# Air Quality Prediction by Machine Learning Methods

by

Huiping Peng

B.Sc., Sun Yat-sen University, 2013

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies

(Atmospheric Science)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2015

# Abstract

As air pollution is a complex mixture of toxic components with considerable impact on humans, forecasting air pollution concentration emerges as a priority for improving life quality. In this study, air quality data (observational and numerical) were used to produce hourly spot concentration forecasts of ozone ($O_3$), particulate matter 2.5$\mu$m ($PM_{2.5}$) and nitrogen dioxide ($NO_2$), up to 48 hours for six stations across Canada – Vancouver, Edmonton, Winnipeg, Toronto, Montreal and Halifax. Using numerical data from an air quality model (GEM-MACH15) as predictors, forecast models for pollutant concentrations were built using multiple linear regression (MLR) and multi-layer perceptron neural networks (MLP NN). A relatively new method, the extreme learning machine (ELM), was also used to overcome the limitation of linear methods as well as the large computational demand of MLP NN. In operational forecasting, the continuous arrival of new data means frequent updating of the models is needed. This type of learning, called online sequential learning, is straightforward for MLR and ELM but not for MLP NN. Forecast performance of the online sequential MLR (OSMLR) and online sequential ELM (OSELM), together with stepwise MLR, all updated daily were compared with MLP NN updated seasonally, and the benchmark, updatable model output statistics (UMOS) from Environmental Canada. Overall OSELM tended to slightly outperform the other models including UMOS, being most successful with ozone forecasts and least with $PM_{2.5}$ forecasts. MLP NN updated seasonally was generally underperforming the linear models MLR and OSMLR, indicating the need to update a nonlinear model frequently.

# Preface

This thesis contains research conducted by the candidate, Huiping Peng, under the supervision of Dr. William Hsieh, Dr. Alex Cannon. The air quality and meteorological data sets used in this study were provided by Dr. Andrew Teakles (Environment Canada). Fig 3.1 was reproduced using the station data from Environment Canada. The ELM and MLP NN model in this thesis were based on R packages developed by Aranildo Lima and Alex Cannon. The supervisory committee provided the original research topic, direction and critical feedback on the research methods. The development of statistical air quality models and the analysis of results were primarily the work of the candidate, but William Hsieh, Alex Cannon and Andrew Teakles contributed substantially by suggesting specialized analysis techniques, by helping to interpret the results and by carefully editing the manuscript. Currently no part of this thesis has been published, but a paper based on this thesis is undergoing preparation for submission.

# Table of Contents

# List of Tables

# List of Figures

# List of Abbreviations

**AQHI** Air quality health index

**ELM** Extreme Learning Machine

**GEM-MACH15** Global environmental multi-scale model - modeling air quality and chemistry with 15-km grid spacing

**LST** Local sidereal time

**MAD** Mean absolute deviation

**MAE** Mean absolute error

**MLP NN** Multilayer perceptron neural network

**MLR** Multiple linear regression

**$NO_2$** Nitrogen dioxide

**$O_3$** Ozone

**OSELM** Online sequential extreme learning machine

**OSMLR** Online sequential multiple linear regression

**$PM_{2.5}$** Particulate matter 2.5 $\mu$m

**r** Pearson correlation coefficient

**RMSE** Root mean square error

**SS** Skill score

**UMOS** Updatable model output statistics

**UTC** Coordinated universal time

# Acknowledgements

I would like to thank Prof. William Hsieh for his never-ending guidance and support. The time that was put into this research could not have been done without his coordination and exceptional knowledge of Atmospheric Science and machine learning methods. His constant support and considerate attitude allowed the completion of this research without any pressure or tension. I would also like to thank Dr. Alex Cannon for his commitment to learning and advisement for this research. I would also like to acknowledge and thank the members of the UBC climate prediction group: Aranildo Rodrigues, Jian Jin and Andrew Snauffer. Not only were they my fellow students who helped and supported, but they were also my friends. I would like to thank Andrew Teakles and Jonathan Baik, for providing data and detailed instruction in each phase of the research, for approving my research and inspiring me through their excellent experience with Environment Canada. I would like to thank Prof. Roland Stull and Prof. Susan Allen, who taught me courses, and the knowledge gained was able to contribute to my research. Lastly I would like to thank my parents and all of my friends for their love and encouragement, they are always my best support no matter where I am.

# Chapter 1

# Introduction

With economic development and population rise in cities, environmental pollution problems involving air pollution, water pollution, noise and the shortage of land resources have attracted increasing attention. Among these, air pollution's direct impact on human health through exposure to pollutants has resulted in an increased public awareness in both developing and developed countries (Kim et al., 2013; Kurt and Oktay, 2010; McGranahan and Murray, 2003). Air pollution is usually caused by energy production from power plants, industries, residential heating, fuel burning vehicles, natural disasters, etc. Human health concern is one of the important consequences of air pollution, especially in urban areas. The global warming from anthropogenic greenhouse gas emissions is a long-term consequence of air pollution (Nordiska, 2008; Ramanathan and Feng, 2009; Kumar and Goyal, 2013). Accurate air quality forecasting can reduce the effect of a pollution peak on the surrounding population and ecosystem, hence improving air quality forecasting is an important goal for society.

## 1.1 Background

Air pollution is the introduction of particulates, biological molecules, or other harmful materials into the Earth's atmosphere, causing disease, death to humans, damage to other living organisms such as food crops, or damage to the natural or man-made environment. An air pollutant is a substance in the air that can have adverse effects on humans and the ecosystem. The substance can be solid particles, liquid droplets, or gases. Pollutants are classified as primary or secondary. Primary pollutants are usually produced from a process, such as ash from a volcanic eruption. Other examples include carbon monoxide gas from motor vehicle exhaust, or sulfur dioxide released from factories. Secondary pollutants are not emitted directly. Rather, they form in the air when primary pollutants react or interact. Ground level ozone is a prominent example of a secondary pollutant. The six "criteria pollutants" are ground level ozone ($O_3$), fine particulate matter ($PM_{2.5}$), carbon monoxide (CO), nitrogen dioxide ($NO_2$),

sulfur dioxide ($SO_2$), and lead, among which ground level $O_3$, $PM_{2.5}$ and $NO_2$ (main component of $NO_x$) are the most widespread health threats.

Ground level $O_3$, a gaseous secondary air pollutant formed by complex chemical reactions between $NO_x$ and volatile organic compounds (VOCs) in the atmosphere, can have significant negative impacts on human health (Chen et al., 2007; Brauer and Brook, 1997). Prolonged exposure to $O_3$ concentrations over a certain level may cause permanent lung damage, aggravated asthma, or other respiratory illnesses. Ground level $O_3$ can also have detrimental effects on plants and ecosystems, including damage to plants, reductions of crop yield, and increase of vegetation vulnerability to disease (EPA, 2005).

Particle pollution (also called particulate matter or PM) is the term for a mixture of solid particles and liquid droplets found in the air. Some particles, such as dust, dirt, soot, or smoke, are large or dark enough to be seen with the naked eye. Others are so small they can only be detected using an electron microscope. Fine particulate matter ($PM_{2.5}$) consisting of particles with diameter 2.5 $\mu$m or smaller, is an important pollutant among the criteria pollutants. The microscopic particles in $PM_{2.5}$ can penetrate deeply into the lungs and cause health problems, including the decrease of lung function, development of chronic bronchitis and nonfatal heart attacks. Fine particles can be carried over long distances by wind and then deposited on ground or water through dry or wet deposition. The wet deposition is often acidic, as fine particles containing sulfuric acid contribute to rain acidity, or acid rain. The effects of acid rain include changing the nutrient balance in water and soil, damaging sensitive forests and farm crops, and affecting the diversity of ecosystems. $PM_{2.5}$ pollution is also the main cause of reduced visibility (haze) (EPA, 2005).

Nitrogen dioxide ($NO_2$) is one of a group of highly reactive gases known as "nitrogen oxides" ($NO_x$). US Environmental Protection Agency (EPA) Ambient Air Quality Standard uses $NO_2$ as the indicator for the larger group of nitrogen oxides. $NO_2$ forms quickly from emissions of automobiles, power plants, and off-road equipment. In addition to contributing to the formation of ground-level ozone, and fine particle pollution, current scientific evidence links short-term $NO_2$ exposures, ranging from 30 minutes to 24 hours, with adverse respiratory effects including airway inflammation in healthy people and increased respiratory symptoms in people with asthma (EPA, 2005).

The Air Quality Health Index (AQHI) is a public information tool designed in Canada to help understand the impact of air quality on health. Basically, the AQHI is defined as an index or rating scale range from 1 to 10+ based on mortality study to indicate the level of health risk associated with local air quality (Chen and Copes, 2013). The higher the number, the greater the health risk and the need to take precautions. The formulation of Canadian national AQHI is based on three-hour average concentrations of ground-level ozone ($O_3$), nitrogen dioxide ($NO_2$), and fine particulate matter ($PM_{2.5}$). The AQHI is calculated on a community basis, each community may have one or more monitoring stations and the average concentration of 3 substances is calculated at each station within a community for the 3 preceding hours. AQHI is a meaningful index protecting residents on a daily basis from the negative effects of air pollution. Our study gives direction to predicting individual pollutants of one hour average concentration instead of AQHI (or its maximum) as the formulation of AQHI is based on health related science and may evolve over time. Building a forecast system based on individual pollutants and one hour average concentration will make it more flexible to future changes in health indices. Our result can also be beneficial to external clients and meteorologists.

