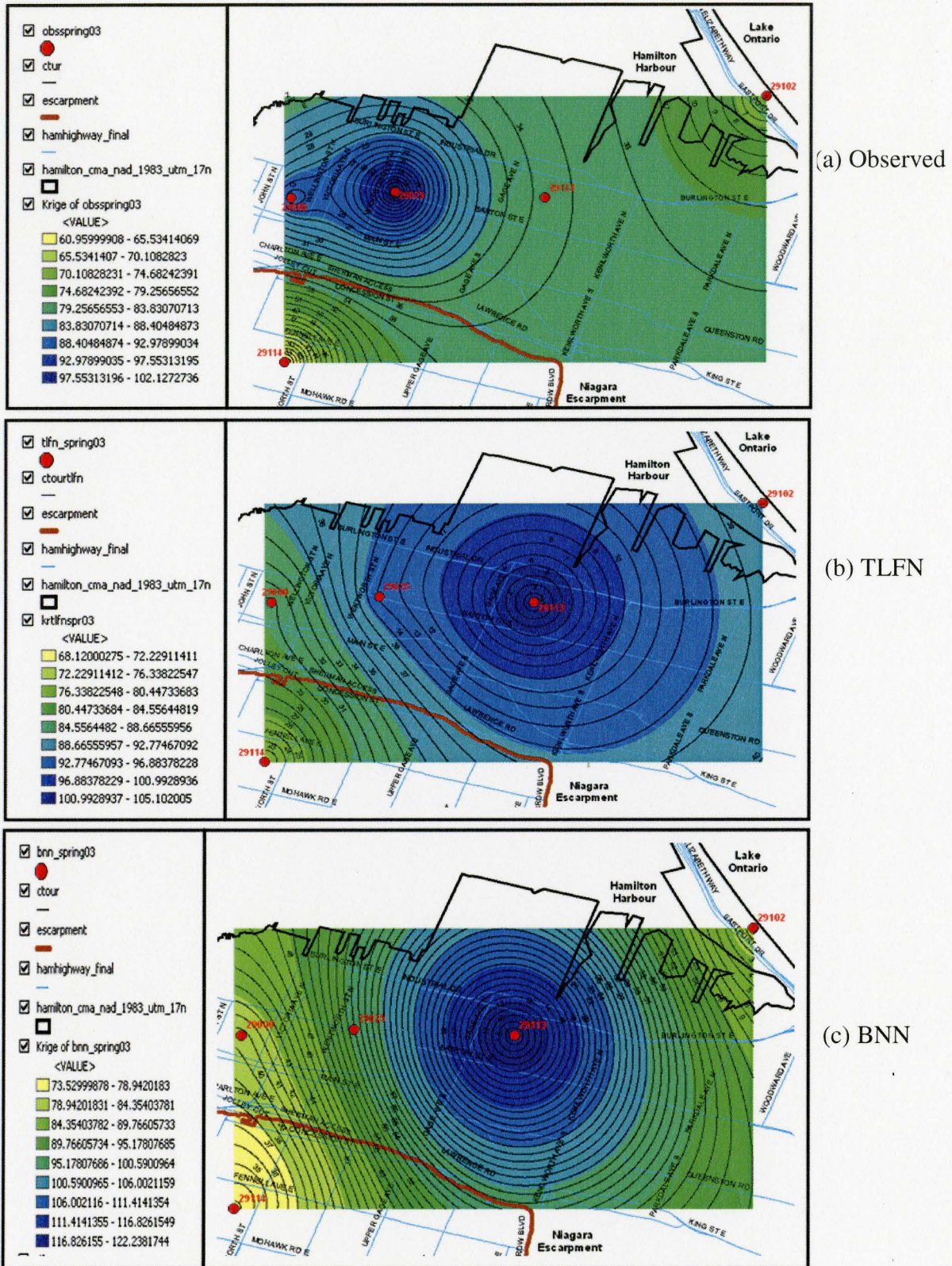
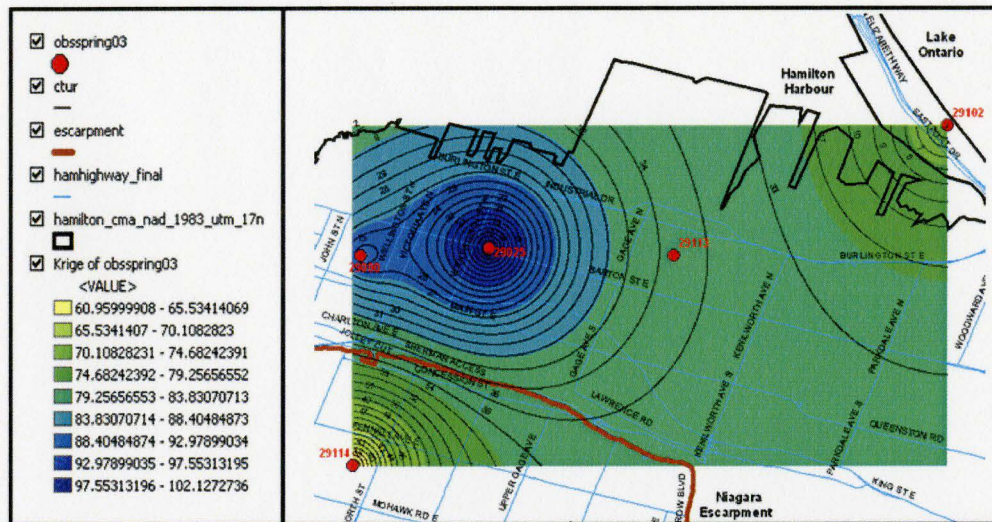
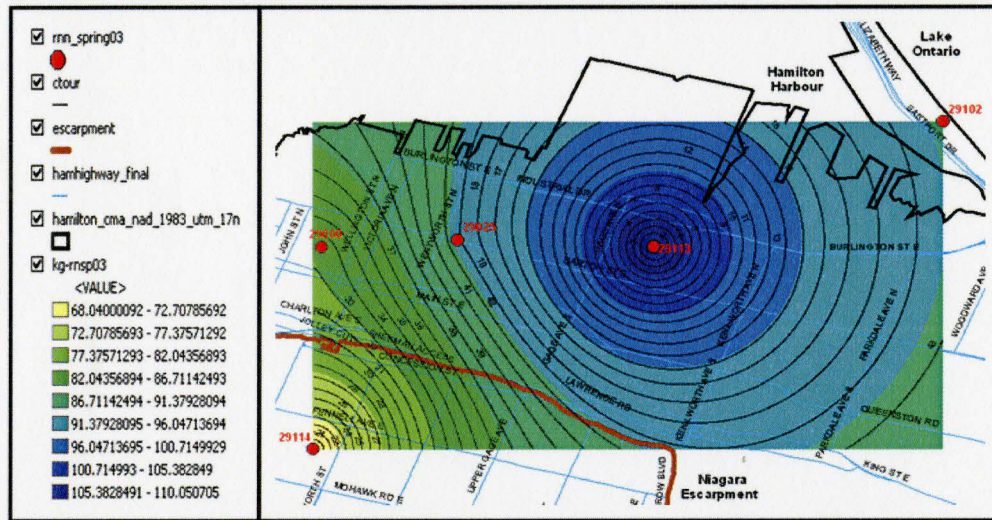


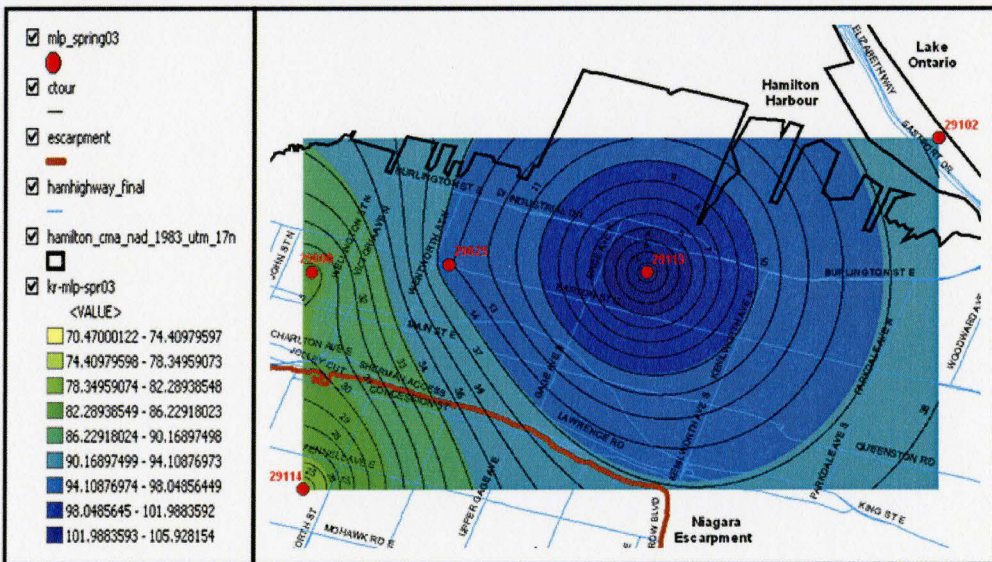
Fig. 7.3 (a): Comparison of observed and simulated TSP values over space during spring 2003



(a) Observed



(b) RNN



(c) MLP

Fig. 7.3 (b): Comparison of observed and simulated TSP values over space during spring 2003

7.2 Simulation of PM₁₀

7.2.1 Combination of input variables

Similar to TSP, the different combinations of input variables have been tested to get optimal input variables. But in this case two stations have been tested with TLFN model only. A comparison of these model performances has been presented in Table 7.4. Like TSP, the model simulating statistics for site 300 shows that Case 1, (i.e. using nearby sample sites as input variables) gave best performance. Interestingly for site 302, a combination of meteorological variables, nearby site's PM₁₀ concentrations together with land use variables have been able to largely improve the models than other combinations. The r values increased almost 50% compared to other combinations, and the RMSE value improved by 25% compared to other cases.

Table 7.4: Combination of input variables for PM₁₀: TLFN model performance

Target stations	Input variables	Model performance statistics				
		RMSE	r	R ²	MAE	NMSE
300	Case 1: Stations together	9.00	0.72	0.52	6.78	0.50
	Case 2: Met variables	12.17	0.36	0.13	9.96	0.91
	Case 3: Land use + stations	9.00	0.71	0.5	5.95	0.5
	Case 4: Met variables + land use + stations	10.00	0.63	0.4	6.45	0.62
302	Case 1: Stations together	20.63	0.4	0.16	13.96	0.91
	Case 2: Met variables	20.76	0.29	0.08	13.85	0.92
	Case 3: Land use + stations	20.81	0.36	0.13	15.3	0.93
	Case 4: Met variables +land use + stations	16.18	0.71	0.5	12.13	0.56

7.2.2 .Model performance statistics

The performances of TLFN, BNN, RNN and MLP models have been compared at sites 300, 302 and 313. The PM₁₀ monitoring sites used in this study are identical to the 3 sites collecting TSP concentrations: site 300 corresponds to the Hamilton Downtown site (ID 29000 in case of O₃, NO₂ and TSP), 302 stands for Beach Boulevard (29102 for TSP) and 313 stands for Gertrude/Depew (29313) station. Table 7.4 represents a comparative

assessment of the performances of the 4 models tested. At site 300, the performances of all the models are quite competitive; overall RNN models appeared better than other 3 models in simulating the PM_{10} values at this site. At site 302, TLFN model outperformed BNN, RNN and MLP models, the performances show that only TLFN model has been able to produce an r value greater than 0.70, the $RMSE$ and $NMSE$ values were also quite less ($RMSE$: $16.18 \mu\text{ gm/ m}^3$ and $NMSE$: 0.45). Finally site 313 also generated similar types of results as site 302; a comparison of model statistics shows that TLFN model performed slightly better than rest 3 models.

Table 7.5: Comparison of model performance for PM_{10}

Target station	Model	Model performance statistics					
		RMSE	r	R^2	MAE	NMSE	RB
300	TLFN	9.00	0.72	0.52	6.78	0.50	0.057
	BNN	9.34	0.68	0.46	6.37	0.53	-0.010
	RNN	8.57	0.74	0.55	6.35	0.45	0.027
	MLP	8.89	0.72	0.52	5.63	0.49	-0.037
302	TLFN	16.18	0.71	0.50	12.13	0.56	-0.144
	BNN	17.42	0.62	0.38	11.92	0.64	-0.115
	RNN	17.51	0.66	0.43	12.15	0.66	-0.149
	MLP	16.73	0.68	0.46	10.98	0.59	-0.141
313	TLFN	12.19	0.66	0.44	9.02	0.58	0.056
	BNN	12.99	0.64	0.41	9.16	0.64	0.080
	RNN	13.09	0.59	0.35	9.96	0.66	0.032
	MLP	12.99	0.63	0.40	9.99	0.65	0.107

The visual assessment of the performances of each model is then performed using scatter plots. The observed versus simulated values for the test period i.e. 2003-2004 are plotted in Fig. 7.6 and Fig. A. 4. It is seen that all the models had suffered some under-prediction or over-prediction in predicting higher and lower concentrations. The under-prediction is predominant at site 300 and 302. On the other hand in case of site 313, some over-prediction is caused by TLFN and BNN models especially for simulating the low concentrations.

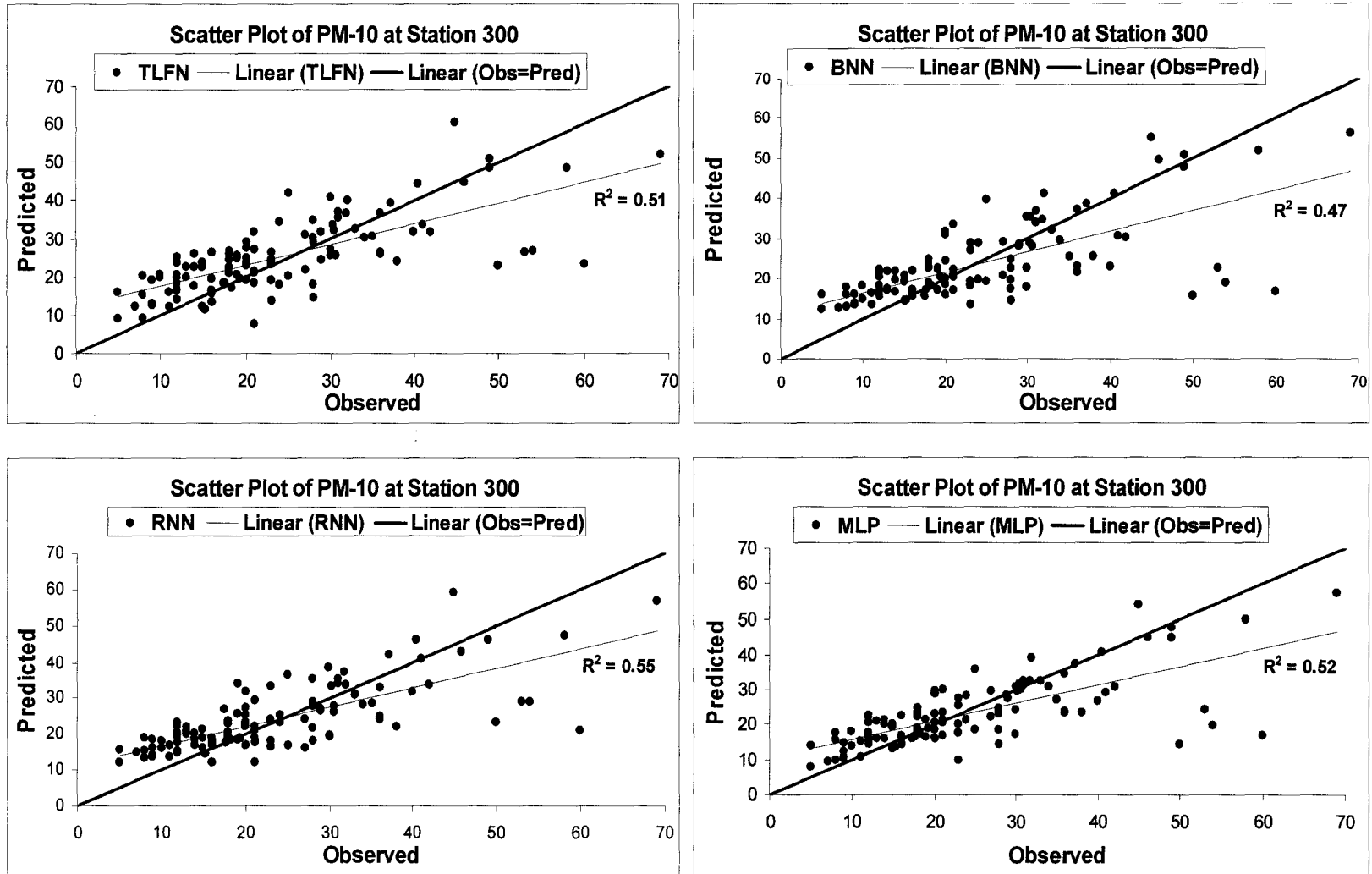


Fig. 7.4 (a): Scatter plots for PM₁₀ concentration at station 300

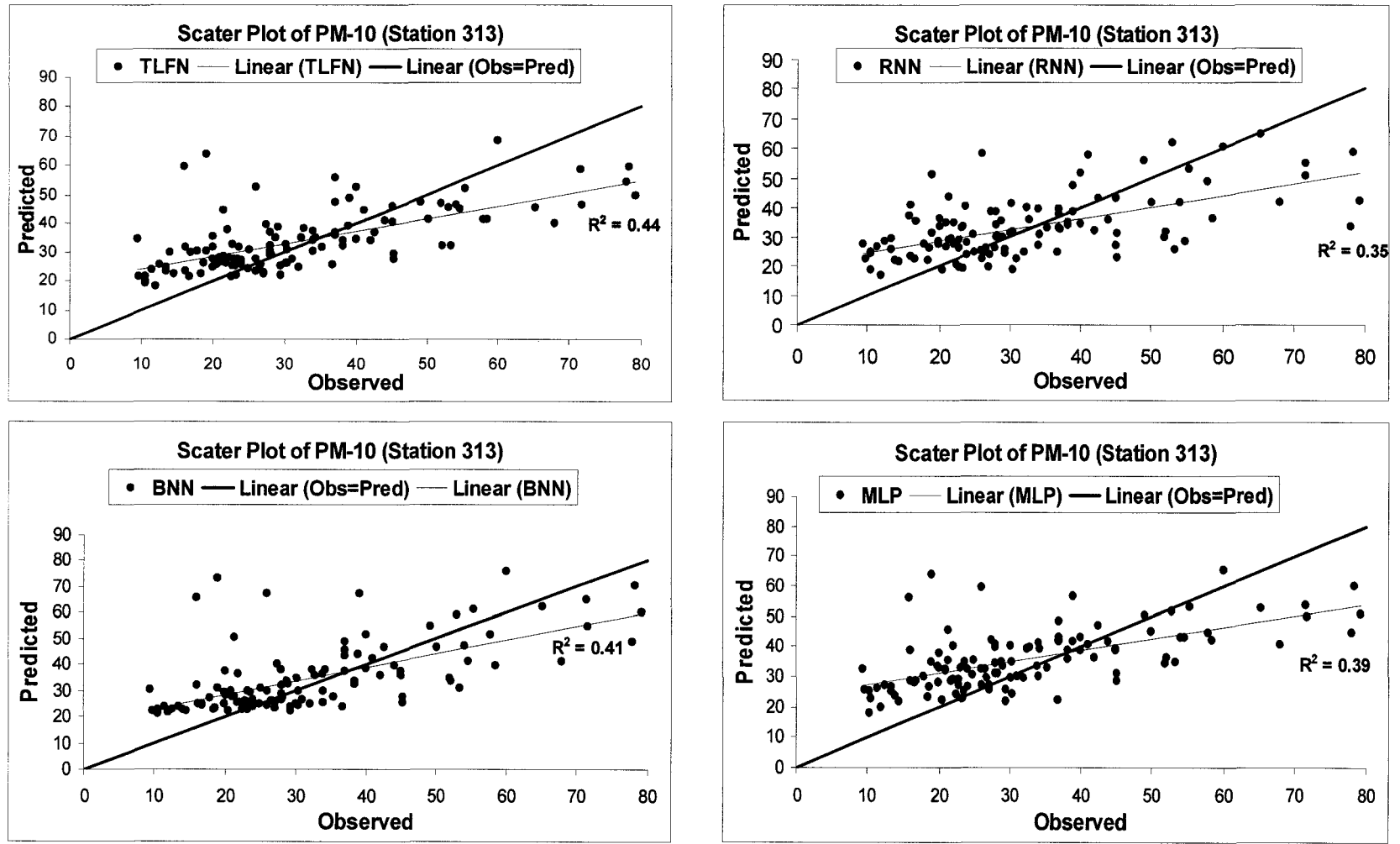


Fig. 7.4 (b): Scatter plots for PM₁₀ concentration at station 313

7.2.3 Seasonal variation

The performances of the models are further compared in Table 7.6 and Fig. A. 7 (Appendix) based on their performance by season. A comparative assessment of the seasonal bias generated by each models have been presented. The maximum acceptable bias within 15% is specified by the author in this study. At site 300 it is seen that during spring, the performances of the models are inferior compared to other seasons. All models generated lower simulated values than the observed ones. Especially the performances of MLP and BNN were poor causing more than 15% bias in this season. During summer, the MLP, RNN and BNN models performed quite well, producing less than 7% of biases. Only TLFN model generated comparatively higher bias (13.56%). Again during fall, the bias level was high; only MLP has been able to generate a bias less than 10%. Finally during winter, MLP again outperformed other three models with almost 50% less biases than those models. So finally, MLP model performed better during summer, fall and winter while RNN showed good performances during spring and summer.

At site 302 which is a residential area, RNN outperformed other three models during spring, summer and fall while TLFN model showed better performance during winter with only 6.57% bias. Finally at the industrial site 313, the bias generated by the models are quite high compared to two other sites especially during summer and fall with under-predictions ranging between 15-26% during summer and 17%-26% during fall. Spring is the only season where the models performed well with only RNN model exceeding more than 10% biases.

Table 7.6: Comparison of seasonal bias at station 300, 302 and 313

Station	Model	Seasonal bias (%)			
		Spring	Summer	Fall	Winter
300	TLFN	-9.43	13.56	20.42	13.12
	BNN	-16.24	6.63	13.79	13.47
	RNN	-8.29	5.87	13.87	12.63
	MLP	-15.55	3.86	8.29	6.75
302	TLFN	7.86	0.83	10.59	6.57
	BNN	12.2	2.46	9.41	13.84
	RNN	4.39	0.43	3.91	13.58
	MLP	10.88	6.37	15.03	20.55
313	TLFN	0.36	-19.74	-20.38	-9.28
	BNN	4.13	-15.04	-17.07	-13.46
	RNN	14.42	-25.22	-25.08	-10.65
	MLP	1.81	-20.77	-18.97	-11.7

7.2.4 Confidence interval with BNN

The performances of the simulated concentrations during 2004 have been compared with the 95% confidence interval of the BNN models in Fig. 7.7 and Fig. A. 5. From the figures it is seen that in case of site 300, the simulated values of TLFN, RNN and MLP models are within the upper and lower limits of BNN models. At site 302, none of the models have been able to capture the higher concentrations, even the upper confidence values predicted by BNN model are well below the observed peak concentrations. Finally at site 313, the observed peak concentrations are slightly above the upper limit, but not severe as site 302. Importantly, the relatively wider bands generated by the BNN model at sites 302 and 313 indicates higher uncertainty. This may be due to the fewer number of extreme values in the training sample.

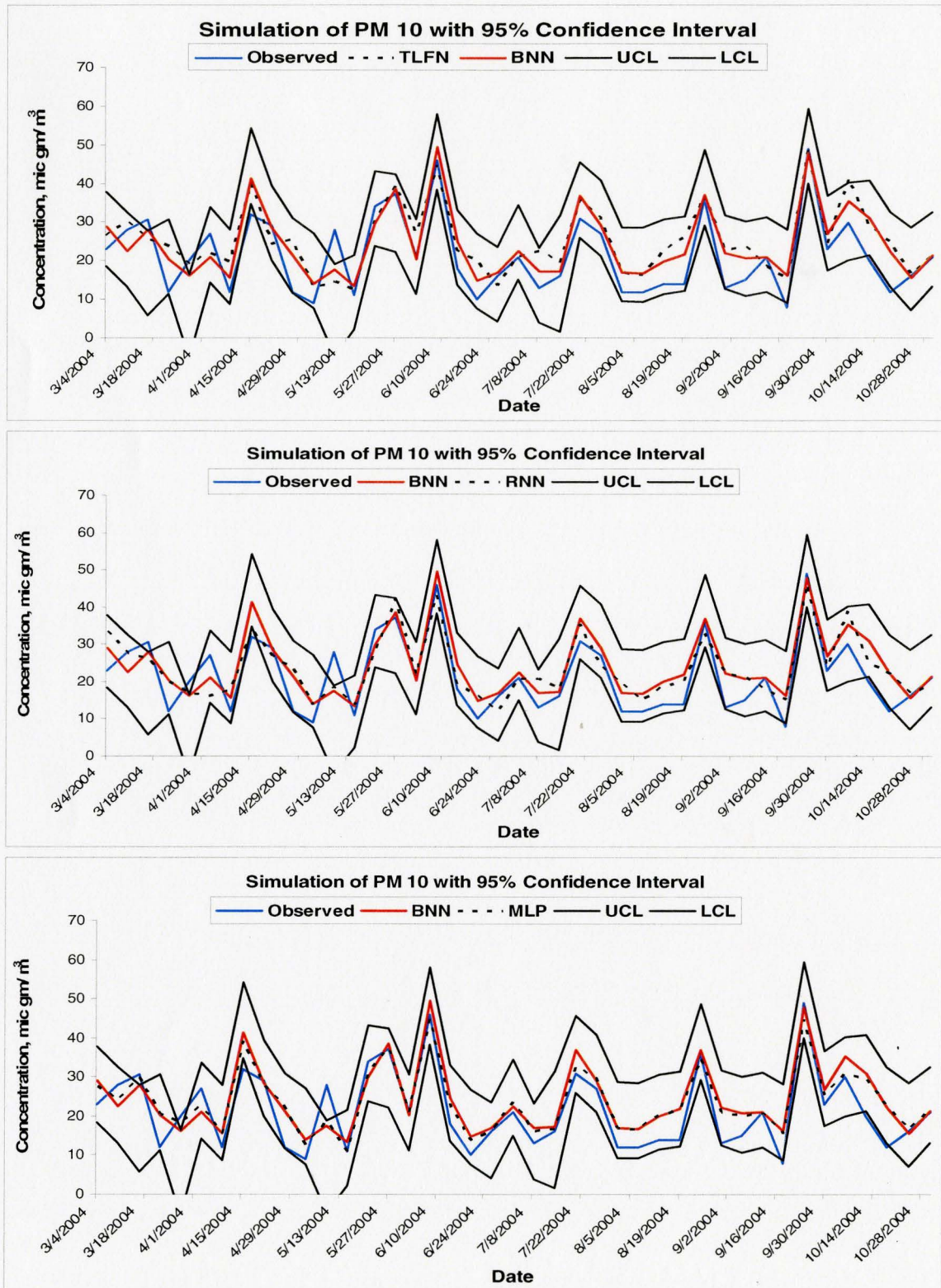


Fig. 7.5: Simulation result of PM₁₀ at station 300 with 95% confidence interval

7.3 Summary of TSP and PM₁₀ results

Overall performances of the four neural network models applied in simulating total suspended particulates and PM₁₀ concentrations at some selected sites in Hamilton are very competitive. Firstly TSP concentrations at 5 monitoring locations located in different land use type areas have been analyzed. In order to get best possible outputs, the input variables are carefully chosen by taking several combinations of input variables. It is seen that except site 29102, the remaining four stations located nearer to each other behaved in a similar manner indicating that the TSP concentrations fluctuate rapidly and are greatly influenced by the sites around which are to a large extent driven by the meteorology. The performances for each season at the stations are different: TLFN model performed well at site 29000; RNN model was good during spring and autumn at site 29025 and RNN model again outperformed all other models in all seasons at site 29114. Inversely, all the models provided higher biased results at site 29113 where the seasonal bias exceeded 15%. The overall performances of the sites further indicate that the performances of all the models are pretty close to each other. MLP model performed very well, even in some cases it outperformed TLFN and RNN models. So it can be said that this model should be able to simulate acceptable concentrations at those sites.

In simulating PM₁₀ concentrations at 3 monitoring sites, RNN and MLP worked well at site 300 while RNN again produced less bias at site 302. Similarly at site 313, higher biases are generated indicating that none of these models have been able to simulate the concentrations well. Similar to TSP, it also appears that it is possible to achieve competitive simulation results using MLP model; hence there may not be any need of applying more complex neural network models in the cases where performances of MLP is similar to the other three models.

Chapter 8

Conclusions and Recommendations

8.1 Conclusions

This research provides the most extensive evaluation of neural network models currently available for the prediction and simulation of some major air pollutants such as ground level ozone (O₃), nitrogen dioxide (NO₂), total suspended particulates (TSP) and inhalable particulate matters (PM₁₀) concentrations. The study is unique because of the variety of neural network models evaluated and the amount of the input combinations used. In the past, several studies have compared linear statistical methods with widely used static multilayer perceptron models to predict or simulate pollutant concentrations and have concluded that the MLP model has the ability to capture the complex non-linear relationships of the pollutants and meteorological variables which most linear models cannot do.

This study provides a more comprehensive comparison of NN based models. The widely used MLP model has been set as a bench-mark and three emergent neural network models such as time lagged feed-forward neural network (TLFN), Bayesian neural networks (BNN) and recurrent neural network (RNN) models have been tested for short term ground level O₃ and NO₂ forecasting and for simulating TSP and PM₁₀ concentrations. The major aim of the study was to investigate and characterize the complex nonlinear temporal and spatial variability of the pollutants and their dependencies on the surrounding meteorology, and develop an air quality forecasting tool based on most appropriate neural networks. Due to data constraint, only O₃ and NO₂ forecasting tools have been developed. The newly developed models were then used to forecast up to 24 hour ahead. To achieve this goal, the TLFN, BNN, RNN and MLP network architectures and different input combinations were compared in order to

identify the optimal model structure. The optimal models identified are then used to forecast or to simulate values for each monitoring stations which are then interpolated spatially to explore their capability to simulate pollutant values at stations where sampling has not been done.

Firstly, the forecasting models have been run for three ozone monitoring sites (Hamilton Downtown (29000), Hamilton Mountain (29114) and Hamilton West (29118)) and at two nitrogen dioxide monitoring sites (Hamilton Downtown (29000) and Hamilton West (29118) within Hamilton census metropolitan area. In this case because of limited number of stations, spatial interpolation could not be done. Overall performances of the O₃ forecasting models show that all 4 models gave competitive results up to 6 steps i.e. 12 hour ahead. Overall TLFN and RNN model performed better than BNN and static MLP model. Seasonal comparison of model forecasting performances showed that none of the models perform very well in the summer season. So in order to improve summer time forecasting, it was separately modeled and the results obtained from this summer model were then compared with the annual model performances. The results indicated slight improvement, hence it can be concluded that the annual model itself has the ability to project the underlying seasonal fluctuations in concentrations. However, the reason behind similar performances of annual and summer model may be due to the absence of more appropriate variables such as solar radiation and maximum temperature which are believed to be the two most influential parameters behind the high O₃ concentrations during summer period. Furthermore, land use types around the monitoring sites may also affect pollutant concentrations. Hence in addition to the meteorological variables, the land use variables were added to improve the model forecasting performance. Surprisingly the inclusion of land use variables with meteorology could not to improve the model performances rather it increased the computational time and cost. This demonstrates that in temporal problems, the inclusion of land use types in a form of logical input may not be the right form to capture the impacts of different land use types rather using emission factors from different sources may be a good alternative to count

the influences of different sources of the pollutants. and finally it is established that the basic model using appropriate meteorological variables are appropriate for short term forecasting. These results further suggest that the inclusion of time delay in and/or adaptive memory (context unit) in MLP have the capacity to improve the results obtained from conventional static neural network (MLP in this case). These performances, however indicate that RNN model has the best generalization performance and suggest that the relationship between ozone and the meteorology can better be represented using these selected predictor variables and a dynamically driven neural network (RNN). Forecasting results for NO₂ shows very similar performances. Interestingly static multilayer perceptron model has performed quite competitively with the advanced TLFN and RNN models. Hence applying MLP with relevant variables would be enough to achieve good forecasting results with much simplicity and less computational cost.