The concentration of air pollutants including ground level ozone, $PM_{2.5}$ and $NO_2$ varies depending on meteorological factors, the source of pollutants and the local topography (Dominick et al., 2012). Among these three factors, the one which most strongly influences variations in the ambient concentration of air pollutants is meteorological factors (Banerjee and Srivastava, 2009). Meteorological factors experience complex interactions between various processes such as emissions, transportation and chemical transformation, as well as wet and dry depositions (Seinfeld and Pandis, 1997; Demuzere et al., 2009). In addition, the spatial and temporal behavior of wind fields are affected by the surface roughness and differences in the thermal conditions (Oke et al., 1989; Roth, 2000), which further influence the dispersion of pollutants. For example, Revlett (1978) and Wolff and Lioy (1978) found that ambient ozone concentration not only depended on the ratio and reactivity of precursor species, but also on the state of the atmosphere - the amount of sunlight, ambient air temperature, relative humidity, wind speed, and mixed layer (ML) depth, while Tai (2012) found that daily variations in meteorology as described by the multiple linear regression (MLR) including nine predictor variables (temperature, relative humidity, precipitation, cloud cover, 850-hPa geopotential height, sea-level pressure tendency, wind speed and wind direction) could explain up to 50% of the daily $PM_{2.5}$ variability in the US. Hence, me-

teorological factors play an important role in air pollutant concentrations, also making them difficult to model.

Most current air quality forecasting uses straightforward approaches like box models, Gaussian models and linear statistical models. Those models are easy to implement and allow for the rapid calculation of forecasts. However, they usually do not describe the interactions and non-linear relationship that control the transport and behaviour of pollutants in the atmosphere (Luecken et al., 2006). With these challenges, machine learning methods originating from the field of artificial intelligence have become popular in air quality forecasting and other atmospheric problems (Comrie, 1997; Hadjiiski and Hopke, 2000; Reich et al., 1999; Roadknight et al., 1997; Song and Hopke, 1996). For instance, several neural network (NN) models have already been used for air quality forecast, in particular for forecasting hourly averages (Kolehmainen et al., 2001; Perez et al., 2000) and daily maximum (Perez, 2001). Although NN have advantages over traditional statistical methods in air quality forecasting, NN-based models still need to improve in order to achieve good prediction performance as effectively and efficiently as possible (Wang et al., 2003). A number of difficulties associated with NN hamper their effectiveness in air quality forecasting. These difficulties include computational expense, multiple local minima during optimization, over-fitting to noise in the data, etc. Furthermore, there are no general rules to determine the optimal size of network and learning parameters, which will greatly affect the prediction performance.

Another key consideration of forecast models is their updatability when doing real-time forecasting. For a forecast model, recently observed data should be used to refine the model. This generally follows a procedure that links the discrepancy between model forecasts and the corresponding latest observation to all or some of the parameters in model. Normally there are two ways for model updating: batch learning and online learning. Whenever new data are received, batch learning uses the past data together with the new data and performs a retraining of the model, whereas online learning only uses the new data to update the model. Batch learning can be computationally expensive in real-time forecasting as the procedure means repeatedly altering a representative set of parameters calibrated over a long historical record. Linear models are generally easy to update online (Wilson and Vallée, 2002), and even with batch learning, linear models are fast and easy to implement. As for non-linear methods, true online learning is difficult for many formulations such as the non-linear kernel method. Furthermore, short time (daily) update via batch learning is too expensive

to implement as a non-linear model tends to have more parameters to train and the training process is much slower compared to linear models. Consequently, there is a need to develop non-linear updatable models for real-time forecasting. This study attempts to use the extreme learning machine (ELM) (Schmidt et al., 1992; Huang et al., 2006b), a non-linear machine learning algorithm using randomized neural networks, to forecast air pollutant concentrations in Canada. The ELM model has an architecture similar to the multi-layer perceptron (MLP) NN model, but it can be used for online sequential learning. ELM has been successfully used in different research areas and has been found to produce good generalization performance with generally less learning time compared with traditional gradient-based NN training algorithms (Huang et al., 2011; Lima et al., 2015).

## 1.2 Research Objectives

The research goal of this study is to develop a non-linear updatable model for real-time air quality forecasting, to potentially replace the updatable linear regression models currently being used. The ultimate goal is to improve air pollution forecasting in Canada and in other countries.

## 1.3 Organization of Thesis

Chapter 2 provides a literature review covering topics related to air quality forecasting, machine learning techniques and updatable model output statistics (UMOS), a linear online updating model from Environment Canada. Background theory on various machine learning and air quality topics will be covered. The reviewed air quality forecasting studies as well as the modeling techniques will be discussed on how they can be applied to this research. Chapter 3 describes the study area and the data sets used in this study. Chapter 4 outlines the methods used to conduct this research and describes the developed forecast models and evaluation methods. In Chapter 5 the results from all developed forecast models for each pollutant are discussed in detail. The thesis concludes with Chapter 6 where the original research objectives are addressed and recommendations for future research are made.

# Chapter 2

# Literature Review

## 2.1 Introduction

Air pollution is major threat to health and exerts a wide range of impacts on biological and economic systems. The purpose of this literature review is to justify the research objectives of this study in light of previous work by investigating past air quality prediction studies and determining where future research is needed. Literature related to air quality prediction and various types of machine learning methods used in this study are reviewed. Machine learning theory and past applications are examined to show why these methods are likely to perform well in air quality forecasting.

## 2.2 Machine Learning Techniques

Machine learning is a major sub-field in computational intelligence (also called artificial intelligence). Its main objective is to use computational methods to extract information from data. Machine learning has a wide spectrum of applications including handwriting and speech recognition, robotics and computer games, natural language processing, brain-machine interface and so on. In the environmental sciences, machine learning methods have been heavily used in data processing, model emulation, weather and climate prediction, air quality forecasting, oceanographic and hydrological forecasting. (Hsieh, 2009).

### 2.2.1 Neural Network

Neural network (NN) methods were originally developed from investigations into human brain function and they are adaptive systems that change as they learn (Hsieh and Tang, 1998). There are many types of NN models, the most common one is the multi-layer perceptron (MLP) NN model shown in Fig 2.1

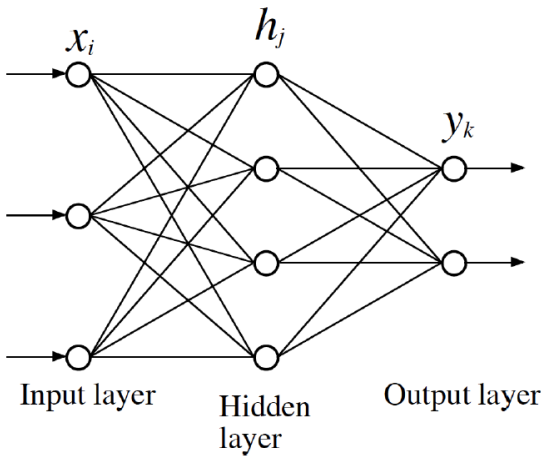



Figure 2.1: The general structure of a MLP NN model (Hsieh, 2009).

The input variables $x_i$ are mapped to a layer of intermediate variables known as "hidden neuron" $h_j$ by

$$h_j = f(\sum_i w_{ji} x_i + b_j), \tag{2.1}$$

and then onto the output variables $y_k$ by

$$y_k = g(\sum_j \beta_{kj} h_j + \beta_{k0}). \tag{2.2}$$

where $f$ and $g$ are "activation" functions in the hidden layer and the output layer, respectively. Normally $f$ can be the logistic sigmoidal or hyperbolic tangent function and $g$ can be linear in NN models for regression. $w_{ji}$ and $\beta_{kj}$ are weight parameters and $b_j$ and $\beta_{k0}$ are offset parameters. Their optimal values are learned by model training (Hsieh and Tang, 1998) where the mean squared error of the model output is minimized.

Numerous studies show NN models have good forecasting performance. Walter et al. (1998) used NN methods to simulate the observed global (and hemispheric) annual mean surface air temperature variations during 1874-1993 using anthropogenic and natural forcing mechanisms as predictors. The two anthropogenic forcings were equivalent $CO_2$ concentrations and tropospheric sulfate aerosol concentrations. The natural forcing were volcanism, solar activity and ENSO (El Niño Southern Oscillation). The NN explained up to 83% of the observed temperature variance, significantly more than by multiple regression analysis. Hewitson and Crane (1996) used MLP NN for precipi-

tation forecast with predictors from the general circulation model (GCM) atmospheric data over southern Africa and the surrounding ocean. The six leading PCs (principal components) of the sea level pressure field and the seven leading PCs of the 500 hPa geopotential height field from the GCM were used as inputs to the NN. Cavazos (1997) also used MLP NN to downscale GCM synoptic-scale atmospheric circulation to local $1^\circ \times 1^\circ$ gridded winter daily precipitation over north-eastern Mexico and found the model was able to reproduce the phase and, to some degree, the amplitude of large rainfall events. Marzban and Stumpf (1996) trained an MLP to predict the existence of tornadoes. The approach outperformed other techniques including discriminant analysis, logistic regression and a rule-based algorithm. Neural networks have also been used to solve hydrological problems, such as the prediction of reservoir inflows, stream flow forecasting, downscaling precipitation, and prediction of water resource variables (e.g., flow, water level, nitrate, salinity and suspended sediment concentration) (Tripathi et al., 2006; Chen et al., 2010; Maier et al., 2010; Cannon, 2012b; Rasouli et al., 2012; Thirumalaiah and Deo, 1998). Other applications of neural network methods include remote sensing and GIS related activities, air quality management (Boznar et al., 1993), adsorbent beds design (Basheer and Najjar, 1996), and hazardous waste management.

Researchers have shown that neural networks have salient advantages over traditional statistical methods in environment forecasting problems. However, a number of difficulties associated with NN hamper their effectiveness, efficiency and general acceptability (Wang et al., 2003). One of the main challenges in developing a NN model is how to address the problem of over-fitting. An over-fitted NN model could fit the data very well during training, but produce poor forecast results during testing (Hsieh and Tang, 1998). Over-fitting occurs when a model fits to the noise in the data and it will not generalize well to new data sets as shown in Fig 2.2.