Finally TSP concentrations from five monitoring sites at Hamilton Downtown (29000), Hamilton Mountain (29114), Barton/Sanford (29025), Beach Boulevard (29102) and Gertrude/Depew (29113) were modeled. The model identification process followed is similar to the one used for O₃ and NO₂ models. After optimal models have been identified and calibrated, the simulated values achieved from TLFN, BNN, RNN and MLP models were then interpolated over space to obtain concentrations at unsampled sites and to compare their performances. In case of inhalable particulate matters (PM₁₀), data from three stations such as Hamilton Downtown (29000), Beach Boulevard (29102) and Gertrude/Depew (29113) were considered, and model performances were compared to obtain the optimal model results. Overall performances shown by the four models for both pollutants are pretty close. It has been noticed that the particulate concentrations are largely dependent on the surrounding pollutant concentrations which are triggered up due to meteorological conditions. Like NO₂ models, MLP also performed competitively with other NN models, even in some cases it outperformed TLFN and RNN models. So it can be said that this model should be able to simulate acceptable concentrations at the selected sites in Hamilton. Seasonal analysis revealed that spring has been the most

challenging season for simulating concentrations. Overall it appears that the weather based particulate simulation is promising, which highly depends on data availability. The results obtained also indicate that the day-to-day fluctuations of the particulates concentrations caused by several factors in the Hamilton CMA are to a large extent driven by meteorological factors and their sources.

The interpolated TSP surfaces indicate that when enough spatial data is available, the neural network models, even a simple MLP can be a good tool to generate concentrations at unsampled sites.

In conclusion, the neural network models used in this study have performed well since they are unconstrained and allowed arbitrary interactions and nonlinear relationships between predictor variables. To be more precise, time lagged feed-forward network (TLFN) and recurrent neural network (RNN) model has shown better performance in case of ground level ozone forecasting at three ozone monitoring sites in Hamilton which further suggests that the inclusion of time delay and/or adaptive memory (context unit) in MLP have the capacity to improve the results obtained from conventional static neural network. Interestingly the multilayer perceptron (MLP) model competed well with the rest three dynamic neural network models for nitrogen dioxide forecasting; hence it appears that applying MLP with most relevant variables can be enough to get better NO₂ forecasting result. Similar to the NO₂ forecasting results, the TSP and PM₁₀ model simulation results revealed that MLP model performed competitively and even in some cases it outperformed TLFN and RNN models. So it can be concluded that this model should be able to simulate acceptable TSP and PM₁₀ concentrations at the study sites in Hamilton metropolitan area.

8.2 Recommendations for future work

This thesis has concentrated on neural network architectures and configurations since no other work with similar objectives has been conducted for the study area. However, this has been a preliminary assessment process in order to develop a robust and appropriate neural network model for air quality modeling. Due to time constraints all kinds of neural network models could not be tested. It would be ideal if an uncertainty analysis could have been conducted to further assess model results accuracy. This kind of test could be useful in air quality modeling where uncertainty can be of significant concern.

Secondly, the present research concentrated only on NN models; comparing the present results with other statistical models would make the present results more complete. Further investigation could also be aimed towards employing other neural network types with different settings and options like self organizing maps.

Apart from these, there are several issues that can be tested further:

1. Try to improve ozone forecasting performance using more appropriate weather variables such as max temp, min temp and solar radiation and emission factors.
2. In case of TSP and PM₁₀, it is very important that a large database be available in order to capture the random variation of weather-pollutant relationships.
3. In order to improve the forecasting ability of the pollutants, it is necessary to incorporate additional climatic information, such as weather classification indicators, cloud cover height, boundary layer information, wind profile, opacity, discomfort index, and non-climatic factors such as traffic levels and indices of heavy and low traffic conditions.

References

- Abdul-Wahab, S. A., Bakheit, C.S., and Al-Alawi, S. M., (2005), Principal component and multiple regression analysis in modeling of ground level ozone and factors affecting its concentration, *Environmental Modeling and Software*, Vol. 20, pp.1263-1271.
- Agirre-Basurko, E., Ibarra-Berastegi, and G., Madariaga, I., (2006), Regression and multilayer perceptron-based models to forecast hourly O₃ and NO₂ levels in the Bilbao area, *Environmental Modelling and Software*, Vol. 21, pp. 430-446.
- Athanasiadis, I.N., Karatzas, K., and Mitkas, P., (2005), Contemporary air quality methods: A comparative analysis between classification algorithms and statistical methods, *Proceedings of 5th Int'l conference on Urban Air Quality Measurement, Modeling and Management*, Valencia, Spain, March 2005.
- Benvenuto, F., and Martini, A., (2000), Neural networks for environmental problems: data quality control and air pollution nowcasting, *Global nest: the intl J.*, Vol. 2 (3), pp. 281-292.
- Bishop, C. M., (1995), *Neural Networks for Pattern Recognitions*, Clarendon, Oxford, U.K.
- Blair, R., (2006), Meteorological variations and their impact on NO₂ concentrations in the Toronto-Hamilton urban air-shed, Canada, M.Sc. Thesis, School of Geography and Earth Sciences, McMaster University, Hamilton, Canada.

- Bodri, L., and Cermak, V., (2000), Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia, *Advances in Engineering Software*, Vol. 31, pp. 311-321.
- Bordignon, S., Gaetan, C., and Lisi, F., (2002), Nonlinear models for ground level ozone forecasting, *Statistical Methods & Applications*, Vol. 11, pp. 227-245.
- Buzzelli, M., and Jerrett, M., (2004), Racial gradients of ambient air pollution exposure in Hamilton, Canada, *Environment and Planning A*, Vol. 36, pp. 1855-1876.
- Chaloulakou, A., Saisana, M., and Spyrellis, N., (2003), Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens, *The science of the Total Environment*, Vol. 313, pp. 1-13.
- Chelani, A.B., (2005), Predicting chaotic time series of PM₁₀ concentration using artificial neural network, *International Journal of Environmental Studies*, Vol. 62 (2), pp. 181-191.
- Chelani, A.B., Rao, C.V.C., Phadke, and K.M., Hasan, M. Z., (2002), Prediction of sulphur dioxide concentration using artificial neural networks, *Environmental Modelling and Software*, Vol. 17 (2), pp. 161-168.
- Comrie, A. C., (1997), Comparing neural networks and regression models for ozone forecasting, *Air and Waste Management Association*, Vol. 47, pp. 653 – 663.
- Coulibaly, P., Abctil, F., and Bobée, B., (2001) a, ANN modeling of water table depth fluctuations, *Water Resources Research*, Vol. 37 (4), pp. 885-869.

- Coulibaly, P., Anctil, F., and Bobee, B., (2001) b, Multivariate reservoir inflow forecasting using temporal neural networks, *Journal of Hydrologic Engineering*, Vol. 6 (5), pp. 367-376.
- Daliakopoulous, J., (2004), Groundwater level forecasting using artificial neural networks, M.A.Sc. thesis, McMaster University, Hamilton, Ontario, Canada.
- Derwent, R.G., Simmonds, P. G., Seuring, S., and Dimmer, C., (1998), Observation and interpretation of the seasonal cycles in the surface concentrations of ozone and carbon monoxide at Mace Head, Ireland from 1990 to 1994. *Atmospheric Environment*, Vol. 31 (13), pp. 145-157.
- Dutot, A.L., Rynliewicz, J., Steiner, F.E., and Rude, J., (2006), A 24-h forecast of ozone peaks and exceedance levels using neural classifiers and weather predictions, *Environmental Modelling and Software* (2006), Doi: 10.1016/j.envsoft.2006.08.002.
- Dibike, Y. B. and Coulibaly, P., (2006), Temporal neural networks for downscaling climate variability and extremes, *Neural Networks*, Vol. 19, pp.135-144.
- Elman, J. L., (1990), Finding structure in time, *Cognitive Science*, Vol. 14, pp. 179-211.
- Elminir, H. K. and Abdel-Galil, (2006), Estimation of air pollutant concentrations from meteorological parameters using artificial neural network, *Journal of Electrical Engineering*, Vol. 57 (2), pp. 105-110.
- Fausett, L. (1994), *Fundamentals of Neural Networks*, Prentice Hall, Englewood Cliffs, NJ.

- Frasconi, P., Gori, M., and Soda, G., (1992), Local feedback multilayered networks, *Neural Computation*, Vol. 4, pp. 120-130.
- Freund, Y., and Schapire, R.E., (1997), A decision- theoretic generalization of on-line learning and an application to boosting, *Journal of Computer and System Sciences*, Vol. 55, pp. 119-139.
- Gardner, M. W., and Dorling, S. R., (1998), Artificial neural networks (the multilayer perceptron)--a review of applications in the atmospheric sciences, *Atmospheric Environment*, Vol. 32 (14-15), pp. 2627-2636.
- Gardner, M. W., and Dorling, S. R., (1999), Neural network modelling and prediction of hourly NO_x and NO₂ concentrations in urban air in London, *Atmospheric Environment*, Vol. 33 (5), pp. 709-719.
- Gardner, M. W., and Dorling, S. R., (2000), Statistical surface ozone models: an improved methodology to account for non-linear behavior, *Atmospheric Environment*, Vol. 34, pp. 21-34.
- Hagan, M., Demuth, H., and Beale, M., (1995), *Neural Network Design*, PWS Publishing Company, Boston.
- Haykin, S., (1999), *Neural Networks: A Comprehensive Foundation*, Upper Saddle River, NJ: Prentice-Hall.
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., and Brasseur, O., (2005), A neural network forecast for daily average PM₁₀ concentrations in Belgium, *Atmospheric Environment*, Vol. 39, pp. 3279-3289.

Intergovernmental Panel on Climate Change (2007), *Climate change 2007: The physical science basis, Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change*, World Meteorological Organization, Geneva, Switzerland.

Jacob, D. J., Horowitz, L. W., Munger, J. W., Heikes, B. G., Dicherson, R. R., Artz, R. S., and Keene, W. C., (1995), Seasonal transition from NO_x – to hydrocarbon-limited ozone production over the eastern United States in September, *J. Geophys. Res.*, Vol. 100, pp. 9315-9324.

Hirsch, A. I., Munger, J. W., Jacob, D. J., Horowitz, L. W., and Goldstein, A. H., (1996), Seasonal variation of the ozone production efficiency per unit NO_x at Harvard forest, Massachusetts, *J. Geophys. Res.*, Vol. 101 (7), pp. 12659-12666.

JANN (Java Artificial Neural Network): A neural tool for air pollution modeling” developed at the University of Catania in the framework of the APPETISE project.,retrieved on 05/04/2007 from <http://www.dees.unict.it/users/gnunnari/appetise/jann/index.html>.

Jerrett, M., Burnett, R., Kanaroglou, P., Eyles, J., Finkelstein, N., Giovis, C., and Brook, J., (2001), A GIS-environmental Justice Analysis of particulate Air Pollution in Hamilton, Canada, *Environment and Planning A*, Vol. 33, pp. 955-973.

Jordan, M. I., (1986). Attractor dynamics and parallelism in a connectionist sequential machine. *8th Annual Conf., Cognitive Sci. Soc.*, MIT Press, Amherst, Mass., pp. 531-546.

Khan, S. M. and Coulibaly, P., (2006), Bayesian neural network for rainfall-runoff modeling, *Water Resources Research*, Vol. 42, pp. 1-18.

Kolehmainen, M., and Doyle, M., (2003), A rigorous inter-comparison of ground-level ozone predictions, *Atmospheric Environment*, Vol. 37, pp. 3237-3253.

Koppen Climate Classification Chart, (2007), Retrieved on 06/02/2007 from <http://geography.about.com/library/weekly/aa011700b.htm>

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, Chatterton, T., Foxall, R., and Cawley, G., (2003), Extensive evaluation of neural network models for the prediction of NO₂ and PM₁₀ concentrations, compared with a deterministic modelling system and measurements in central Helsinki, *Atmospheric Environment*, Vol. 37 (32), pp. 4539 - 4550.

Kuylenstierna, J., Hicks, W. K., and Chadwick, M.J., (2002), A perspective on global air pollution problems. *Issues in environmental science and technology, No. 17, Global Environmental Change*, The Royal Society of Chemistry, 2002.

Lefohn, A.S., Knudsen, H. P., and Shadwick, D. S., (2006), Using ordinary kriging to estimate the seasonal W126 and N100 24-h concentrations for the years 2004 and 2005, Retrieved on 06/03/2007 from http://199.128.173.141/ozone/spatial/2004/contractor_2004_2005.pdf

McKean, S. A., Hsie, E.-Y., and Liu, S. C., (1991), A regional model study of the ozone budget in the eastern United States, *J. Geophys. Res.*, Vol. 96, pp. 10809-10845.

Nabney, I. T., (2004), *Netlab Algorithms for Pattern Recognition*, Springer, New York.

NeuroSolutions, (2004), *Neurosolutions: The Neural Network Simulation Environment*. NeuroSolutions getting started manual version 4, Gainesville, FL.

Niranjan, M., Lopes, S., Kukkonen, J., Karppinen, A., Greig, A., Zickus, M., Pelilan, E. and Eben, K., (2001), Exploratory studies on spatial-temporal modeling report on exploratory studies on algorithms, Air Pollution Episodes: Modelling Tools for Improved Smog Management, A project funded by the European Community under the “*Information Society Technology*” program (1998 - 2002).

NIST/SEMATECH, (2006), *E-handbook of Statistical Methods*, retrieved on 10/04/2007 from <http://www.itl.nist.gov/div898/handbook/>

Nunnari, G., Bertucco, L., Dorling, S., Schlink, U. and Ruuskanen, J., (2001), Literature review sulphur dioxide modeling at a point, Air Pollution Episodes: Modelling Tools for Improved Smog Management, A project funded by the European Community under the “*Information Society Technology*” program (1998 - 2002).

O’Hare, G. P., and Wilby, R., (1995), A review of ozone pollution in the United Kingdom and Ireland with an analysis using lamb weather types, *The geographical Journal*, Vol. 161 (1), pp. 1-20.

Ontario Ministry of Environment, (2004), Air quality in Ontario 2004 report, Retrieved on 07/2/2007 from <http://www.ene.gov.on.ca/envision/techdocs/5383e.pdf>

Ontario Ministry of Environment, (2005), Air quality in Ontario 2005 report, Retrieved on 09/02/2007 from <http://www.ene.gov.on.ca/envision/techdocs/6041e.pdf>

Ordieres, J. B., Vergara, E. P., Capuz, R. S., and Salazar, R. E., (2005), Neural network prediction model for fine particulate matter (PM_{2.5}) on the US-Mexico border in El Paso (Texas) and Ciudad Juárez (Chihuahua), *Environmental Modeling and Software*, Vol. 20, pp. 547-559.

Pagina, T., (2005), “JANN (Java Artificial Neural Network): A Neural Tool for Air Pollution Modeling”, <http://www.dees.unict.it/users/gnunnari/appetise/jann/index.html>
Developed at the University of Catania in the framework of the APPETISE project.

Perez, P. and Reyes, J., (2006), An integrated neural network model for PM₁₀ forecasting, *Atmospheric Environment*, Vol. 40, pp. 2845-2851.

Pouliou, T. (2005), Air pollution and respiratory health: re-Analysis of Hamilton children’s cohort study, M.Sc. thesis, School of Geography and Earth Sciences, McMaster University, Hamilton, Canada.

Principe, J. C., Euliano, N. R., and Lefebvre, W. C., (2000), *Neural Network and Adaptive Systems*, John Wiley & sons, Inc., USA.