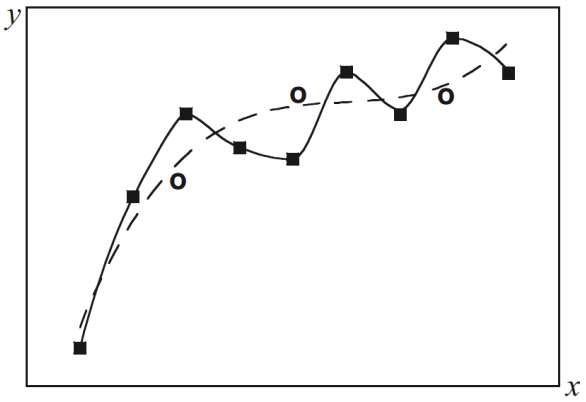



Figure 2.2: A diagram illustrating the problem of over-fitting. The dash curve shows a good fit to noisy data (squares), while the solid curve illustrate over-fitting, where the fit is perfect on the training data (squares), but is poor on the test data (circles) (Hsieh and Tang, 1998).

Typically regularization (i.e. the use of weight penalty) is used to prevent over-fitting (Golub et al., 1979; Haber and Oldenburg, 2000). This usually requires some of the training data to be used as validation data to determine the optimal regularization parameter to prevent over-fitting. Yuval (2000) introduced generalized cross-validation (GCV) to control overfitting/underfitting automatically in MLP NN and applied the method to forecasting the tropical Pacific SST anomalies. Yuval (2001) used bootstrap resampling of the data to generate an ensemble of MLP NN models and used the ensemble spread to estimate the forecast uncertainty.

Another issue is the computational expense involved during the training process in neural networks. Training the NN model to learn from the target data, we need to minimize the objective function $J$, defined here to be mean squared error (MSE) between the model output $y$ and the target $t$. Normally the back-propagation algorithm is used to perform the training tasks, using a gradient-descent approach to reduce the MSE iteratively (Hsieh, 2009), which could be time-consuming.

### 2.2.2 Extreme Learning Machine (ELM)

The extreme learning machine (ELM) is a randomized neural network method proposed by Schmidt et al. (1992) and popularized by Huang et al. (2006b). The ELM algorithm has the same architecture as a single-hidden layer feed-forward neural network (SLFN), but it is generally fast to train. The ELM randomly chooses the weights leading to the

hidden nodes or neurons (HN) and analytically determines the weights at the output layer by solving a linear least squares problem. The only hyper-parameter to be tuned in the ELM is the number of HN (Lima et al., 2015). Extensions to ELM includes online sequential ELM (OS-ELM)(Liang et al., 2006), incremental ELM (I-ELM)(Huang et al., 2006a; Huang and Chen, 2007, 2008), ELM ensembles, pruning ELM (P-ELM) (Rong et al., 2008) and error minimized ELM (EM-ELM) (Feng et al., 2009). In online sequential learning, new data arrive continuously and the model is repeatedly updated with the new data. OS-ELM is readily updated using only the new data, without the need to retrain using the complete historical record.

ELM has also been successfully used in different research areas (Huang et al., 2011). An integration of several ELMs was proposed by Sun et al. (2008) to predict the future sales amount. Several ELM networks were connected in parallel and the average of the ELMs outputs was used as the final predicted sales amount, with better generalization performance. Heeswijk et al. (2009) investigated the adaptive ensemble models of ELM on the application of one-step ahead prediction in stationary and non-stationary time series. They found that the method worked well on stationary time series and the adaptive ensemble model achieved a test error comparable to the best methods on the non-stationary time series, while keeping adaptivity with low computational cost. Handoko et al. (2006) found that the ELM was as good as, if not better than, the MLP NN in terms of computing time, accuracy deviations across experiments and prevention of overfitting. Using the ELM as a mechanism for learning the stored digital elevation information to allow multi-resolution access in terrain models, Yeu et al. (2006) found that to achieve the same MSE during access, the memory needed in ELM was significantly lower than that needed by Delaunay triangulation (DT). Additionally, the offline training time for the ELM network was much less than that for the MLP NN, DT and support vector machines (SVM).

As a randomized neural network, ELM is controlled by hyper-parameters such as the range of the random weights and the number of HN. The optimal number of HN is problem dependent and unknown in advance. We have to ensure the network structure is balanced between generalization ability and network complexity. Low network complexity (i.e. too few HN) might be unable to capture the true non-linear relationship, whereas too high a network complexity might decrease model generalization ability due to overfitting to noise in the data, and increase the model training time. In general, the number of HN is selected empirically based on model performance over indepen-

dent validation data not used in model training. For an ELM, the optimal number of HN found can be much greater (sometimes by orders of magnitude) than MLP using iterative non-linear optimization, so an automatic procedure is needed to select the number of HN. Lima et al. (2015) used the hill climbing method to find the optimal number of HN and the test results on nine environmental regression problems showed that among the non-linear models, the ELM method, with often the fastest computing time for model training, tended to perform well in prediction skills.

Random initialization of the weights is another important prerequisite for good convergence of NN models. A balance must exist to ensure that the activation function does not remain linear nor become saturated near the asymptotic limits of -1 and 1 in the case of a hyperbolic tangent function. If the range of random weight distribution is too small, both activation and error signals will die out on their way through the network. If it is too large, the saturated activation function will block the back-propagated error signals from passing through the node (Lima et al., 2015). Normally the range of weight interval is simply a constant, but Parviainen and Riihimaki (2013) raised questions about meaningfulness of choosing model complexity based on HN only and also found that using an appropriate weight range can improve ELM performance, achieving a similar effect as regularization in traditional neural network models.

## 2.3 Air Quality Forecasting Models

An air quality model is a numerical tool used to describe the causal relationship between emissions, meteorology, atmospheric concentration, deposition and other factors. It can give a complete deterministic description of the air quality problem (Nguyen, 2014). The most commonly used air quality models include dispersion models, photochemical models and regression models. Various neural network models, as non-linear regression models, have also been shown to be effective in air quality forecasting. In this section, different models and their applications will be introduced.

### 2.3.1 Dispersion Models

Dispersion models normally use mathematical formulations to simulate the atmospheric process after pollutants were emitted by a source. Data needed for dispersion models

vary in their complexity. At a minimum, most of the models require meteorological data, emissions data, and details about the facilities in question (such as stack height, gas exit velocity, etc.). Some of the more complex models require topographic information, individual chemical characteristics and land use data. The output is predicted concentration at selected downwind receptor locations. There are different types of dispersion models with specific requirement and special scales. The most commonly used dispersion models are the box model, Gaussian plume model, Lagrangian model, Eulerian model, computational fluid dynamics model and Gaussian puff model. The processes included in those models are building wake effects, topography, street canyon, intersections, plume rise and chemistry (Holmes and Morawska, 2006).

Two of the most common models used to calculate the dispersion of vehicle emissions are CALINE4 (California Department of Transportation) and HIWAY2 (US EPA). Both models are based on a Gaussian plume model. Yura et al. (2007) explored the range of CALINE4's $PM_{2.5}$ modeling capabilities by comparing previously collected $PM_{2.5}$ data with CALINE4 predicted values. Two sampling sites, a suburban site and an urban were used for this study. Model predicted concentrations are graphed against observed concentrations and evaluated against the criterion that 75% of the points fall within the factor-of-two prediction envelope. However, only the suburban site results by CALINE4 met the criterion. For urban site, several factors including street canyon effects likely contributed to an inaccuracy of the emission factors used in CALINE4, and therefore, to the overall CALINE4 predictions. The study suggested that CALINE4 might not perform well in densely populated areas and differences in topography may be a decisive factor in determining when CALINE4 may be applicable to modeling $PM_{2.5}$.

Colvile et al. (2002) applied the ADMS-urban atmospheric dispersion model system to review air quality in central London in 1996-1997. The model performance was validated by monitoring data and showed that model precision was 10% with 0-12% bias for the annual mean $NO_2$ and $PM_{10}$ concentrations. Wallace and Kanaroglou (2008) used the Integrated Model of urban Land-use and Transportation for Environmental Analysis to estimate emission and concentrations of $NO_x$ from traffic sources in the Hamilton census metropolitan area. The results showed a prominent triangle area of high pollution, which is defined by major roads and highways along the Hamilton Harbour during peak hour. The resulting dispersion surfaces characterized the spatial distribution of traffic emissions and thus provide a way for assessing population expo-

sure over the Hamilton area.

Although dispersion models consider many processes that affect air pollutant concentration, they have some limitations such as the simplified treatment of turbulence and meteorology, and cannot take into account any formation of pollutants. Even $NO_x$ and $SO_x$, which are fundamental to determining particles and ozone concentrations, are often only calculated using a simple exponential decay (Holmes and Morawska, 2006). Gaussian models have also been shown to consistently over predict concentrations in low wind conditions (Benson, 1984), as Gaussian models are not designed to model the dispersion under low wind conditions.

### 2.3.2 Photochemical Models

Photochemical models have become widely utilized as tools in air pollution control strategies. Photochemical models simulate the changes of pollutant concentrations in the atmosphere using a set of mathematical equations characterizing the chemical and physical processes in atmosphere. These models are applied at multiple spatial scales from local, regional, national, and global (Nguyen, 2014).

Photochemical models have been formulated in both the Lagrangian and Eulerian reference frames (Russell and Dennis, 2000). Eulerian models include both single box models and multi-dimensional grid-based models. Box models were used early and are still used today in studies focusing on atmospheric chemistry alone. The limitation of box models is a lack of significant physical realism such as horizontal and vertical transport, and spatial variation. Grid models are potentially the most powerful photochemical model (Hansen et al., 28; Dennis et al., 1996), but are also the most computationally intensive. They solve a finite approximation by dividing the modeling region into a large number of cells, horizontally and vertically, which interact with each other to simulate the various processes that affect the evolution of pollutant concentrations, including chemistry, diffusion, advection, sedimentation (for particles), and deposition.