Rao, S. G., (2000) a, Artificial neural networks in hydrology I: preliminary concepts, *Journal of Hydrologic Engineering*, Vol. 5 (2), pp. 115-123.

Rao, S. G., (2000) b, Artificial neural networks in hydrology II: hydrologic applications, *Journal of Hydrologic Engineering*, Vol. 5 (2), pp. 124-136.

Sahsuvaroglu, T. and Jerrett, M., (2003), A public health assessment of mortality and hospital admissions attributable to air pollution in Hamilton, City of Hamilton: Hamilton Air Quality Initiative. Retrieved on 06/02/2007 from [http://www.cleanair.hamilton.ca/downloads/Health-Study-\(Full-Report\).pdf](http://www.cleanair.hamilton.ca/downloads/Health-Study-(Full-Report).pdf)

Schlink, U., Chatterton, T., Costa, A., Dorling, S., Eben, K., Faila, J., Haase, P., Keder, J., Kolehmainen, M., Mandic, D., Nunnari, G., Nucifora, G., Pelikan, E., and Palus, M., (2001), Literature review of statistical approaches to modelling ground level ozone at a point, Air Pollution Episodes: Modelling Tools for Improved Smog Management, A

project funded by the European Community under the “*Information Society Technology*” program (1998 - 2002).

Schlink, U., Dorling, S., Pelikan, E., Nunnari, G., Cawley, G., Junninen, H., Greig, M., Foxall, R., Eben, K., Chatterton, T., Vondracek, J., Richter, M., Dostal, M., Bertuccio, L., Kolehmainen, M., and Doyle, M., (2003), A rigorous inter-comparison of ground-level ozone predictions, *Atmospheric Environment*, Vol. 37, pp. 3237-3253.

Schlink, U., Herbarth, O., Richter, M., Dorling, S., Nunnari, G., Cawley, G., and Pelikan, E., (2006), Statistical models to assess the health effects and to forecast ground-level ozone, *Environmental Modelling & Software*, Vol. 21, pp. 547-558.

Slini, T., Kaprara, A., Karatzas, K., and Moussiopoulos, N., (2006), PM₁₀ forecasting for Thessaloniki, Greece, *Environmental Modelling and Software*, Vol. 21 (4), pp. 559-565.

Source: Google map, 2007 / Imaginary Digital Globe, Earth Sat. Retrieved on 02/05/2007 from <http://maps.google.ca/?hl=en>

Sousa, S. I.V., Martins, F. G., Alvim-Ferraz, M.C.M., and Pereira, M.C., (2007), Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations, *Environmental Modelling and Software*, Vol. 22 (1), pp. 97-103.

Statistics Canada, (2005), Canada’s national statistical agency, Retrieved on 05/02/2007 from <http://www40.statcan.ca/101/cst01/demo05a.htm>

US-EPA, The Clean Air Act Amendments of 1990, Retrieved on 05/02/2007 from <http://www.epa.gov/air/caa/overview.txt>

Walker, A.M., (1969), On the asymptotic behavior of posterior distributions, *J.R. Stat. Soc.*, Vol. 31, pp. 80-88.

Yap, D, Reid, N., Brou, G.D., and Bloxam, R., (2005), Transboundary air pollution in Ontario, Report published by Ontario Ministry of Environment.

Yi, J., and Prybutok, R., (1996), A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area, *Environmental Pollution*, Vol. 92 (3), pp. 349-357.

Zickus, M., Greig, A. J., and Niranjana, M., (2002), Comparison of four machine learning methods for predicting PM₁₀ concentrations in Helsinki, Finland, *Water, Air, & Soil Pollution: Focus*, Vol. 2 (5-6), pp. 717-729.

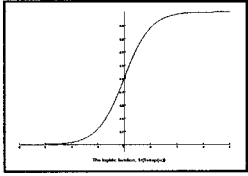
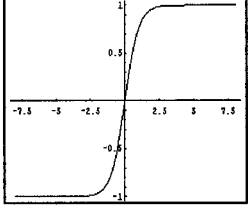
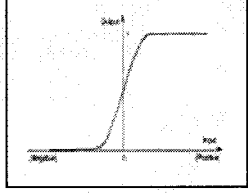
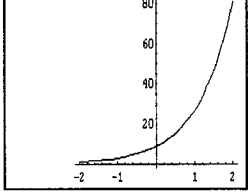
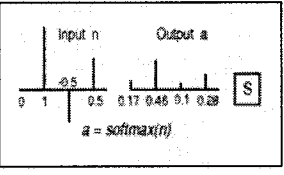
APPENDIX

Table A.1: Sources and impact of air pollutants

Pollutant Type	Major sources	Impact
<i>Total Suspended Particulates (TSP):</i> Particles larger than 10 micrometer in diameter that are suspended in the air	Use of fossil fuel: coal, oil, peat, biomass; Construction process; Dust emissions.	Human health such as: irritation in eyes, nose and throat; Nuisance dust; Soiling; Ecosystem degradation; Reduced crop quality and yield.
<i>PM₁₀:</i> Particulate matter with less than 10 micrometers in diameter which include both fine and coarse dust particles	Diesel vehicles; Aerosols; Industry and land use activity; Predominantly aerosols of sulfate, nitrate and ammonium.	Health effects; Reduced visibility.
Ground level ozone (O ₃)	Photochemical reaction between NO _x and VOCs in the presence of sunlight.	Impact of human health; Impact on crop yield, tree vitality and natural vegetation; Corrosion.
NO _x	Energy use; Any fuel, particularly from transport.	Health effects. Ecosystem acidification and eutrophication. Precursor of photochemical oxidants (O ₃).

Source: (Kuylenstierna et al, 2002)

Table A.2: Transfer functions in neural network model

Function	Definition	Range	Shape
Logistic	$\frac{1}{1 + e^{-x}}$	(0, +1)	
Hyperbolic	$\frac{e^x - e^{-x}}{e^x + e^{-x}}$	(-1, +1)	
Sigmoid	$P(t) = \frac{1}{1 + e^{-t}}$	(0, 1)	
Exponential	e^{-x}	(0, + inf)	
Softmax	$\frac{e^{x_j}}{\sum_i e^{x_i}}$	(0, +1)	

Other Activation Functions:

Linear Tangent hyperbolic Axon: ranges from -1 to 1; It is piecewise linear approximation to Tangent hyperbolic Axon.

Linear Sigmoid Axon: ranges from 0 to 1; It is piecewise linear approximation to Sigmoid Axon.

Bias Axon: It is an infinite linear axon with adjustable slope and adaptable bias.

Linear Axon: It is an infinite linear axon with adaptable bias.

Axon: It is an infinite simplest axon; identity transfer function.

Table A.3: Ground level ozone forecasting statistics: Hamilton Downtown (29000), Hamilton Mountain (29114) and Hamilton west (29118)

Predictand station	Forecasting horizon	Model	Model performance statistics				
			RMSE	R ²	r	NMSE	MAE
Hamilton Downtown	t+1	TLFN	6.28	0.85	0.92	0.15	4.50
		BNN	6.68	0.83	0.91	0.12	4.92
		RNN	6.48	0.83	0.91	0.16	4.67
		MLP	6.68	0.83	0.91	0.17	4.82
	t+2	TLFN	9.46	0.66	0.81	0.35	7.05
		BNN	10.01	0.61	0.78	0.26	7.43
		RNN	9.87	0.62	0.79	0.38	7.37
		MLP	9.89	0.62	0.79	0.38	7.33
	t+3	TLFN	11.16	0.52	0.72	0.49	8.43
		BNN	11.96	0.44	0.67	0.48	9.09
		RNN	10.90	0.53	0.73	0.46	8.27
		MLP	11.58	0.48	0.69	0.53	8.75
	t+4	TLFN	12.05	0.44	0.66	0.57	9.21
		BNN	12.94	0.35	0.59	0.59	9.95
		RNN	11.53	0.48	0.69	0.52	8.85
		MLP	12.28	0.40	0.63	0.61	9.54
	t+5	TLFN	12.42	0.40	0.63	0.60	9.54
		BNN	13.03	0.34	0.58	0.62	10.01
		RNN	11.78	0.46	0.68	0.54	9.11
		MLP	12.97	0.35	0.59	0.66	9.96
	t+6	TLFN	12.62	0.38	0.62	0.62	9.73
		BNN	13.32	0.31	0.56	0.65	10.33
		RNN	12.42	0.40	0.63	0.60	9.65
		MLP	12.51	0.31	0.56	0.69	10.27
	t+7	TLFN	12.61	0.38	0.62	0.62	9.78
		BNN	13.84	0.26	0.51	0.69	9.78
		RNN	12.68	0.37	0.61	0.63	9.91
		MLP	12.50	0.30	0.55	0.71	10.45
	t+8	TLFN	12.55	0.38	0.62	0.62	9.73
		BNN	13.75	0.27	0.52	0.69	10.69
		RNN	12.48	0.40	0.63	0.61	9.73
		MLP	12.44	0.30	0.55	0.71	10.37
	t+9	TLFN	12.63	0.38	0.62	0.62	9.82
		BNN	13.74	0.27	0.52	0.66	10.68
		RNN	12.80	0.36	0.60	0.64	9.99

	MLP	13.62	0.29	0.54	0.73	10.51	
t+10	TLFN	12.66	0.37	0.61	0.63	9.85	
	BNN	13.42	0.30	0.55	0.67	10.49	
	RNN	12.82	0.36	0.60	0.64	9.97	
	MLP	13.50	0.30	0.55	0.71	10.48	
t+11	TLFN	12.78	0.36	0.60	0.64	9.96	
	BNN	13.15	0.33	0.57	0.66	10.22	
	RNN	12.89	0.35	0.59	0.65	10.01	
	MLP	13.20	0.32	0.57	0.68	10.24	
t+12	TLFN	12.83	0.36	0.60	0.64	9.94	
	BNN	13.06	0.33	0.58	0.67	10.11	
	RNN	12.78	0.36	0.60	0.64	9.96	
	MLP	13.04	0.34	0.58	0.67	10.10	
Hamilton Mountain	t+1	TLFN	6.13	0.86	0.93	0.13	4.44
		BNN	6.57	0.85	0.92	0.14	4.73
		RNN	6.36	0.86	0.93	0.14	4.60
		MLP	6.62	0.85	0.92	0.15	4.81
	t+2	TLFN	9.73	0.67	0.82	0.33	7.28
		BNN	10.24	0.64	0.80	0.37	7.61
		RNN	10.12	0.64	0.80	0.36	7.63
		MLP	10.17	0.64	0.80	0.36	7.60
	t+3	TLFN	11.70	0.52	0.72	0.48	8.90
		BNN	12.33	0.48	0.69	0.49	9.36
		RNN	11.69	0.52	0.72	0.48	8.96
		MLP	12.17	0.48	0.70	0.52	9.25
	t+4	TLFN	12.52	0.44	0.66	0.57	9.76
		BNN	13.46	0.37	0.61	0.59	10.26
		RNN	12.58	0.45	0.67	0.56	9.72
		MLP	13.35	0.38	0.62	0.63	10.22
	t+5	TLFN	13.28	0.38	0.62	0.62	10.24
		BNN	13.90	0.33	0.58	0.65	10.74
		RNN	13.16	0.40	0.63	0.61	10.27
		MLP	13.98	0.32	0.57	0.69	10.77
	t+6	TLFN	13.39	0.37	0.61	0.63	10.40
		BNN	14.52	0.27	0.52	0.69	11.31
		RNN	13.06	0.40	0.63	0.60	10.20
		MLP	14.24	0.30	0.55	0.72	11.06
	t+7	TLFN	13.31	0.38	0.62	0.63	10.39
		BNN	14.30	0.29	0.54	0.68	11.08
		RNN	13.22	0.38	0.62	0.62	10.33

	MLP	14.25	0.30	0.55	0.72	11.06	
t+8	TLFN	13.14	0.40	0.63	0.61	10.27	
	BNN	14.62	0.26	0.51	0.68	11.31	
	RNN	13.20	0.38	0.62	0.61	10.29	
	MLP	14.30	0.29	0.54	0.72	11.15	
t+9	TLFN	13.10	0.40	0.63	0.60	10.20	
	BNN	14.30	0.29	0.54	0.67	11.16	
	RNN	13.31	0.38	0.62	0.63	10.36	
	MLP	14.20	0.30	0.55	0.71	11.11	
t+10	TLFN	13.18	0.40	0.63	0.61	10.27	
	BNN	14.30	0.30	0.55	0.67	6.78	
	RNN	13.29	0.38	0.62	0.62	10.37	
	MLP	13.98	0.32	0.57	0.69	10.91	
t+11	TLFN	13.26	0.38	0.62	0.62	10.33	
	BNN	13.71	0.35	0.59	0.67	6.61	
	RNN	13.41	0.37	0.61	0.63	10.42	
	MLP	13.59	0.34	0.60	0.65	10.61	
t+12	TLFN	13.42	0.37	0.61	0.64	10.44	
	BNN	13.71	0.59	0.35	0.65	6.58	
	RNN	13.54	0.36	0.60	0.65	10.53	
	MLP	13.58	0.36	0.60	0.65	10.56	
Hamilton West	t+1	TLFN	6.22	0.83	0.91	0.17	4.47
		BNN	6.89	0.80	0.89	0.17	4.96
		RNN	6.51	0.82	0.91	0.18	4.67
		MLP	6.83	0.80	0.89	0.20	4.92
	t+2	TLFN	9.21	0.64	0.80	0.36	6.92
		BNN	10.55	0.53	0.73	0.40	7.87
		RNN	6.64	0.61	0.78	0.39	7.30
		MLP	10.40	0.54	0.74	0.46	7.82
	t+3	TLFN	10.79	0.50	0.71	0.49	8.28
		BNN	12.71	0.32	0.57	0.52	9.70
		RNN	10.79	0.51	0.71	0.49	8.32
		MLP	12.33	0.36	0.60	0.64	9.41
t+4	TLFN	11.43	0.45	0.67	0.55	8.87	
	BNN	13.76	0.20	0.46	0.62	10.55	
	RNN	11.51	0.44	0.66	0.56	8.95	
	MLP	13.44	0.24	0.49	0.76	10.42	
t+5	TLFN	12.19	0.37	0.61	0.63	9.51	
	BNN	14.11	0.16	0.40	0.67	10.98	
	RNN	12.02	0.39	0.63	0.61	9.38	

	MLP	13.96	0.19	0.43	0.83	10.83
t+6	TLFN	12.39	0.35	0.59	0.65	9.71
	BNN	14.17	0.16	0.40	0.69	10.99
	RNN	12.07	0.38	0.62	0.62	9.47
	MLP	13.57	0.23	0.48	0.78	10.54
t+7	TLFN	12.57	0.33	0.57	0.67	9.90
	BNN	13.59	0.22	0.47	0.71	10.58
	RNN	12.70	0.31	0.56	0.68	9.97
	MLP	13.21	0.27	0.52	0.74	10.26
t+8	TLFN	12.63	0.32	0.57	0.68	9.98
	BNN	13.26	0.26	0.51	0.68	10.36
	RNN	12.73	0.31	0.56	0.69	10.02
	MLP	12.94	0.29	0.54	0.71	10.11
t+9	TLFN	12.76	0.31	0.56	0.69	10.07
	BNN	13.01	0.28	0.53	0.69	10.22
	RNN	12.81	0.30	0.55	0.70	10.11
	MLP	12.92	0.29	0.54	0.71	10.14
t+10	TLFN	12.72	0.31	0.56	0.69	10.06
	BNN	13.13	0.27	0.52	0.68	10.35
	RNN	12.89	0.30	0.55	0.70	10.20
	MLP	13.20	0.27	0.52	0.74	10.42
t+11	TLFN	12.74	0.31	0.56	0.69	10.08
	BNN	13.06	0.28	0.53	0.71	8.72
	RNN	12.89	0.30	0.55	0.70	10.18
	MLP	13.11	0.27	0.52	0.73	10.36
t+12	TLFN	12.76	0.31	0.56	0.69	10.11
	BNN	13.00	0.29	0.53	0.70	10.24
	RNN	12.96	0.29	0.54	0.71	10.26
	MLP	13.01	0.28	0.53	0.72	10.25

Legends: TLFN: Time lagged feedforward network
MLP: Multi-layer Perceptron

RNN: Recurrent neural network
TDNN: Time delay neural network

Table A. 4 (a): Seasonal variation of model performance of O₃: Hamilton Mountain (29114)

Forecasting period	Season	Model	Model performance statistics					
			RMSE	r	R ²	NMSE	MAE	SB
t+1	Fall	TLFN	4.48	0.90	0.81	0.20	3.35	-0.01
		BNN	4.67	0.89	0.79	0.21	3.44	0.00
		RNN	4.70	0.89	0.79	0.22	3.51	0.01
		MLP	4.78	0.88	0.78	0.22	3.55	-0.02
	Summer	TLFN	7.49	0.93	0.87	0.14	5.60	-0.03
		BNN	8.21	0.92	0.84	0.16	6.12	-0.05
		RNN	7.75	0.93	0.86	0.15	5.76	-0.04
		MLP	8.15	0.92	0.84	0.16	6.09	-0.04
	Spring	TLFN	6.75	0.89	0.79	0.21	4.95	-0.02
		BNN	7.03	0.88	0.77	0.23	5.20	-0.03
		RNN	6.91	0.88	0.78	0.22	5.07	-0.03
		MLP	7.14	0.88	0.77	0.24	5.32	-0.04
	Winter	TLFN	4.48	0.90	0.81	0.20	3.35	-0.01
		BNN	4.67	0.89	0.79	0.21	3.44	0.00
		RNN	4.70	0.89	0.79	0.22	3.51	0.01
		MLP	4.78	0.88	0.78	0.22	3.55	-0.02
t+2	Fall	TLFN	8.74	0.81	0.66	0.34	6.55	0.00
		BNN	9.18	0.79	0.62	0.38	6.86	-0.01
		RNN	9.13	0.79	0.63	0.37	6.89	-0.01
		MLP	9.27	0.79	0.62	0.38	6.93	-0.01
	Summer	TLFN	12.08	0.81	0.65	0.36	9.26	-0.06
		BNN	13.04	0.78	0.61	0.42	9.89	-0.09
		RNN	12.63	0.79	0.63	0.39	9.69	-0.07
		MLP	12.74	0.79	0.62	0.40	9.73	-0.07
	Spring	TLFN	10.41	0.72	0.52	0.51	8.03	-0.07
		BNN	10.66	0.70	0.50	0.53	8.23	-0.06
		RNN	10.70	0.71	0.50	0.54	8.36	-0.09
		MLP	10.74	0.70	0.49	0.54	8.34	-0.07
	Winter	TLFN	7.10	0.72	0.52	0.49	5.42	0.00
		BNN	7.24	0.71	0.50	0.51	5.57	0.01
		RNN	7.36	0.69	0.48	0.53	5.71	0.00
		MLP	7.22	0.71	0.51	0.51	5.55	0.00
t+6	Fall	TLFN	11.88	0.61	0.37	0.63	9.35	0.023
		BNN	12.73	0.53	0.28	0.72	10.05	0.022
		RNN	11.65	0.63	0.40	0.60	9.25	0.031
		MLP	12.64	0.54	0.29	0.71	9.89	0.003

Summer	TLFN	17.05	0.57	0.32	0.71	13.43	-0.099
	BNN	18.71	0.47	0.22	0.86	14.58	-0.157
	RNN	16.68	0.58	0.33	0.68	13.18	-0.062
	MLP	18.12	0.50	0.25	0.80	14.30	-0.126
Spring	TLFN	13.70	0.43	0.19	0.88	11.08	-0.096
	BNN	14.82	0.33	0.11	1.03	12.18	-0.138
	RNN	13.35	0.43	0.19	0.84	10.79	-0.059
	MLP	14.78	0.36	0.13	1.03	11.98	-0.144
Winter	TLFN	9.88	0.35	0.12	0.95	7.86	-0.008
	BNN	10.55	0.19	0.04	1.08	8.59	0.013
	RNN	9.48	0.39	0.16	0.88	7.65	-0.009
	MLP	10.33	0.29	0.09	1.04	8.23	-0.038

Table A. 4 (b): Seasonal variation of model performance of O₃: Hamilton West (29118)

Forecasting period	Season	Model	Model performance statistics					
			RMSE	r	R ²	NMSE	MAE	SB
t+1	Fall	TLFN	5.47	0.92	0.84	0.16	3.93	0.014
		BNN	6.09	0.90	0.81	0.19	4.36	0.018
		RNN	5.83	0.91	0.82	0.18	4.19	0.017
		MLP	6.12	0.90	0.80	0.20	4.41	0.019
	Summer	TLFN	7.59	0.92	0.84	0.16	5.58	-0.010
		BNN	8.56	0.90	0.80	0.20	6.28	-0.040
		RNN	7.93	0.91	0.83	0.17	5.82	-0.027
		MLP	8.38	0.90	0.81	0.19	6.15	-0.022
	Spring	TLFN	7.08	0.86	0.74	0.26	5.13	-0.011
		BNN	7.57	0.84	0.71	0.30	5.60	-0.040
		RNN	7.22	0.86	0.73	0.27	5.23	-0.030
		MLP	7.55	0.84	0.71	0.30	5.52	-0.031
	Winter	TLFN	4.59	0.89	0.79	0.21	3.44	-0.005
		BNN	4.91	0.87	0.76	0.24	3.72	0.012
		RNN	4.76	0.88	0.78	0.22	3.57	0.008
		MLP	4.93	0.87	0.76	0.24	3.74	0.006
t+2	Fall	TLFN	8.05	0.81	0.66	0.33	6.07	0.033
		BNN	9.44	0.73	0.54	0.46	7.05	0.039
		RNN	8.73	0.78	0.60	0.39	6.58	0.013
		MLP	9.32	0.74	0.55	0.45	6.99	0.050
	Summer	TLFN	11.45	0.80	0.65	0.35	8.71	-0.018
		BNN	13.45	0.72	0.52	0.49	10.15	-0.069
		RNN	11.89	0.79	0.62	0.38	9.14	-0.024
		MLP	13.01	0.74	0.55	0.45	9.92	-0.054
	Spring	TLFN	9.92	0.70	0.49	0.51	7.68	-0.027
		BNN	10.99	0.63	0.39	0.63	8.56	-0.067
		RNN	10.22	0.68	0.46	0.54	7.93	-0.040
		MLP	11.03	0.63	0.40	0.63	8.63	-0.075
	Winter	TLFN	6.85	0.74	0.54	0.46	5.37	-0.005
		BNN	7.40	0.68	0.47	0.53	5.87	0.0241
		RNN	7.15	0.71	0.50	0.50	5.65	-0.024
		MLP	7.54	0.67	0.45	0.55	5.90	0.017
t+6	Fall	TLFN	10.86	0.62	0.39	0.62	8.60	0.055
		BNN	12.62	0.42	0.18	0.83	10.00	0.089
		RNN	10.81	0.63	0.40	0.61	8.59	0.066

	MLP	12.25	0.48	0.24	0.78	9.67	0.091
Summer	TLFN	15.73	0.58	0.34	0.67	12.34	-0.056
	BNN	18.65	0.35	0.12	0.94	14.33	-0.185
	RNN	15.05	0.63	0.39	0.61	11.74	-0.052
	MLP	17.29	0.48	0.23	0.81	13.29	-0.144
Spring	TLFN	12.67	0.43	0.19	0.84	10.34	-0.059
	BNN	13.82	0.28	0.08	1.00	11.30	-0.133
	RNN	12.31	0.49	0.24	0.79	9.90	-0.050
	MLP	13.51	0.35	0.12	0.95	10.88	-0.111
Winter	TLFN	9.39	0.41	0.17	0.86	7.67	-0.015
	BNN	10.01	0.25	0.06	0.98	8.35	0.081
	RNN	9.38	0.43	0.19	0.86	7.74	-0.054
	MLP	10.14	0.29	0.08	1.01	8.37	0.064

Table A. 5: Comparison of annual and summer models for O₃: Hamilton Downtown (29000)

Forecasting period	Model	Model performance statistics									
		Annual model					Summer model				
		RMSE	r	R ²	NMSE	MAE	RMSE	r	R ²	NMSE	MAE
t+1	TLFN	7.86	0.91	0.84	0.17	5.77	7.88	0.91	0.83	0.17	5.89
	BNN	8.63	0.90	0.81	0.20	6.33	8.43	0.90	0.81	0.19	6.31
	RNN	8.06	0.91	0.83	0.17	5.91	8.08	0.91	0.83	0.17	6.10
	MLP	8.40	0.90	0.82	0.19	6.21	8.37	0.90	0.81	0.19	6.30
t+2	TLFN	11.96	0.79	0.62	0.38	9.12	12.38	0.77	0.59	0.41	9.52
	BNN	12.77	0.76	0.58	0.44	9.60	13.09	0.74	0.54	0.46	10.09
	RNN	12.49	0.77	0.59	0.42	9.40	12.73	0.75	0.56	0.43	9.86
	MLP	12.51	0.77	0.59	0.42	9.36	12.72	0.75	0.56	0.43	9.75
t+3	TLFN	14.24	0.69	0.47	0.54	10.94	15.11	0.62	0.38	0.61	11.82
	BNN	15.40	0.63	0.40	0.63	11.73	15.74	0.58	0.34	0.66	12.25
	RNN	13.99	0.69	0.48	0.52	10.82	13.72	0.71	0.50	0.50	10.57
	MLP	14.78	0.66	0.44	0.58	11.29	15.49	0.60	0.36	0.64	12.01
t+4	TLFN	15.55	0.61	0.37	0.65	12.07	16.07	0.56	0.31	0.69	12.60
	BNN	16.86	0.54	0.30	0.76	12.92	17.20	0.47	0.22	0.79	13.34
	RNN	14.92	0.65	0.42	0.59	11.61	14.71	0.65	0.42	0.58	11.57
	MLP	16.04	0.59	0.35	0.69	12.31	16.69	0.51	0.26	0.75	12.99
t+5	TLFN	15.95	0.58	0.34	0.68	12.37	16.80	0.50	0.25	0.76	13.23
	BNN	16.98	0.52	0.27	0.77	13.15	18.05	0.38	0.14	0.87	14.29
	RNN	15.04	0.64	0.41	0.60	11.72	15.28	0.61	0.37	0.62	11.97
	MLP	16.65	0.55	0.30	0.74	12.85	17.32	0.46	0.21	0.80	13.57
t+6	TLFN	16.25	0.57	0.32	0.71	12.61	17.02	0.48	0.23	0.78	13.42
	BNN	17.33	0.50	0.25	0.80	13.45	18.33	0.35	0.12	0.90	14.30
	RNN	16.14	0.56	0.31	0.70	12.78	15.80	0.58	0.34	0.67	12.36
	MLP	17.23	0.51	0.26	0.79	13.37	17.66	0.42	0.18	0.83	13.93

Table A. 6: Model forecasting performance with land use variable for O₃: Hamilton Downtown (29000)

Forecasting period	Model	Model performance statistics					
		RMSE	r	R ²	NMSE	MAE	RB
t+1	TLFN	6.34	0.92	0.85	0.16	4.55	-0.011
	BNN	6.65	0.91	0.83	0.17	4.77	-0.011
	RNN	6.89	0.9	0.81	0.19	5.06	-0.011
	MLP	6.69	0.91	0.83	0.18	4.8	-0.011
t+2	TLFN	9.62	0.80	0.64	0.36	7.15	-0.029
	BNN	10.32	0.76	0.58	0.42	7.65	-0.036
	RNN	10.20	0.77	0.59	0.41	7.64	-0.023
	MLP	10.00	0.78	0.61	0.39	7.43	-0.038
t+3	TLFN	11.09	0.72	0.52	0.48	8.41	-0.022
	BNN	11.68	0.69	0.47	0.53	8.82	-0.053
	RNN	11.84	0.67	0.45	0.55	9.05	-0.044
	MLP	11.66	0.69	0.48	0.53	8.76	-0.053
t+4	TLFN	11.75	0.68	0.46	0.54	9.03	-0.006
	BNN	12.65	0.62	0.38	0.63	9.60	-0.059
	RNN	12.37	0.63	0.40	0.60	9.66	-0.021
	MLP	12.44	0.63	0.40	0.61	9.47	-0.040
t+5	TLFN	12.10	0.65	0.42	0.57	9.36	-0.010
	BNN	13.12	0.58	0.33	0.67	10.05	-0.061
	RNN	12.72	0.61	0.37	0.63	9.91	-0.021
	MLP	12.98	0.59	0.35	0.66	9.99	-0.054
t+6	TLFN	12.26	0.64	0.41	0.59	9.51	-0.012
	BNN	13.69	0.52	0.27	0.73	10.52	-0.067
	RNN	12.65	0.61	0.37	0.63	9.88	-0.018
	MLP	13.25	0.56	0.31	0.69	10.22	-0.050

Table A. 7 (a): Model forecasting performance of NO₂: Hamilton Downtown (29000)

Forecasting horizon	Model	Model forecasting statistics				
		RMSE	r	R ²	NMSE	MAE
t+1	TLFN	5.76	0.84	0.71	0.29	4.16
	BNN	5.79	0.84	0.70	0.29	4.22
	RNN	5.77	0.84	0.71	0.29	4.19
	MLP	5.85	0.84	0.71	0.30	4.25
t+2	TLFN	8.07	0.65	0.42	0.57	6.12
	BNN	8.02	0.66	0.43	0.57	6.07
	RNN	8.22	0.64	0.41	0.59	6.29
	MLP	8.09	0.65	0.42	0.58	6.12
t+3	TLFN	9.07	0.53	0.28	0.72	6.98
	BNN	8.91	0.55	0.30	0.70	6.87
	RNN	8.95	0.54	0.29	0.70	6.89
	MLP	9.00	0.54	0.29	0.71	6.93
t+4	TLFN	9.27	0.50	0.25	0.75	7.18
	BNN	9.41	0.47	0.22	0.78	7.27
	RNN	9.30	0.49	0.24	0.76	7.25
	MLP	9.35	0.48	0.23	0.77	7.27
t+5	TLFN	9.49	0.45	0.20	0.79	7.43
	BNN	9.48	0.46	0.21	0.79	7.39
	RNN	9.48	0.46	0.21	0.79	7.40
	MLP	9.56	0.44	0.19	0.80	7.45
t+6	TLFN	9.78	0.40	0.16	0.84	7.65
	BNN	9.72	0.41	0.17	0.83	7.62
	RNN	9.82	0.40	0.16	0.84	7.65
	MLP	9.74	0.41	0.17	0.83	7.61
t+7	TLFN	10.00	0.35	0.12	0.87	7.86
	BNN	9.98	0.35	0.12	0.88	7.83
	RNN	9.87	0.38	0.14	0.86	7.74
	MLP	9.96	0.36	0.13	0.87	7.81
t+8	TLFN	10.12	0.32	0.10	0.90	7.96
	BNN	10.18	0.30	0.09	0.91	8.00
	RNN	9.92	0.37	0.14	0.86	7.79
	MLP	10.18	0.31	0.10	0.91	7.98
t+9	TLFN	10.17	0.30	0.09	0.91	8.03
	BNN	10.22	0.29	0.08	0.92	8.07
	RNN	10.00	0.35	0.12	0.88	7.89

	MLP	10.23	0.29	0.08	0.92	8.06
t+10	TLFN	10.08	0.33	0.11	0.89	7.93
	BNN	10.26	0.28	0.08	0.90	8.09
	RNN	9.97	0.35	0.12	0.87	7.86
	MLP	10.24	0.28	0.08	0.92	8.06
t+11	TLFN	9.97	0.35	0.12	0.87	7.82
	BNN	10.06	0.33	0.11	0.89	7.90
	RNN	9.95	0.36	0.13	0.87	7.84
	MLP	10.07	0.33	0.11	0.89	7.92
t+12	TLFN	9.97	0.35	0.12	0.87	7.84
	BNN	10.07	0.33	0.11	0.87	7.90
	RNN	10.00	0.34	0.12	0.88	7.83
	MLP	10.00	0.35	0.12	0.88	7.82

Table A. 7 (b): Model forecasting performance for NO₂: Hamilton West (29118)