Photochemical models have been widely used to assess the relative importance of VOC and $NO_x$ controls in reducing ozone levels. Milford et al. (1989) used the CIT model to show the spatial variation of ozone isopleths and found a negative response of

ozone to $NO_x$ controls in the downtown region of Los Angeles. Flemming et al. (2001) have employed the regional Eulerian model with 3 chemistry mechanisms (REM3), to operationally forecast ozone since 1997 at the Freie University, Berlin. The model has been used for making 1, 2, and 3 day advance ozone forecasts with data over Germany from 1997 to 1999. The resulted correlation coefficient ($r$) ranged from 0.77 to 0.90. The disadvantage of this model was that it tended to underestimate the low ozone concentrations. Wotawa et al. (1998) developed a Lagrangian photochemical box model for providing ozone forecasts for Vienna, Austria. This model consisted of up to 8 vertical and up to 5 horizontal boxes. It simulated emission, chemical reactions, horizontal diffusion, vertical diffusion, dry deposition, wet deposition and synoptic scale vertical exchange. Model input data included a trajectory term, which was calculated using forecast meteorological data. The model predictions for 1995 $O_3$ season underestimated $O_3$ concentrations on most days and $r$ was greater than 0.6 for most of the study cases.

### 2.3.3 Regression Models

Both linear regression and non-linear regression models have been employed for air quality forecasting. The general purpose of a linear regression model is to learn about the linear relationship between several independent variables (predictors) and a dependent variable (predictand).

Prybutok et al. (2000) built a simple linear regression model for forecasting the daily peak $O_3$ concentration in Houston. The final model used four meteorological and $O_3$ precursor parameters: $O_3$ concentration at 9:00 a.m., maximum daily temperature, average $NO_2$ concentration between 6:00 a.m. and 9:00 a.m. and average surface wind speed between 6:00 a.m. and 9:00 a.m. The correlation coefficient $r$ of this model was 0.47. Chaloulakou et al. (1999) proposed a multiple regression model to forecast the next day's hourly maximum $O_3$ concentration in Athens, Greece. The set of input variables consisted of eight meteorological parameters and three persistence variables, which were the hourly maximum $O_3$ concentrations of the previous three days. Testing this linear regression model on four separate test data sets, the mean absolute error (MAE) ranged from 19.4% to 33.0% of the corresponding average $O_3$ concentrations.

Non-linear regression models are superior to simple linear regression models because they capture the non-linear relationships between air pollutant and meteorological pa-

rameters. Bloomfield et al. (1996) described a non-linear regression model to explain the effects of meteorology on $O_3$ in the Chicago area. The model input variables consisted of a seasonal term, a linear annual trend term, and twelve meteorological variables. The observed ozone and meteorological data in 1981-1991 were divided into subsets for model development and validation. The model error were within $\pm 5$ ppb about half the time, and within $\pm 16$ ppb about 95% of the time. Bloomfield et al. (1996) demonstrated that the meteorological data accounted for at least 50% of the ozone concentration variance.

As the reference model in this thesis, the updatable model output statistics - air quality (UMOS-AQ) system applies multiple linear regression (MLR) to forecast air quality predictands. UMOS-AQ is a statistical post-processing system for air quality forecasting in Canada. The current Environment Canada (EC) operational AQ forecast model is the GEM-MACH15 (global environmental multi-scale model - modeling air quality and chemistry with 15-km grid spacing). GEM-MACH15 runs twice daily at 00 and 12 UTC to give 48-hour AQ forecasts (Anselmo et al., 2010). UMOS-AQ is based on post-possessing the GEM-MACH15 forecasts. The UMOS post-processing package has been used by EC to forecast meteorological predictands such as surface temperature and probability of precipitation since 1995 (Wilson and Vallée, 2002, 2003). UMOS-AQ uses the existing UMOS framework and became operational in July 2010. Three predictands are currently considered by UMOS-AQ: $O_3$, $PM_{2.5}$ and $NO_2$. Possible MLR predictors include $O_3$, $PM_{2.5}$ and $NO_2$ hourly concentrations at a station for each hour of the previous day (i.e., persistence) plus 84 other chemical, meteorological, and physical predictors (e.g., solar flux, sine of scaled Julian day). Two seasons (summer and winter) are considered with a transitional period of 6 weeks. A minimum of 250 observation-model pairs per season are needed to generate robust MLR equations and the equations are regenerated with the latest model data every week (Moran et al., 2014). One of UMOS-AQ's main advantages is its ability to adapt to the model changes, as its equations are updated four times per month. However, UMOS-AQ can only be constructed for locations where historical AQ measurements are available (Wilson and Vallée, 2002; Moran et al., 2014). This becomes a limitation because most AQ stations are not co-located with public weather forecast stations. A solution is to blend the UMOS-AQ point forecasts with GEM-MACH15 gridded forecast fields. This is now being done by optimal interpolation (OI) using MIST (Moteur d'Interpolation STatistique), an EC statistical interpolation package that uses the OI algorithm described by Mahfouf et al. (2007).

### 2.3.4 Neural Network Models

Although many approaches such as box models, Gaussian plume models, persistence and regression models are commonly applied to characterize and forecast air pollutants concentration, they are relatively straightforward with significant simplifications (Luecken et al., 2006).

A promising alternative to these models is the neural network model (Lal and Tripathy, 2012; Nejadkoorki and Baroutian, 2012; Gardner and Dorling, 1998). Several NN models have already been used for different air pollutant concentration forecast. Gardner and Dorling (2000) used MLP NN to forecast the hourly ozone concentration at five cities in UK and they found that NN outperformed both CART (classification and regression tree) and linear regression (LR). The predictors used included the amount of low cloud, base of lowest cloud, visibility, dry bulb temperature, vapour pressure, wind speed and direction. To account for seasonal effect, they had two extra predictors in model 2, $\sin(2\pi d/365)$ and $\cos(2\pi d/365)$, with $d$ the Julian day of the year, thereby informing the model where in the annual cycle the forecast was made. Ballester et al. (2002) used a finite impulse response NN model to make 1-day advance predictions of 8-hr average ozone concentrations in eastern Spain. The input variables were observed 2h lagged observed values of air quality and meteorological inputs. The models were evaluated using data from the 1996 to 1999 ozone seasons (July to September). The statistics of the model fits for three sampling sites ranged from 6.39 to 8.8 ppb for MAE and from 0.73 to 0.79 for $R$.

For particulate matter (PM), Kukkonen et al. (2003) compared the performance of five different NN models for the prediction of $PM_{10}$ concentrations in Helsinki. Results obtained showed that NN models performed better than linear models. In addition, Perez et al. (2000) constructed an NN $PM_{2.5}$ forecast model to make predictions of hourly averaged $PM_{2.5}$ concentrations in the downtown area of Santiago, Chile. Three forecast models, NN, LR, and persistence, were developed to predict $PM_{2.5}$ concentrations at any hour of the day, using the 24 hourly averaged concentrations measured on the previous day as the input variables. The normalized MAE (NMAE) of the predictions for 1994-1995 ozone season (May 1 to September 30) ranged from 30% to 60%. These authors found that $PM_{2.5}$ formation strongly depended on weather conditions, with the $PM_{2.5}$ concentrations negatively correlated with wind speed and relative humidity. $NO_2$ concentration have also been investigated using NN (Gardner and Dorling,

1999).

Several authors compared different approaches when applied to different pollutants and prediction time lags (Boznar et al., 1993; Lu and Wang, 2005; Yi and Prybutok, 2002). In the overview of NN application in the atmospheric sciences, Gardner and Dorling (1998) concluded that NN generally gives as good or better results than linear methods.

## 2.4 Summary

Studies from the fields of machine learning and air quality models show that much effort has been put into air quality forecasting, including the use of various machine learning methods. Machine learning methods have been widely used in environmental science problems and the applications of the MLP NN tend to provide some advantages over linear methods based on the results of the previous studies. In air quality forecasting, machine learning methods are promising when compared with the linear regression model and the photochemical dispersion model. The ELM method has been introduced to overcome some of the drawbacks in the popular MLP NN model, e.g. in computing time and the local minima problem.

# Chapter 3

# Data

## 3.1 Study Area

The updatable model output statistics - air quality (UMOS-AQ) model uses observations from more than 250 stations across Canada (Fig 3.1). The stations belong to the National Air Pollution Surveillance Network (NAPS), where each station measures all or a combination of the concentrations of ozone ($O_3$), fine particulates ($PM_{2.5}$) and nitrogen dioxide ($NO_2$) (Antonopoulos et al., 2012). Six stations across Canada are used for model testing: Vancouver International Airport (British Columbia), Edmonton Central (Alberta), Winnipeg (Manitoba), Toronto Downtown (Ontario), Montreal Airport (Quebec) and Halifax (Nova Scotia). These six stations include the largest cities of Canada, the coastal cities and the major center for oil and gas industry in Canada. They all have different topography, weather conditions and major pollution sources.

## 3.2 Data Set

The data set used in this study covers the period 2009/07-2014/07 and was provided by UMOS-AQ model of Environment Canada. The first two years of data (2009/07-2011/07) were for model training and validation and the final three years (2011/08-2014/07) were used for model testing as well as model updating. The model will be evaluated using these 3-year data sets. As mentioned before, UMOS-AQ is a post possessing system that combines multiple sources of information: AQ forecasts, meteorological forecasts, AQ measurements and physical variables. Hence, the input data sets consisted of observational and numerical data.

The observational air pollutant data were from automated near-real-time (NRT) hourly reports of local $O_3$, $PM_{2.5}$, and $NO_2$ concentrations from around 250 urban and rural AQ measurement stations located across Canada. The near-real-time data have

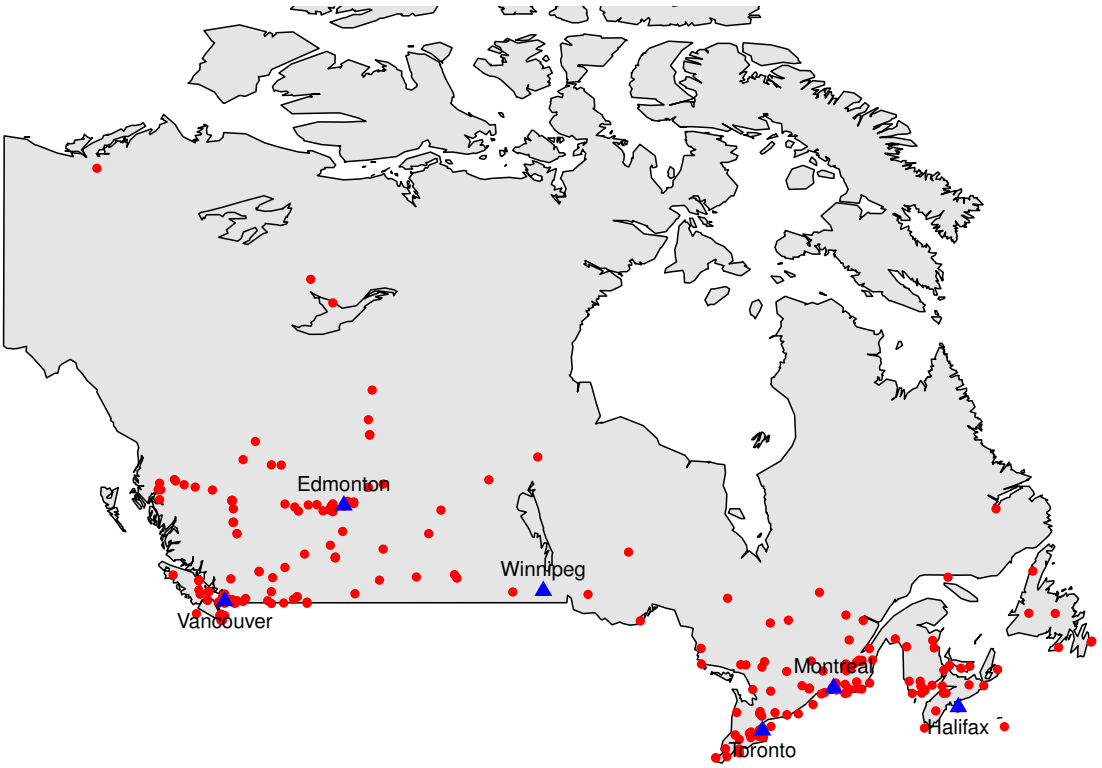



Figure 3.1: Geographical distribution of the UMOS-AQ stations (red dots), with the six stations selected in this study shown as blue triangles.

uncertainty as they have not been verified and may not adequately reflect representative air quality values. Numerical data came from an air quality model (GEM-MACH15), which was used to produce a set of direct and calculated predictors for our machine learning models. GEM-MACH15 produced both chemical and meteorological fields with its twice daily 48-h forecasts starting at 00 and 12 UTC. Using these predictors, our models produced hourly spot concentration forecasts up to 48 h.

The predictors used in this study includes persistence predictors, meteorological predictors, chemical and physical predictors. The persistence predictors included observed ozone, $PM_{2.5}$, and $NO_2$ concentration at the time the model was initiated (00 UTC and 12 UTC). Meteorological predictors consisted of model dry bulb temperature, wind component, geopotential height, relative humidity, dew point depression, surface model pressure, cumulative precipitation, average rate of snowfall in water equivalent, cloud cover, model boundary layer height, wind speed and dew point temperature. Chemical variables were the maximum and average ozone, $PM_{2.5}$, and $NO_2$ concentrations during 3, 6 and 24 hour interval. Physical variables included the downward solar flux, calculated mixing height, day of week, sine of the Julian day and calculated mixing height.

Three predictands (target data) were considered, namely the observed ozone, $PM_{2.5}$, and $NO_2$ hourly average concentrations in the six stations across Canada. Nine antecedent predictors were also considered: 1) the pollutant concentration valid at the same local sidereal time (LST) as the forecast but from 24 hours prior to the model initialization, 2) the maximum hourly average pollutant concentration observed within the 24 hour period prior to the model initialization, 3) the minimum hourly average pollutant concentration observed within the 24 hour period prior to the model initialization. As UMOS did not include antecedent predictors, we did not use the antecedent predictors when comparing against the UMOS benchmark.

Some statistical properties of the ozone, $PM_{2.5}$, and $NO_2$ concentrations during the study period (2009/07 - 2014/07) are shown in Table 3.1. The mean and standard deviation were calculated over the 5-year period, while the maximum values were the median of each year's maximum concentration with the median used to avoid influence from extreme events.

| Station | Ozone (ppb) | | | $PM_{2.5}$ ($\mu g/m^3$) | | | $NO_2$ (ppb) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 16.0 | 10.3 | 49.0 | 4.4 | 3.0 | 23.0 | 14.5 | 7.5 | 45.5 |
| Edmonton | 18.2 | 10.4 | 63.0 | 9.9 | 6.6 | 68.0 | 17.7 | 9.2 | 58.5 |
| Winnipeg | 25.9 | 11.7 | 66.5 | 5.8 | 3.7 | 37.0 | 6.5 | 6.3 | 39.5 |
| Toronto | 25.4 | 11.4 | 78.0 | 6.8 | 5.0 | 37.0 | 14.9 | 6.8 | 45.0 |
| Montreal | 23.6 | 10.9 | 67.0 | 9.5 | 6.1 | 47.0 | 9.8 | 6.9 | 52.0 |
| Halifax | 21.5 | 9.7 | 53.0 | 5.4 | 3.2 | 33.5 | 3.2 | 3.1 | 18.0 |

Table 3.1: Statistical properties of ozone concentration in 6 stations.

Table 3.1 shows that Toronto and Winnipeg have the highest mean ozone values. All six stations have less than 10 $\mu g/m^3$ mean $PM_{2.5}$ concentration, with Montreal and Edmonton being about 30% larger than others. The $NO_2$ mean value varies greatly among the six stations, as the mean in Halifax is only 3.2 ppb, whereas Toronto, Vancouver and Edmonton all have over 14 ppb mean. These statistics cannot provide a full assessment of the air pollutant concentration in each city as the statistics may be strongly influence by the location of each station within the city.

# Chapter 4

# Methods and Models Set up

In this chapter, several different methods and models are introduced. Each method produces a different model per station per forecast hour per pollutant. The model development can be separated into two phase: 1) training and 2) testing and updating. Models are first trained and validated using 2-year data sets (2009/07-2011/07) and after the initialization phase, the models are used to predict the air pollutant concentration with newly arrived single datum or a chunk of data during 2011/08-2014/07. Model updating is conducted by either batch learning algorithm or an online-sequential learning algorithm from the newly arrived data.

When data become available, batch learning performs a complete retraining of the model using all past data plus the new data. It can be used to update the multiple linear regression (MLR), multi-layer perceptron neural network (MLP NN) and extreme learning machine (ELM) methods. Depending on computation resources, batch updating can be applied daily, monthly or seasonally. Batch learning can be computationally intensive for nonlinear models as it may involve many iterations through the training data. There are many applications where online-sequential learning algorithms are preferred over batch learning algorithms as sequential learning algorithms do not require retraining with the full dataset whenever new data arrive (Liang et al., 2006).

A versatile online-sequential learning algorithm means the data for training are sequentially presented (singly or as a chunk of data) to the learning algorithm. At any time, only the newly arrived data (instead of all past data) are needed to update the model. The new data, once learned by the model, can be discarded (Liang et al., 2006). The learning algorithm has no prior knowledge as to how many training dataset will be presented. A comparison will be made between the online-sequential extreme learning machine (OS-ELM), the online-sequential multiple linear regression (OS-MLR) and the reference model, UMOS which is also online-sequential.

## 4.1 Input Data Preprocessing

Without properly transforming or scaling the input data, machine learning methods may be trained inefficiently and the resulting model may perform poorly. If the input variables in the training dataset vary greatly in magnitude, the model weights have to adapt to the differences. The resulting weights will also have a large spread in magnitude, rendering the training algorithm inefficient (Rasouli et al., 2012). Input data preprocessing/scaling is an efficient way to solve the problem. The commonly used scaling methods include: (i) linear transformation, (ii) statistical standardization, and (iii) nonlinear transformation (e.g. the logarithmic transformation). Input data in this study is standardized, i.e. data have the mean value subtracted, then divided by the standard deviation, yielding variables with zero mean and unit standard deviation. As separate forecast models are developed for the different hours of the day, there is no need to remove the diurnal cycle from the input data. Control filters with minimum concentration, maximum concentration and rate of change criteria were applied here to remove unrealistic low/high observations and to ensure reasonable rates of changes in the measurements.

## 4.2 Multiple Linear Regression (MLR)

Multiple linear regression models were developed in the free R software (R Development Core Team, 2011) environment for statistical computing with the package "stats". MLR is a statistical technique for finding the linear relation between the independent variables (predictors) and the dependent or response variable (Kumar and Goyal, 2013). The general MLR model is built from N observations of the multiple predictor variables $x_k$ $(k = 1, \ldots, m)$ and the observed target data $y$. The MLR output variable $\hat{y}$ can be written in terms of the input predictor variables as

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 ... + \beta_m x_m, \tag{4.1}$$

where $\beta_j (j = 0, \ldots, m)$ are the regression coefficients or parameters determined by minimizing the MSE between the model output and the target data using a linear least squares algorithm. Stepwise regression is applied here using the R software to choose relevant predictor variables by an automatic procedure (going both forward and backward). The model was trained with two years of data (2009/07-2011/07), while the testing and updating were performed daily by batch learning using the 3-year data

(2011/08-2014/07). Predictors were re-selected and the linear regression was recalculated during each model update.