Forecasting period	Model forecasting statistics					
	Model	RMSE	r	R ²	NMSE	MAE
t+1	TLFN	5.77	0.87	0.77	0.24	4.22
	BNN	5.99	0.86	0.75	0.25	4.38
	RNN	5.83	0.87	0.76	0.24	4.26
	MLP	6.04	0.86	0.74	0.26	4.45
t+2	TLFN	8.04	0.74	0.55	0.47	6.16
	BNN	8.39	0.71	0.50	0.51	6.44
	RNN	8.23	0.72	0.52	6.44	6.33
	MLP	8.39	0.71	0.50	0.51	6.45
t+3	TLFN	9.23	0.63	0.40	0.61	7.20
	BNN	9.51	0.60	0.36	0.65	7.39
	RNN	9.28	0.63	0.40	0.62	7.26
	MLP	9.47	0.61	0.37	0.65	7.38
t+4	TLFN	9.83	0.56	0.31	0.70	7.75
	BNN	10.15	0.52	0.27	0.72	7.86
	RNN	9.86	0.56	0.31	0.70	7.81
	MLP	10.00	0.54	0.29	0.73	7.91
t+5	TLFN	10.20	0.52	0.27	0.75	8.11
	BNN	10.34	0.49	0.24	0.77	8.18
	RNN	10.18	0.52	0.27	0.75	8.15

	MLP	10.37	0.49	0.24	0.77	8.24
t+6	TLFN	10.45	0.48	0.23	0.79	8.32
	BNN	10.63	0.45	0.20	0.81	8.47
	RNN	10.54	0.47	0.22	0.80	8.48
	MLP	10.65	0.45	0.20	0.82	8.51
t+7	TLFN	10.74	0.44	0.19	0.83	8.63
	BNN	10.89	0.40	0.16	0.85	8.70
	RNN	10.60	0.45	0.20	0.81	8.52
	MLP	10.93	0.40	0.16	0.86	8.76
t+8	TLFN	10.73	0.44	0.19	0.83	8.64
	BNN	11.16	0.36	0.13	0.90	9.00
	RNN	10.69	0.44	0.19	0.82	8.62
	MLP	11.12	0.36	0.13	0.89	8.95
t+9	TLFN	10.73	0.44	0.19	0.83	8.64
	BNN	11.26	0.33	0.11	0.91	9.10
	RNN	10.90	0.41	0.17	0.85	8.83
	MLP	11.25	0.33	0.11	0.91	9.09
t+10	TLFN	10.81	0.42	0.18	0.84	8.73
	BNN	11.19	0.35	0.12	0.90	9.06
	RNN	10.82	0.42	0.18	0.84	8.72
	MLP	11.22	0.34	0.12	0.90	9.06
t+11	TLFN	10.84	0.41	0.17	0.84	8.73
	BNN	11.00	0.38	0.14	0.87	8.89
	RNN	10.87	0.40	0.16	0.85	8.77
	MLP	11.03	0.38	0.14	0.88	8.92
t+12	TLFN	10.96	0.39	0.15	0.86	8.83
	BNN	10.99	0.39	0.15	0.87	8.87
	RNN	10.07	0.37	0.14	0.88	8.94
	MLP	10.99	0.39	0.15	0.87	8.87

Table A. 8: Seasonal variation of model performance: Hamilton Downtown (29000)

Forecasting period	Season	Model	Model performance Statistics					
			RMSE	r	R ²	NMSE	MAE	SB
t+1	Fall	TLFN	5.14	0.84	0.71	0.29	3.71	0.009
		BNN	5.18	0.84	0.70	0.30	3.77	0.009
		RNN	5.14	0.84	0.71	0.29	3.74	0.008
		MLP	5.24	0.83	0.70	0.30	3.81	0.009
	Summer	TLFN	6.10	0.82	0.67	0.34	4.44	0.034
		BNN	6.18	0.81	0.66	0.35	4.53	0.035
		RNN	6.13	0.81	0.66	0.34	4.51	0.030
		MLP	6.15	0.81	0.66	0.34	4.51	0.033
	Spring	TLFN	7.15	0.81	0.66	0.34	5.21	-0.029
		BNN	7.16	0.81	0.66	0.34	5.22	-0.027
		RNN	7.17	0.81	0.66	0.34	5.24	-0.025
		MLP	7.23	0.81	0.66	0.35	5.29	-0.025
	Winter	TLFN	4.90	0.87	0.75	0.26	3.63	-0.028
		BNN	4.93	0.86	0.75	0.26	3.67	-0.025
		RNN	4.85	0.87	0.75	0.25	3.61	-0.022
		MLP	4.98	0.86	0.74	0.26	3.71	-0.023
t+2	Fall	TLFN	7.15	0.65	0.43	0.58	5.45	0.026
		BNN	7.10	0.66	0.43	0.57	5.41	0.029
		RNN	7.36	0.63	0.39	0.61	5.62	0.027
		MLP	7.16	0.65	0.42	0.58	5.44	0.016
	Summer	TLFN	8.26	0.63	0.40	0.62	6.39	0.065
		BNN	8.24	0.63	0.40	0.62	6.33	0.058
		RNN	8.37	0.61	0.37	0.64	6.50	0.041
		MLP	8.28	0.62	0.39	0.62	6.40	0.059
	Spring	TLFN	8.74	0.63	0.40	0.60	6.55	-0.016
		BNN	8.68	0.64	0.41	0.59	6.50	-0.021
		RNN	8.90	0.62	0.38	0.62	6.76	-0.023
		MLP	8.77	0.63	0.40	0.61	6.55	-0.025
	Winter	TLFN	7.17	0.68	0.46	0.55	5.51	-0.043
		BNN	7.10	0.69	0.48	0.54	5.46	-0.047
		RNN	7.35	0.66	0.43	0.58	5.67	-0.033
		MLP	7.16	0.69	0.47	0.55	5.49	-0.055
t+6	Fall	TLFN	8.90	0.36	0.13	0.88	7.06	0.032
		BNN	8.81	0.38	0.15	0.86	7.02	0.036
		RNN	8.80	0.38	0.15	0.86	7.02	0.023
		MLP	8.88	0.37	0.14	0.87	7.02	0.025

Summer	TLFN	9.98	0.35	0.12	0.91	8.09	0.075
	BNN	9.84	0.37	0.14	0.88	7.98	0.067
	RNN	9.89	0.36	0.13	0.89	8.05	0.073
	MLP	9.91	0.37	0.14	0.89	8.03	0.084
Spring	TLFN	11.78	0.34	0.12	0.92	9.02	-0.085
	BNN	11.79	0.35	0.12	0.92	9.05	-0.099
	RNN	11.82	0.35	0.12	0.93	9.03	-0.106
	MLP	11.67	0.36	0.13	0.90	8.92	-0.088
Winter	TLFN	8.86	0.43	0.18	0.84	6.88	-0.055
	BNN	8.88	0.43	0.18	0.84	6.87	-0.057
	RNN	8.99	0.41	0.17	0.86	6.94	-0.071
	MLP	8.92	0.42	0.18	0.85	6.90	-0.067

Fig. A. 1 (a): Scatter plots of observed vs. simulated values of TSP values at station 29113 (left) and 29114 (right)

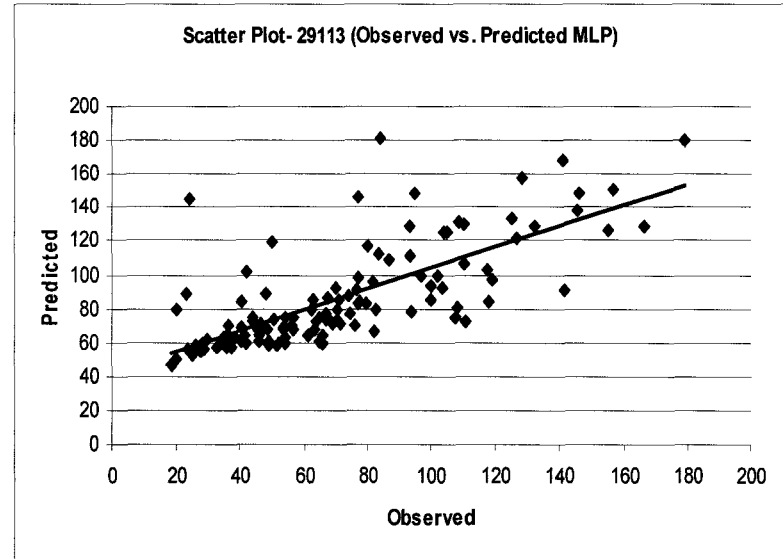
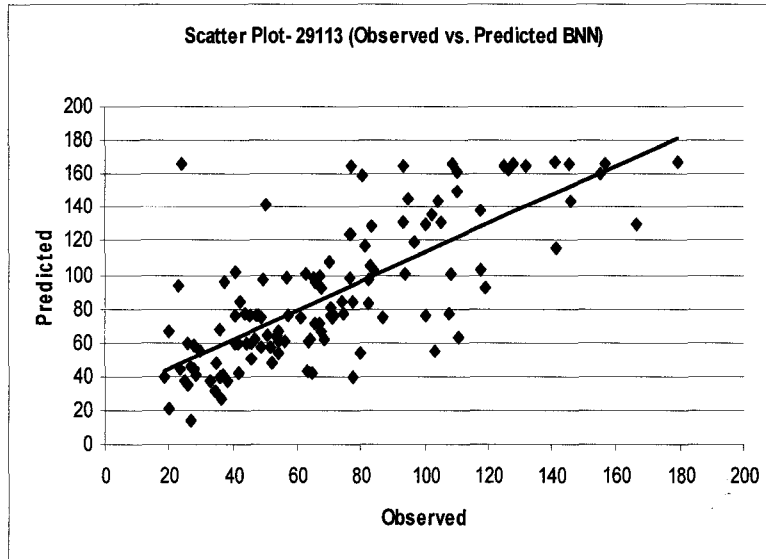
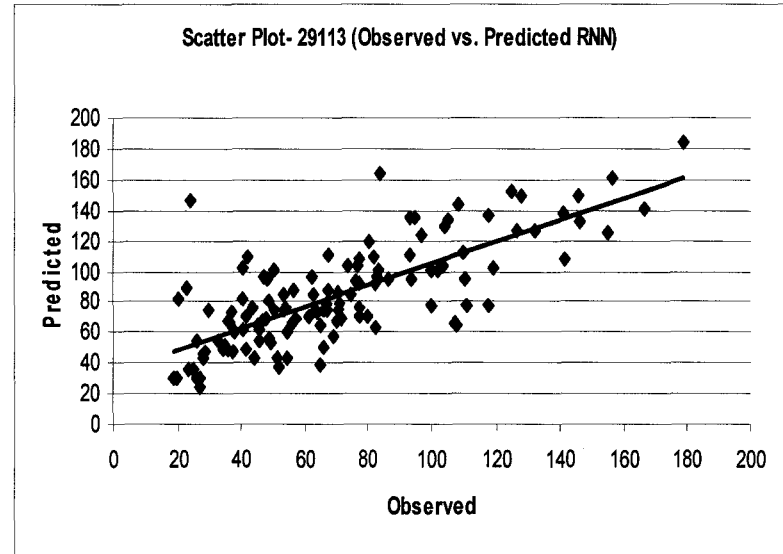
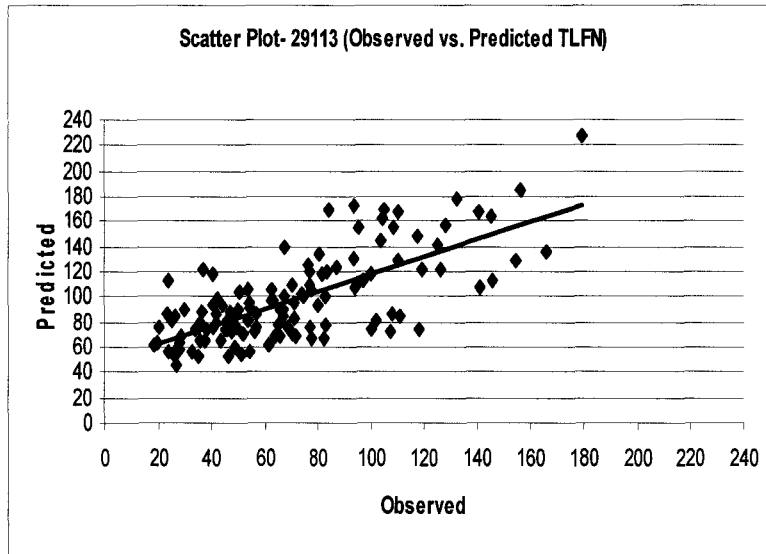


Fig. A. 1 (b): Scatter plots of observed vs. simulated values of TSP values at station 29113 (left) and 29114 (right)

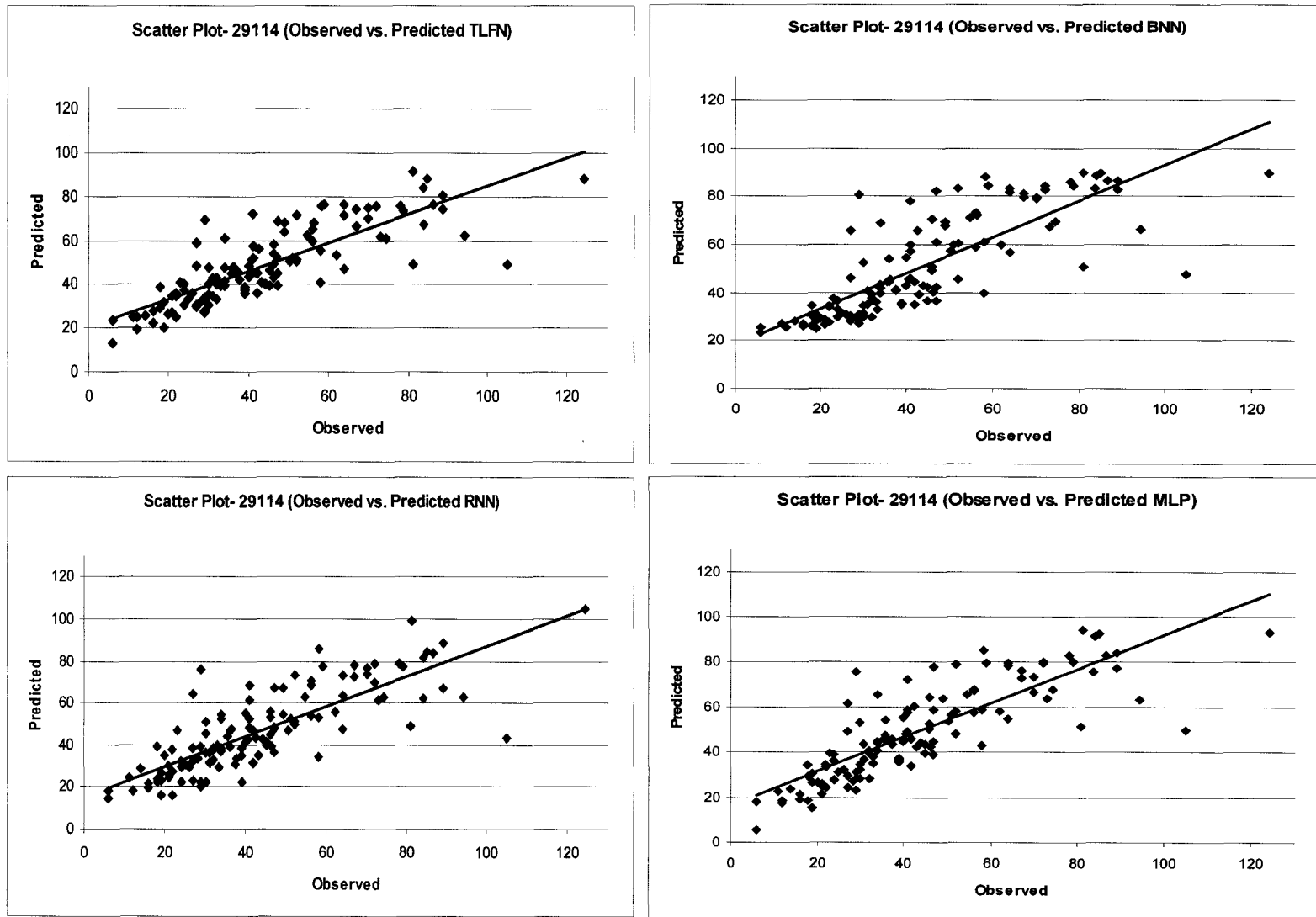


Fig. A. 1 (c): Scatter plots of observed vs. simulated values of TSP concentration at station 29102