## 4.3 Online-Sequential Multiple Linear Regression (OS-MLR)

To facilitate the rapid and frequent updating of large number of equations from a linear statistical model, OS-MLR models are developed using the sums-of-squares-and-cross-products matrix (SSCP) (Wilson and Vallée, 2002). The idea of the updating is to do part of the necessary recalculation of regression coefficients in near-real time by updating the SSCP matrix and storing the data in that form rather than as raw observations. The MLR model in (4.1) involves finding the least squares solution of the linear system

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{Y}, \tag{4.2}$$

where the input data matrix $\mathbf{X}$ of dimension $N \times (m+1)$ is

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & \cdots & x_{Nm} \end{bmatrix}, \tag{4.3}$$

and the $\boldsymbol{\beta}$ parameter vector of length $m+1$ and the target data vector $\mathbf{Y}$ of length $N$ are, respectively,

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \tag{4.4}$$

Minimizing $\|\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}\|^2$ leads to the solution

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}, \tag{4.5}$$

where $\hat{\boldsymbol{\beta}}$ is the least squares estimate of the regression coefficients and $\mathbf{K} = \mathbf{X}^T\mathbf{X}$ is the SSCP or data covariance matrix.

First, start with an training set with $N_0$ data points and the solution is given by $\beta^{(0)} = \mathbf{K}_0^{-1}\mathbf{X}_0^T\mathbf{Y_0}$, where $\mathbf{K}_0 = \mathbf{X}_0^T\mathbf{X}_0$.

Next, suppose a new chunk of data containing $N_1$ data points has arrived, updating the model requires minimizing

$$\left\| \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix} \boldsymbol{\beta} - \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \end{bmatrix} \right\|^2, \tag{4.6}$$

yielding the new parameter vector

$$\boldsymbol{\beta}^{(1)} = \mathbf{K}_1^{-1} \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \end{bmatrix}, \text{with } \mathbf{K}_1 = \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix}. \tag{4.7}$$

For sequential learning, $\boldsymbol{\beta}^{(1)}$ would need to be expressed only in terms of $\boldsymbol{\beta}^{(0)}, \mathbf{K}_1, \mathbf{X}_1$ and $\mathbf{Y}_1$. $\mathbf{K}_1$ can be written as

$$\mathbf{K}_1 = \begin{bmatrix} \mathbf{X}_0^T & \mathbf{X}_1^T \end{bmatrix} \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix} = \mathbf{K}_0 + \mathbf{X}_1^T\mathbf{X}_1. \tag{4.8}$$

In (4.7),

$$\begin{aligned} \begin{bmatrix} \mathbf{X}_0 \\ \mathbf{X}_1 \end{bmatrix}^T \begin{bmatrix} \mathbf{Y}_0 \\ \mathbf{Y}_1 \end{bmatrix} &= \mathbf{X}_0^T\mathbf{Y}_0 + \mathbf{X}_1^T\mathbf{Y}_1 \\ &= \mathbf{K}_0\mathbf{K}_0^{-1}\mathbf{X}_0^T\mathbf{Y}_0 + \mathbf{X}_1^T\mathbf{Y}_1 \\ &= \mathbf{K}_0\boldsymbol{\beta}^{(0)} + \mathbf{X}_1^T\mathbf{Y}_1 \\ &= (\mathbf{K}_1 - \mathbf{X}_1^T\mathbf{X}_1)\boldsymbol{\beta}^{(0)} + \mathbf{X}_1^T\mathbf{Y}_1 \\ &= \mathbf{K}_1\boldsymbol{\beta}^{(0)} - \mathbf{X}_1^T\mathbf{X}_1\boldsymbol{\beta}^{(0)} + \mathbf{X}_1^T\mathbf{Y}_1. \end{aligned} \tag{4.9}$$

Substituting (4.9) into (4.7), we get

$$\begin{aligned} \boldsymbol{\beta}^{(1)} &= \mathbf{K}_1^{-1}(\mathbf{K}_1\boldsymbol{\beta}^{(0)} - \mathbf{X}_1^T\mathbf{X}_1\boldsymbol{\beta}^{(0)} + \mathbf{X}_1^T\mathbf{Y}_1) \\ &= \boldsymbol{\beta}^{(0)} + \mathbf{K}_1^{-1}\mathbf{X}_1^T(\mathbf{Y}_1 - \mathbf{X}_1\boldsymbol{\beta}^{(0)}). \end{aligned} \tag{4.10}$$

Generalizing the recursive algorithm for updating, when the $(k+1)$th chunk of new data arrives, the least squares solution can be written as

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} &= \boldsymbol{\beta}^{(k)} + \mathbf{K}_{k+1}^{-1}\mathbf{X}_{k+1}^T(\mathbf{Y}_{k+1} - \mathbf{X}_{k+1}\boldsymbol{\beta}^{(k)}), \\ \mathbf{K}_{k+1} &= \mathbf{K}_k + \mathbf{X}_{k+1}^T\mathbf{X}_{k+1}. \end{aligned} \tag{4.11}$$

Eq.(4.11) gives the recursive formula for $\boldsymbol{\beta}^{(k+1)}$ in OS-MLR. In summary, our initial model was trained by MLR using a 2-year data set (2009/07-2011/07). After that, prediction and model updating were conducted daily by the OS-MLR algorithm using a 3-year data set (2011/08-2014/07).

## 4.4 Multi-layer Perceptron Neural Network (MLP NN)

To construct the MLP NN model (Fig.2.1), the neural network is considered to be a system receiving information from $m$ input variables $x_i$ $(i = 1, \ldots, m)$, namely meteorological, physical and chemical predictors, and produces a single output, in our case the concentration of ozone, $PM_{2.5}$ or $NO_2$. No prior knowledge about the relationship between input and output variables is assumed. The MLP NN forecast model is developed in R software using the "monmlp" package (Cannon, 2012a). The activation function used is the hyperbolic tangent function for the hidden layer and the identity function for output layer. Hence, the MLP NN with $L$ hidden nodes or neurons is mathematically modeled by

$$\hat{y_i} = \sum_{j=1}^{L} \beta_j f(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) + \beta_0, \ (i = 1, ..., N), \tag{4.12}$$

where $f$ is the tanh function, $\mathbf{x}_i$ and $\hat{y_i}$ are the model input and output, respectively, $N$ is the number of data points, $\mathbf{w}_j = [w_{j1}, w_{j2}, ..., w_{jm}]^T$ and $b_j$ are the weights or parameters connecting the input layer to the $j$th hidden node, $\beta_j = [\beta_1, \beta_2, ..., \beta_m]^T$ and $\beta_0$ are the weights/parameters connecting the $j$th hidden node to the output.

Training the MLP NN model involves adjusting the parameters or weights to minimize the objective function $J$, defined here to be the mean squared error (MSE) between the model output and the target data $y_i$:

$$J = \frac{1}{N} \sum_{i=1}^{N} (\hat{y_i} - y_i)^2. \tag{4.13}$$

The minimization of $J$ involves using back-propagation (Hsieh, 2009). A common problem in the development of a neural network is determining the optimal number of hidden nodes (HN). A sequential, small grid search with *bagging* (abbreviated from Bootstrap Aggregating) is used to select the optimal number of HN (Lima et al., 2015).

Bagging (Breiman, 1996) is an ensemble method developed from the idea of bootstrapping in statistics. Under bootstrap resampling, data are randomly selected repeatedly from a dataset with replacement to form a new training dataset, which has the same number of data points as original dataset. A data point in the original dataset can be selected more than once into the new training dataset. During the random draws, predictor and predictand pairs are drawn together. For autocorrelated data, data segments about the length of the autocorrelation time scale are drawn instead of individual data points. In the bagging approach, one model can be built from one bootstrap sampled set, so an ensemble of models can be derived using a large number of bootstrap sets. By averaging the model output from individual members in the ensemble, a final output is obtained. The data not selected in a bootstrap (the "out-of-bag" data) are used as validation data. NN model training is stopped when the model error calculated from the validation data begins to increase to prevent overfitting to noise in the data (Hsieh, 2009).

During the grid search for the optimal number of HN using data from 2009/07 to 2011/07, we used an ensemble of 30 (bagging) models, and sequentially increased the number of HN one by one until the maximum number of HN was achieved or consecutive increments were without improvements, based on the out-of-bag error. Due to the limitation of computational resources and the time-consuming nature of the MLP NN, models were only batch updated seasonally using the 3-year data from 2011/08-2014/07. In each update, 30 bagging ensemble members were run with the same number of HN as found in the initial training.

## 4.5 Extreme Learning Machine (ELM)

The Extreme Learning Machine was proposed by Schmidt et al. (1992) and Huang et al. (2006b) based on single-hidden layer feed-forward neural network (SLFNs) with random weights in the hidden layer. The ELM algorithm implements a SLFN similar in structure to an MLP NN model (Fig 2.1) and mathematically modeled as in (4.12).

Our ELM uses the same activation functions as our MLP NN model, i.e. the hyperbolic tangent function for the hidden layer and the identity function for the output layer. Huang et al. (2006b) proved that the $\mathbf{w}_i$ and $b_i$ parameter in (4.12) can be randomly assigned if the activation function is infinitely differentiable, so only the $\boldsymbol{\beta}$

parameters need to be optimized when minimizing the mean squared error between the model output $\hat{y}$ and the target data $y$. Thus, in the ELM approach, training an SLFN is equivalent to simply finding the least-squares solution of the linear system

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y}, \tag{4.14}$$

where the hidden layer output matrix $\mathbf{H}$ of dimension $N \times (L+1)$ is

$$\mathbf{H} = \begin{bmatrix} 1 & f(\mathbf{w_1} \cdot \mathbf{x}_1 + b_1) & \cdots & f(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & f(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}, \tag{4.15}$$

and the $\boldsymbol{\beta}$ parameter vector of length $L+1$ and the target data vector $\mathbf{Y}$ of length $N$ are

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_L \end{bmatrix} \text{ and } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}. \tag{4.16}$$

Eqs.(4.14) and (4.16) are mathematically identical to the MLR Eqs.(4.2) and (4.4), hence ELM has transformed an MLP NN model requiring complicated nonlinear optimization to a simple MLR problem. The solution of the linear system for $\boldsymbol{\beta}$ is simply via least squares as in MLR, i.e.

$$\hat{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{Y} \tag{4.17}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose pseudo-inverse (Liang et al., 2006).

We consider the case where rank $\mathbf{H} = L$, the number of hidden nodes, then $\mathbf{H}^{\dagger}$ is given by

$$\mathbf{H}^{\dagger} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T, \tag{4.18}$$

and

$$\hat{\boldsymbol{\beta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{Y}. \tag{4.19}$$

Huang et al. (2006b) did not have the bias parameter $\beta_0$ in the output layer as leaving out $\beta_0$ might make learning more difficult (Thimm and Fiesler, 1997), we in-

cluded it by having a first column of ones in $\mathbf{H}$ and having $\beta_0$ in the top row of the $\boldsymbol{\beta}$ matrix (Romero and Alquẽzar, 2012). Huang and Wang (2006) set the number of hidden neurons empirically and used a uniform random distribution in the range [-1,1] for both the weights and bias parameters in the hidden layer. In order to perform the non-linear relationship of air quality problem efficiently, we need to find a more accurate and automatic way to choose the parameters and the net work structure (i.e. the number of HN).

The hill climbing method was used to decide the optimal number of HN in ELM. Hill climbing is a simple mathematical optimization technique that starts with an arbitrary solution to a problem, and then attempts to find a better solution (smaller MSE) by incrementally changing a single element (number of hidden nodes). If the change produces a better solution, an incremental change (often by a defined step size) is made towards the new solution. Iterations continue until no further improvements can be found. In the hill climbing algorithm used here, the step size is automatically adjusted by the algorithm. Thus it shrinks when the probes do poorly and it grows when the probes do well, helping the algorithm to be more efficient and robust (Yuret, 1994). For each candidate a 10-fold cross-validation within the training set was performed to avoid over-fitting. Cross-validation is a model validation technique for assessing how the results of a statistical analysis will generalize to an independent data set. In $k$-fold cross-validation, the original sample is randomly partitioned into $k$ equal sized subsamples, with a single subsample retained as the validation data for testing the model, and the remaining $k-1$ subsamples as training data. The process is repeated $k$ times so all $k$ subsamples are used as validation data.

As weights are randomly assigned in ELM, diversity is an important factor because model complexity is affected by the variance of the distribution. Typically a uniform distribution, spread over a fixed interval $[-r, r]$ is used for weight distribution. Thimm and Fiesler (1997) gave a review of random weight initialization methods for MLP and found that $r$ should be of the form

$$r = aF^{-0.5} \tag{4.20}$$

where F is the number of predictors in the case of a 1-hidden layer MLP model. Using the hyperbolic tangent as the activation function, they found $a \approx 1$ to be a reasonable value. For our ELM, $a = 1$ was chosen, and the random bias parameter $b_j$ in the hidden

layer was chosen to be uniformly distributed within [-1, 1].

The ELM algorithm was only used to do the initial model training with two-year data sets (2009/07-2011/07). Model updating was then conducted by the OS-ELM algorithm. Like many other learning algorithms the stochastic nature of ELM means that different trials of simulation may yield different results. The random assignment of weight and bias parameters in the hidden layer makes each ELM distinct. To make the ELM model more stable, we use an ensemble of 30 members in the ELM models and the output of the ensemble is the average of the individual ensemble members.

## 4.6 Online-Sequential Extreme Learning Machine (OS-ELM)

As ELM randomly chooses weights for the hidden layer and analytically determines weights in the output layer by linear least squares, the ELM algorithm can be adapted for online sequential learning in the same way as the linear regression model in Sec 4.3.

Given $N_0$ observations in the initial training set with $N_0 \geq L$, the number of HN, if we use batch ELM to train the model, the matrix $\|\mathbf{H}_0\beta - \mathbf{Y}_0\|^2$ is minimized and the solution (4.19) gives $\boldsymbol{\beta}^{(0)} = \mathbf{K}_0^{-1}\mathbf{H}_0^T\mathbf{Y_0}$, where $\mathbf{K}_0 = \mathbf{H}_0^T\mathbf{H}_0$.

Analogous to the online sequential solution (4.11) for OS-MLR, when the $(k+1)$th chunk of new data arrives, the recursive least-squares solution for updating OS-ELM is

$$\begin{aligned} \boldsymbol{\beta}^{(k+1)} &= \boldsymbol{\beta}^{(k)} + \mathbf{K}_{k+1}^{-1}\mathbf{H}_{k+1}^T(\mathbf{Y}_{k+1} - \mathbf{H}_{k+1}\boldsymbol{\beta}^{(k)}), \\ \mathbf{K}_{k+1} &= \mathbf{K}_0 + \mathbf{H}_{k+1}^T\mathbf{H}_{k+1}. \end{aligned} \tag{4.21}$$

To summarize, the OS-ELM approach consists of two parts: an initial training phase and a sequential learning phase. The initialization phase involves batch learning with ELM on the initial training data and analytically solves (4.19). Following the initialization phase, the model is updated in the online sequential learning phase by (4.21) using the newly arrive chunks of data. Once a chunk of data has been used, it can be discarded as it is not used in future model updates (Liang et al., 2006). For our case, the initial model training was conducted by the ELM algorithm using two years

of data (2009/07-2011/07) with 30 ensemble members. Models are then update daily by the OS-ELM algorithm using the 3-year data set from 2011/08 to 2014/07. The number of hidden node was chosen in the initial learning phase using the hill climbing method and was not changed anymore during the online-sequential learning phase.

## 4.7 Model Evaluation

Several statistical scores were used to evaluate the performance of $O_3$, $NO_2$ and $PM_{2.5}$ model, including the Pearson correlation coefficient ($r$), mean absolute error (MAE), MAE/MAD (MAD being the mean absolute deviation), root mean squared error (RMSE) and skill score (SS).

### 4.7.1 Pearson Correlation Coefficient ($r$)

The Pearson correlation coefficient, reflecting the degree of linear relationship between two variables, is defined by

$$r = \frac{\mathrm{cov}(\hat{\mathbf{Y}}, \mathbf{Y})}{\sigma_{\hat{\mathbf{Y}}}\sigma_{\mathbf{Y}}}, \tag{4.22}$$

where $\hat{\mathbf{Y}}$ demotes the model predicted pollutant concentrations, $\mathbf{Y}$ the observed values, cov the covariance and $\sigma$ the standard deviation. This coefficient varies from -1 to 1, with 0 indicating no relationship. While the Pearson correlation is a good measure of the linear association between predictions and observations, it does not take into account the prediction bias, and is sensitive to rare extreme events.

### 4.7.2 Mean Absolute Error (MAE)

The mean absolute error (MAE) is the average absolute value of the forecast errors, with

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i|, \tag{4.23}$$

where $N$ is the number of data points, $y_i$ is the observed value and $\hat{y}_i$ is the predicted value.

### 4.7.3 MAE/MAD

The average air pollutant concentrations can vary from one location to another. Model predictions for areas with large variations in pollutant concentration levels usually have higher MAE than those for areas with smaller variations. Therefore, the MAE is not always useful for comparing model results from different locations. To normalize the errors, MAE is divided by the mean absolute deviation (MAD) of the observations, yielding a relative mean absolute error that allows comparison between different locations.

$$\frac{\text{MAE}}{\text{MAD}} = \frac{\sum_{i=1}^{N} |\hat{y}_i - y_i|}{\sum_{i=1}^{N} |y_i - \overline{y}|}, \tag{4.24}$$

where $\overline{y}$ is the mean of $y$.

### 4.7.4 Root Mean Square Error (RMSE)

The root mean squared error (RMSE) is the square root of the mean squared error between the predictions and observations,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}, \tag{4.25}$$

RMSE is more sensitive to outliers than the MAE.

### 4.7.5 Skill Score (SS)

The skill score is used to determine the skill of a forecast model by comparing it to a base or reference model such as climatology or persistence. In this study the skill score is given by

$$\text{SS} = \frac{(A - A_{\text{ref}})}{(A_{\text{perfect}} - A_{\text{ref}})} \tag{4.26}$$

where $A$ represents the MAE from using the MLR, OS-MLR, MLP NN or OS-ELM model, $A_{\text{ref}}$ = MAE of the UMOS reference model and $A_{\text{perfect}} = 0$ for the MAE of a perfect model. A model with SS $> 0$ means that it performs better than the reference model, whereas SS $< 0$ means it has less skill than the reference model. The skill score

makes it easy to compare the forecast skill of different models.

## 4.8 Summary

In summary, stepwise multiple linear regression (MLR), online-sequential multiple linear regression (OSMLR), multi-layer perceptron neural network (MLP NN) and online-sequential extreme learning machine (OSELM) were developed and updated using a total of five years of air quality data. Model performance was evaluated by several statistical scores and compared with those of the UMOS model. These five models compared include non-linear models (MLP NN, OSELM) and linear models (MLR, OSMLR, UMOS). They included two different approach to model updating, namely batch learning (MLP NN, MLR) and online-sequential learning (OSMLR, OSELM, UMOS).