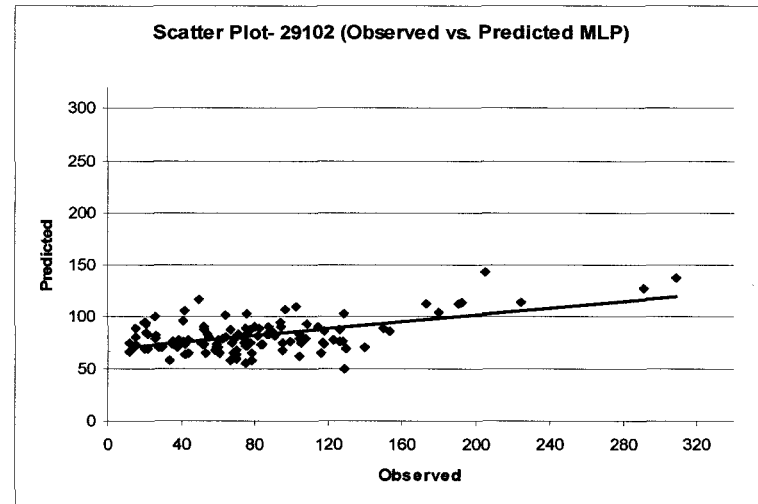
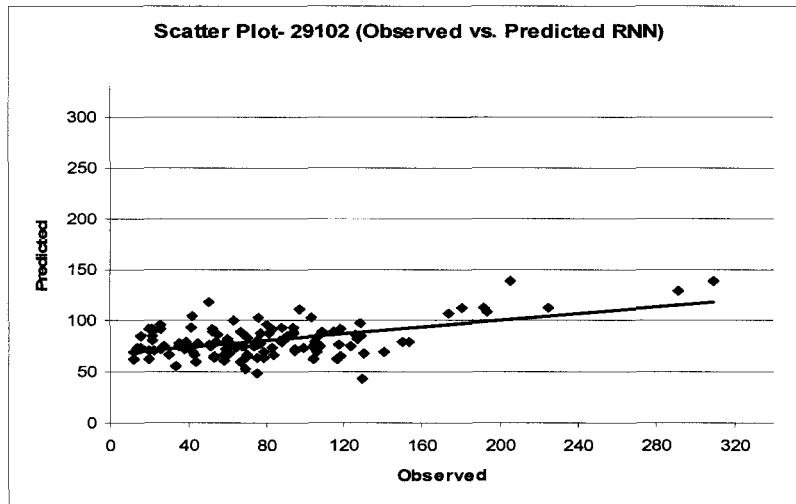
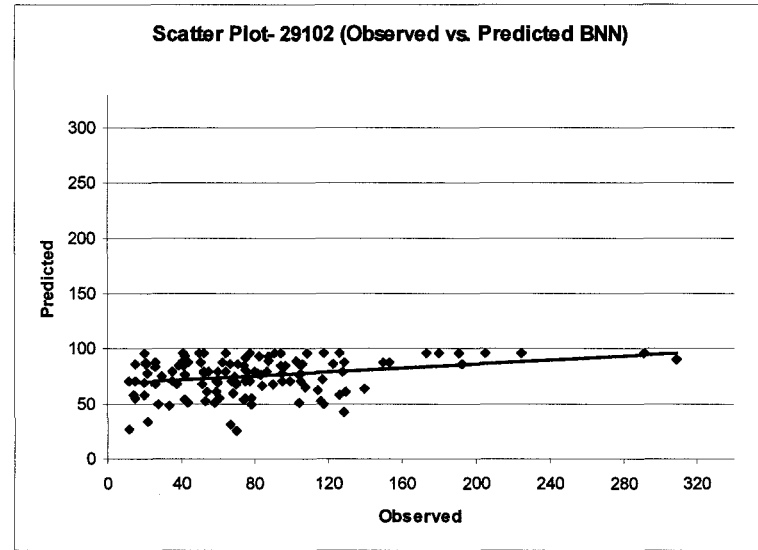
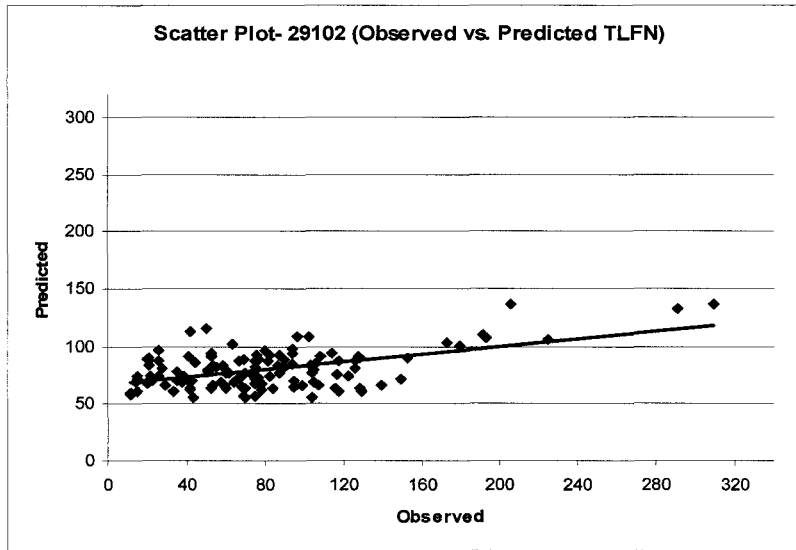


Fig A. 2 (a): Simulation result of TSP at station 29000 with 95% confidence interval

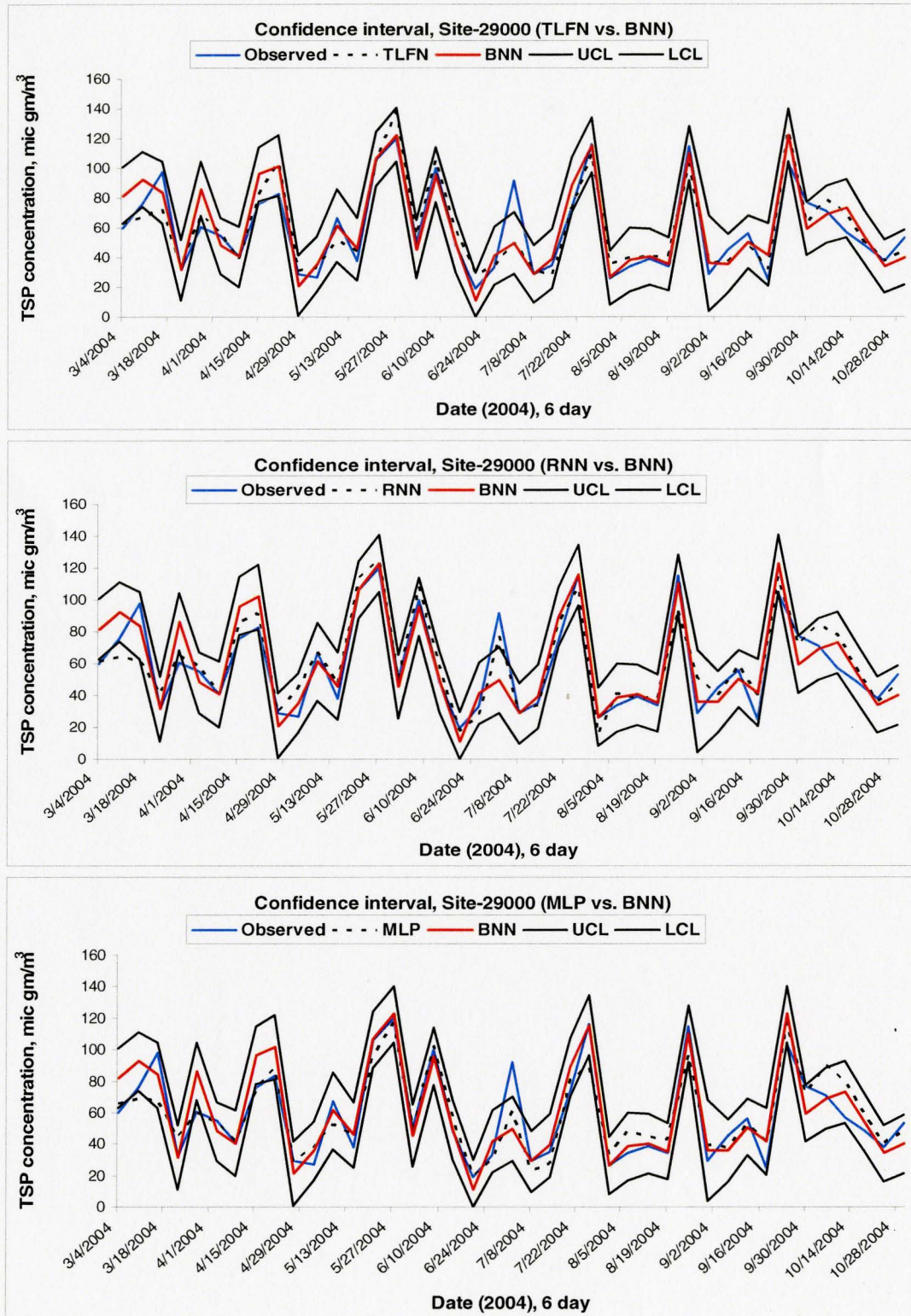


Fig. A. 2 (b): Simulation result of TSP at station 29113 with 95% confidence interval

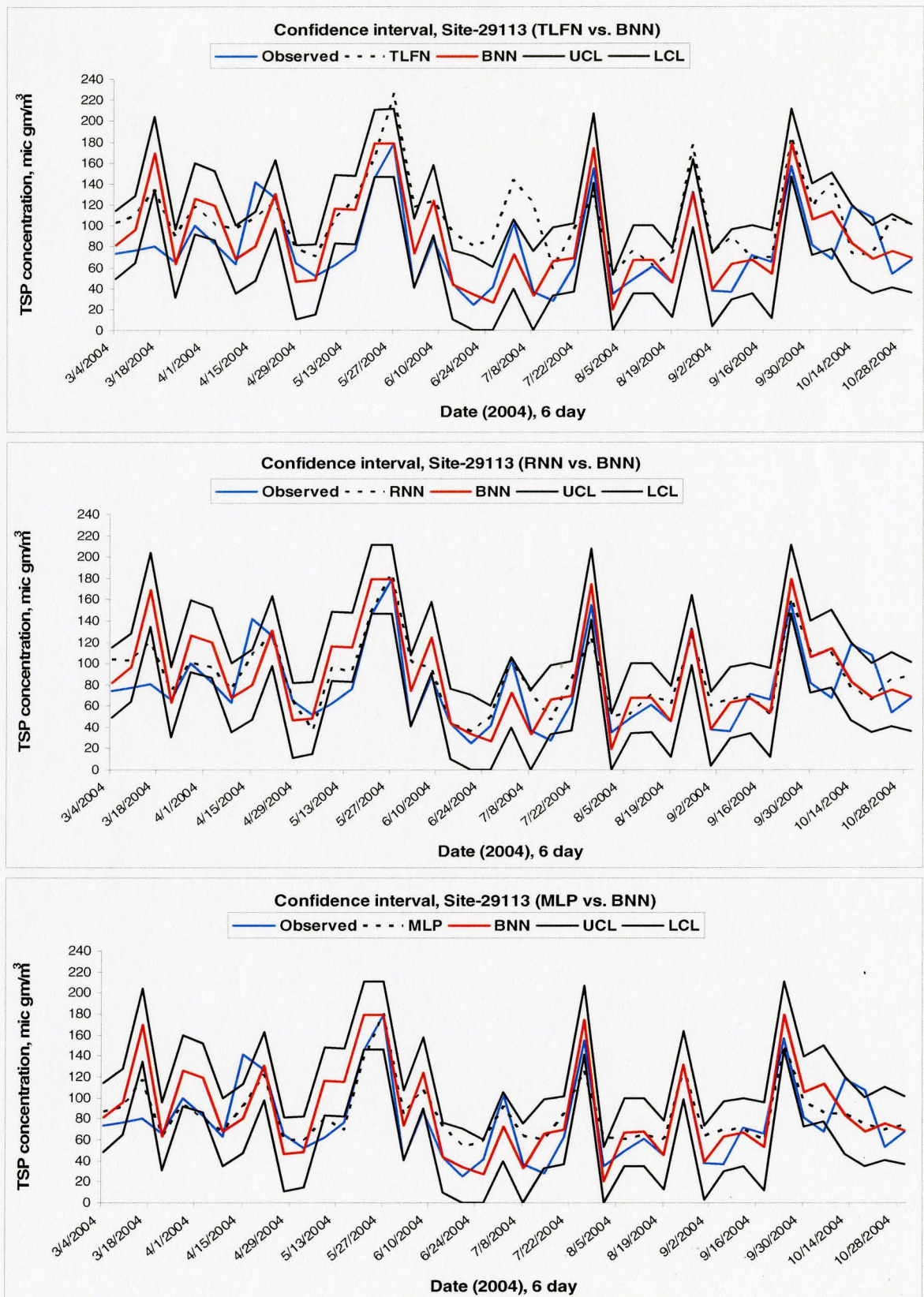


Fig. A. 2 (c): Simulation result of TSP at station 29114 with 95% confidence interval

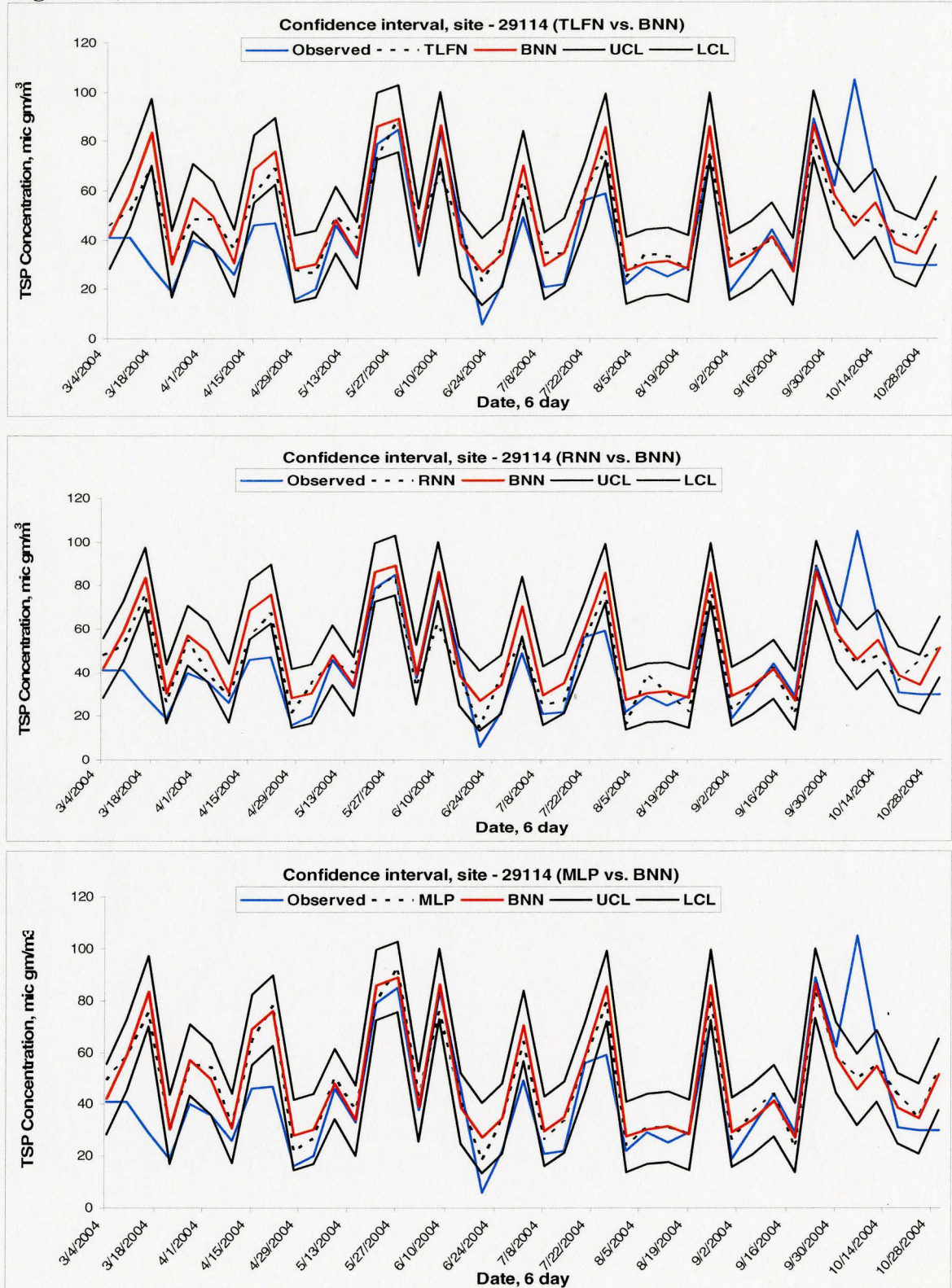
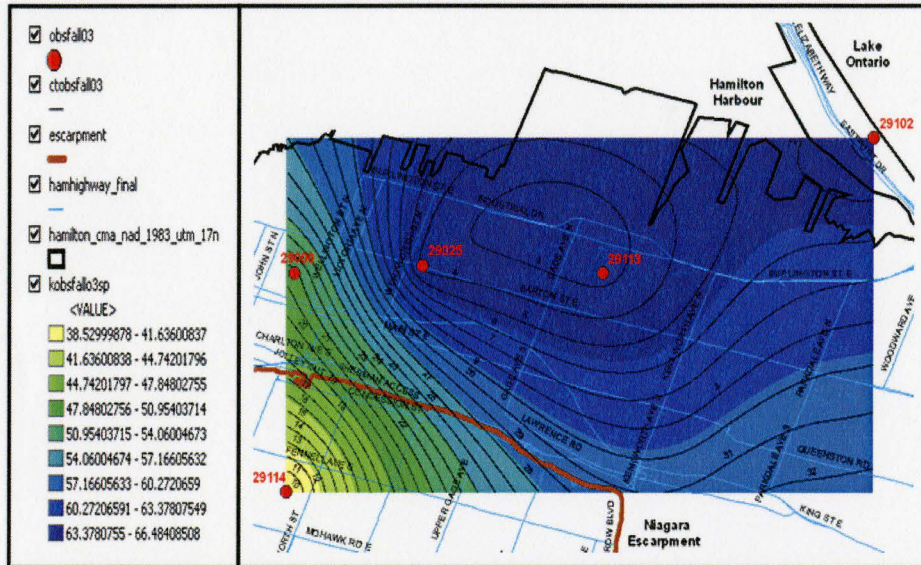
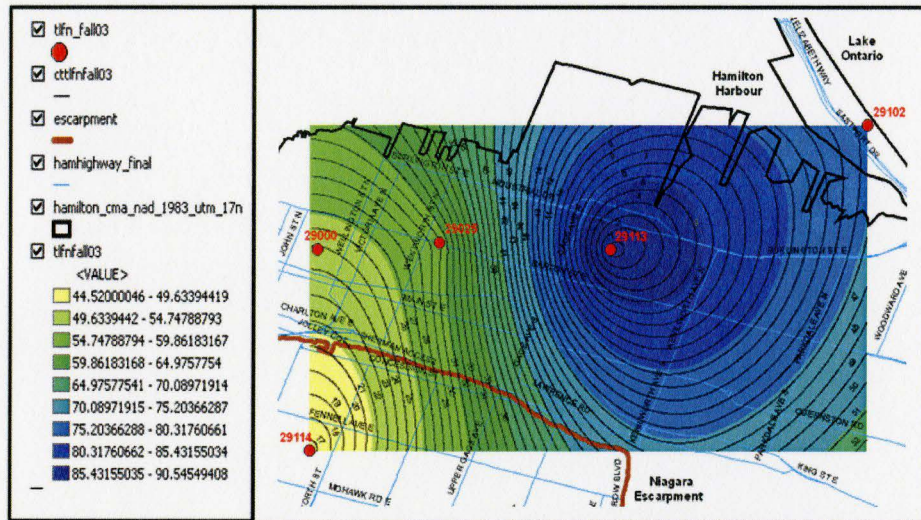


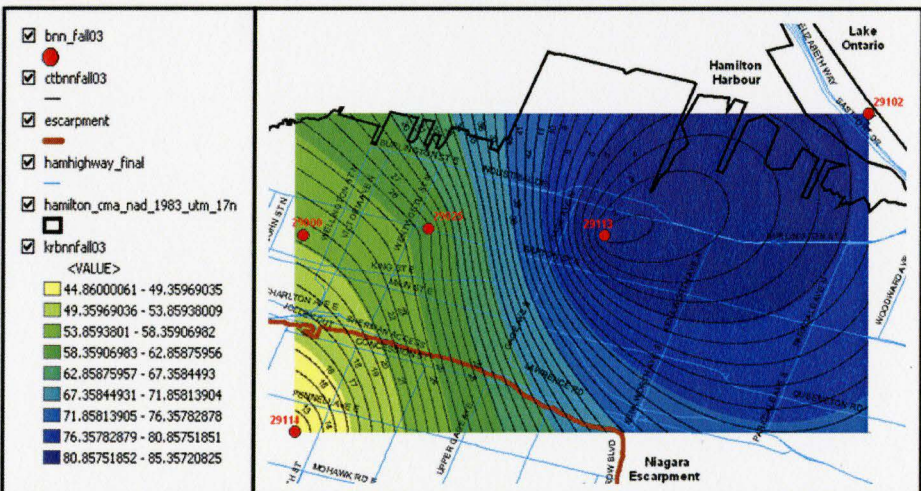
Fig. A. 3 (a): Comparison of observed and simulated values over space during fall 2003



(a) Observed



(b) TLFN



(c) BNN

Fig. A. 3 (b): Comparison of observed and simulated values over space during fall 2003

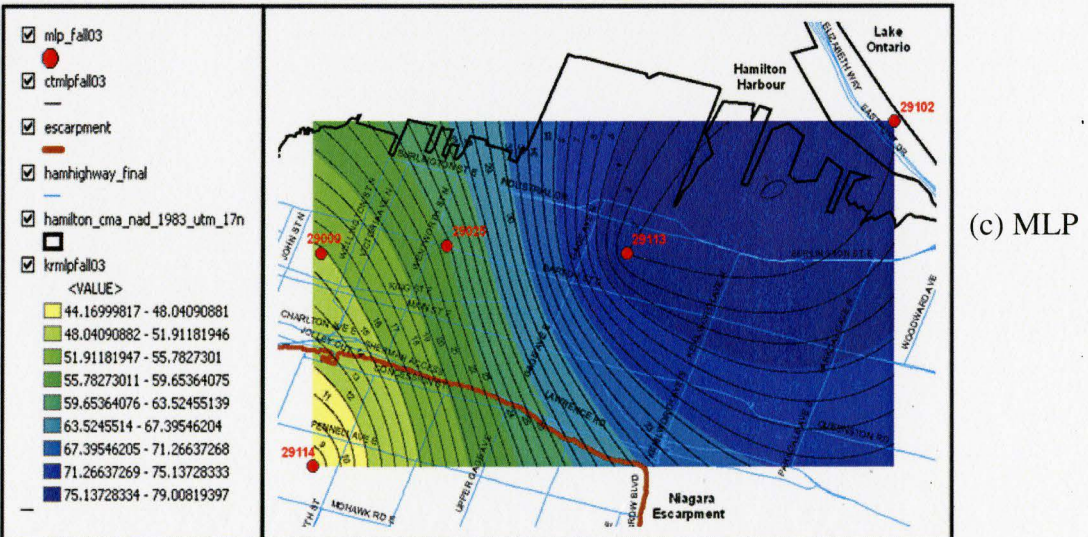
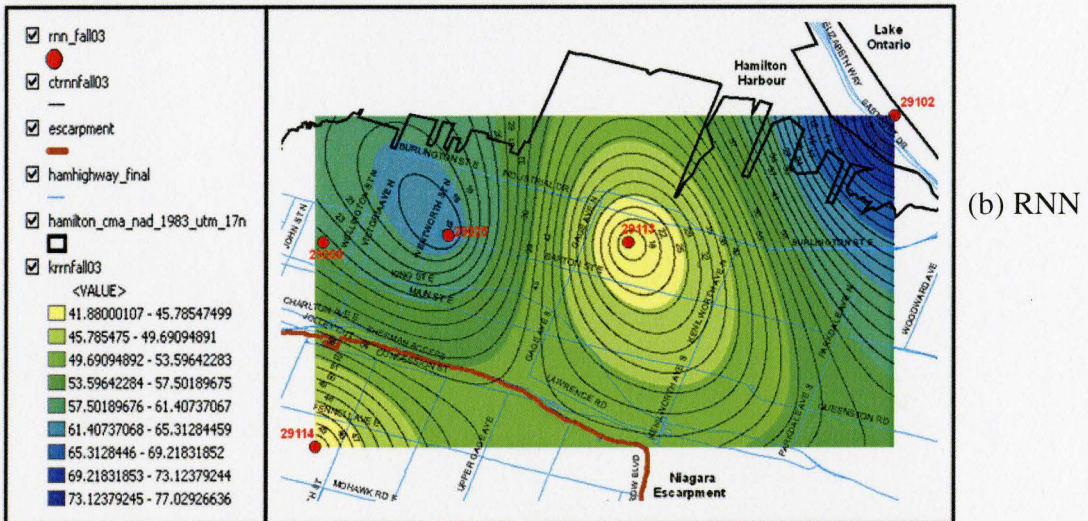
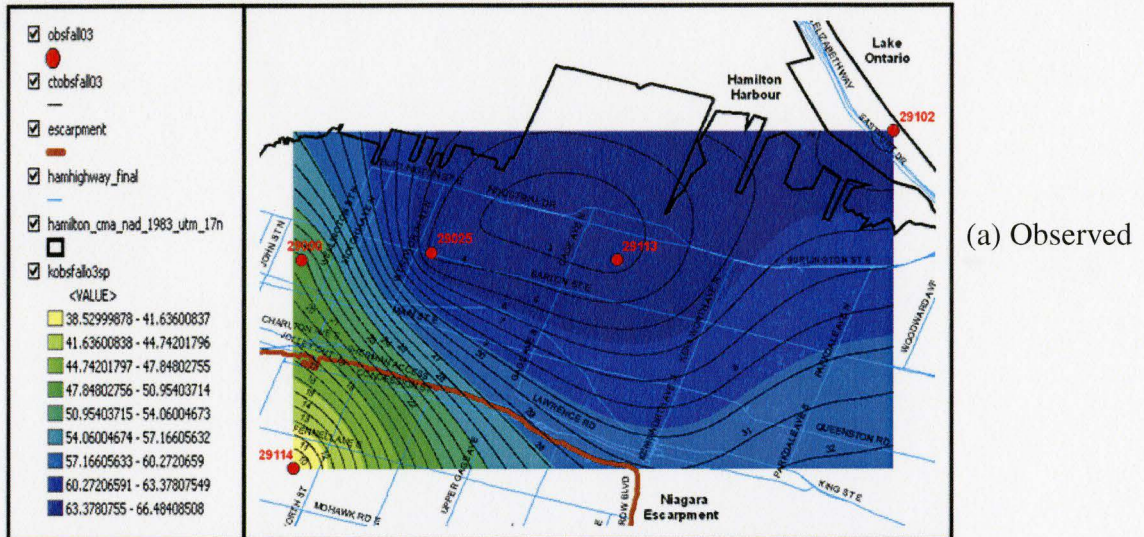


Fig. A. 4: Scatter plots for PM₁₀ concentration at station 302

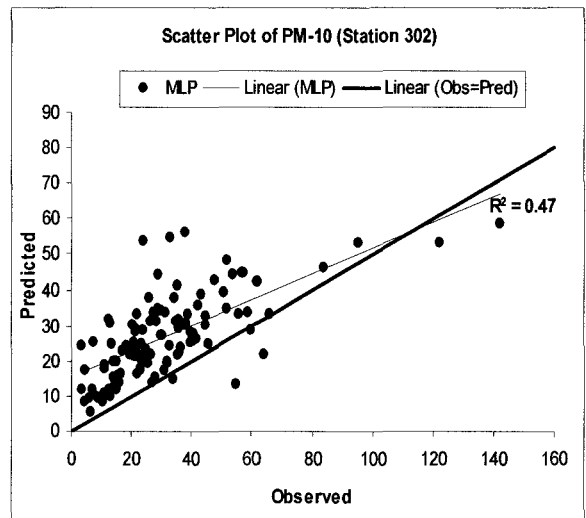
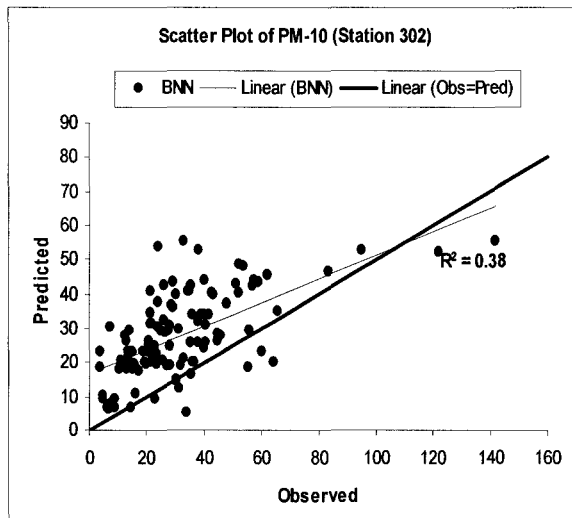
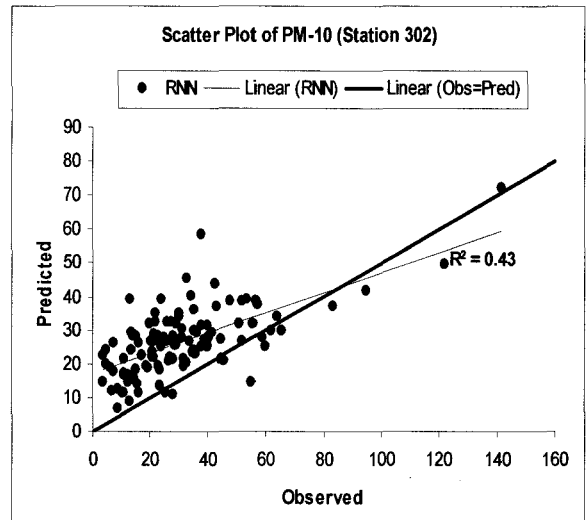
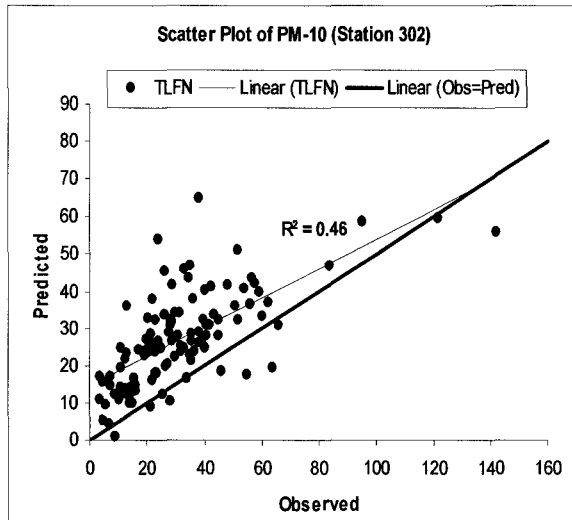


Fig. A. 5: Simulation result of PM₁₀ at station 302 with 95% confidence interval