# Chapter 5

# Results and Discussion

Model performance on testing data was evaluated by comparing the five model estimates (UMOS, MLR, OSELM, OSMLR and MLP NN) with the near-real-time observed ozone concentrations. A comparison between our four models and UMOS would not be completely fair due to differences in model strategy and some fundamental deficiencies in the UMOS implementation. UMOS uses separate models for the warm and cold seasons, whereas our models were developed for the whole year. Furthermore, the integer precision of the UMOS output and the UMOS forecasts made under missing data conditions likely lowered the UMOS forecast scores. Nevertheless, we have included UMOS in our model comparisons below.

## 5.1 Ozone

The boxplot (Fig 5.1) shows the observed and predicted values at the six stations. The boxplot is a convenient way of displaying the distribution of data based on the following statistics: median, first quartile, third quartile, minimum and maximum. In the boxplot, the central rectangle spans the first quartile to the third quartile, while the waistline inside the rectangle shows the median. Distance between the first and third quartiles is the interquartile range (IQR). The upper whisker extends to the hight value within 1.5 IQR from the top of the rectangle, while the lower whisker extends to the lowest value within 1.5 IQR from the bottom of the rectangle. Values beyond the end of the whiskers are considered outliers and are shown as dots. Among the six stations, Winnipeg has highest average ozone value and extreme cases, although the extreme values have not been vetted sufficiently to know if they are valid observations, i.e. they could be related to wildfire events upwind or local smoke issues. Fig 5.1 also indicates the forecast models tended to underestimate the median ozone concentration, and the extremes in Winnipeg.

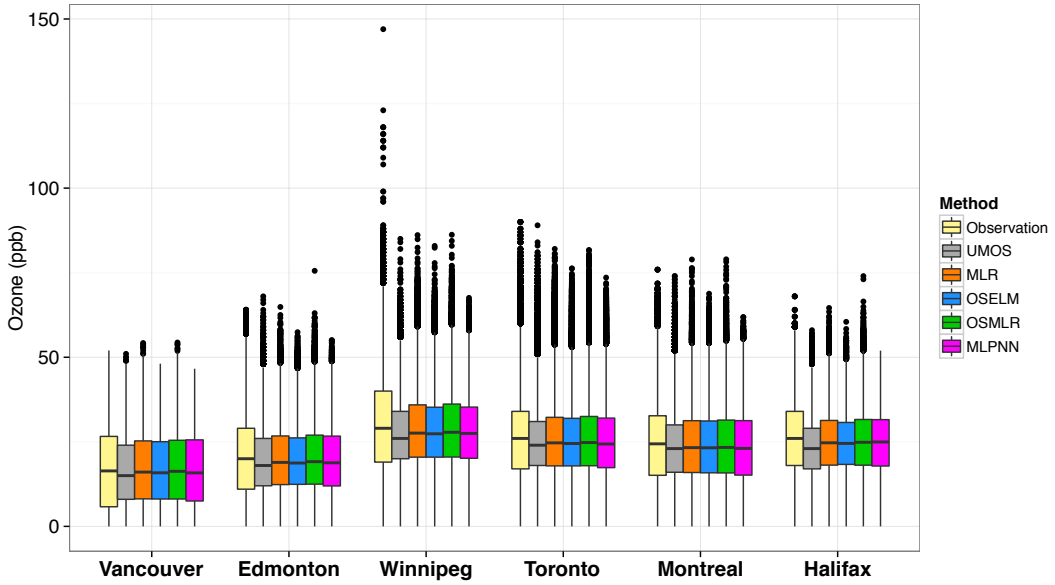



Figure 5.1: Boxplot of the observed ozone values and the predicted values from five methods over all forecast lead times (1 - 48hr) at six stations.

Ozone model forecast error statistic are shown in Fig 5.2. The Pearson correlation ($r$) for the five models range from 0.75 to 0.85. The MAE for the models range from 4.8 ppb to 7.3 ppb and RMSE range from 6.4 ppb to 9.8 ppb. The normalized error MAE/MAD varies from 0.47 to 0.64.

According to the correlation and normalized error, models have best performance in Vancouver with smallest error and highest correlation attained by the OSELM method. Due to the extreme values in Winnipeg (Fig 5.1), all forecast models performed poorer than other stations in terms of MAE and RMSE. All four models had better performance than UMOS, the benchmark, and OSELM generally outperformed the other methods over the six stations. Both OSMLR and stepwise MLR are daily updated linear methods and both performed well in ozone forecasting, with these two linear regression methods showing best skill in Winnipeg. The seasonally updated MLP NN model tended to underperform when compared with MLR, OSELM and OSMLR, as the other models were updated daily and the MLP NN only 3-monthly due to high computational cost - updating the MLP NN seasonally used more than 10 times the cpu time of the OSELM updated daily.

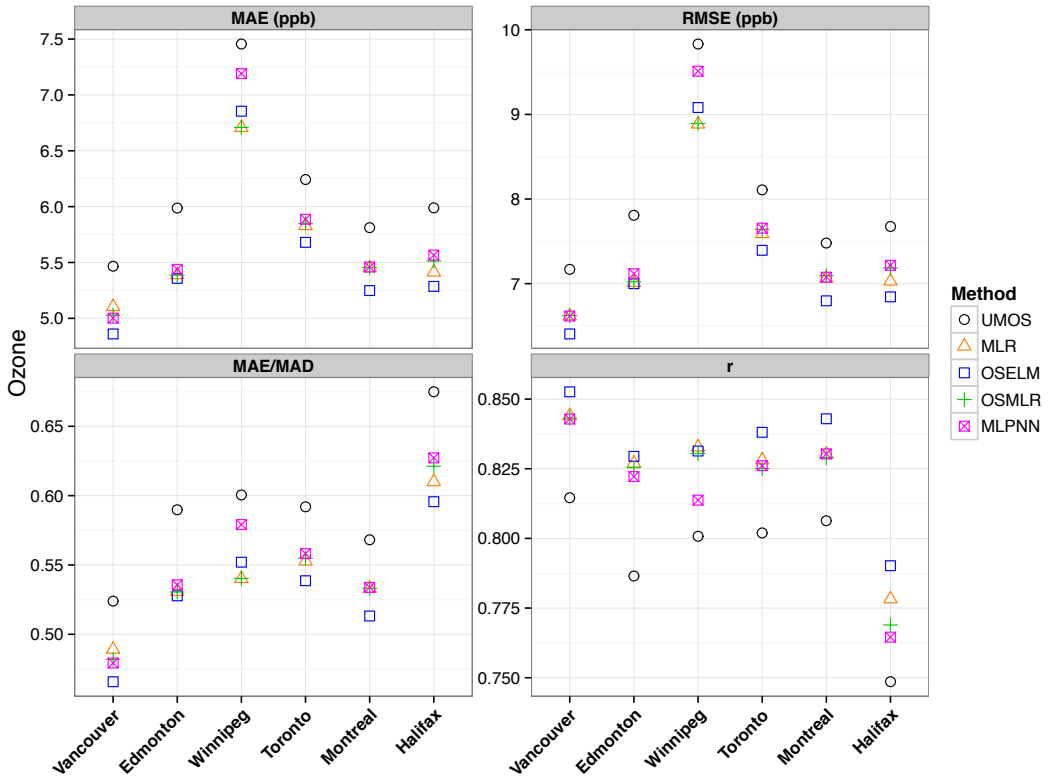


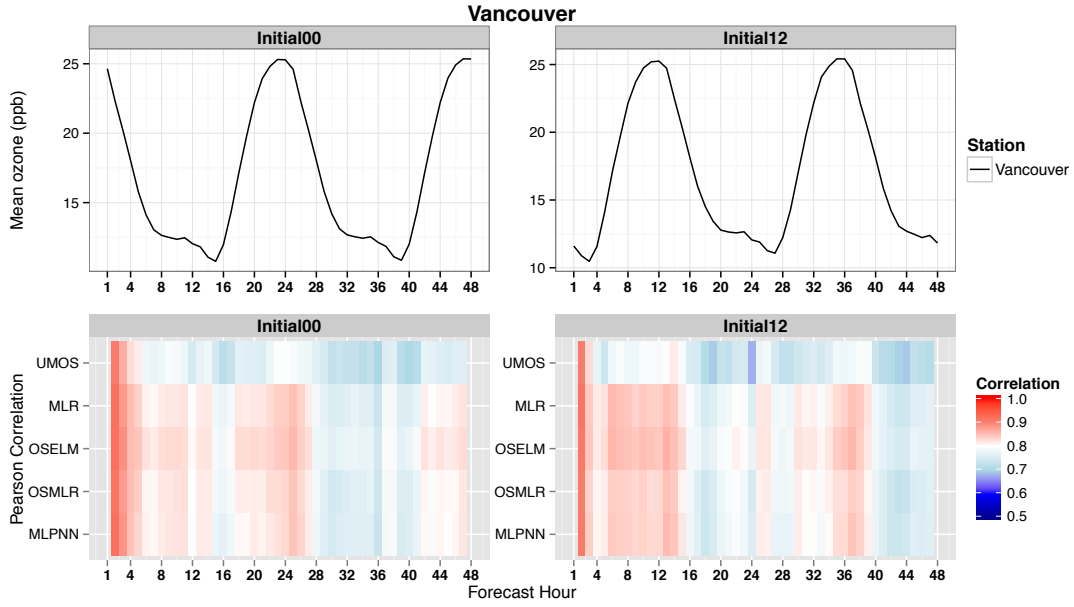

Figure 5.2: Ozone forecast scores of different methods averaged over all forecast lead times (1 - 48hr) at the six stations.

Forecast correlation scores from the five models are shown by heat maps in Fig 5.3 as the forecast lead time varies from 1 - 48 hrs, with the forecast initiated at 00 UTC and 12 UTC. The heat maps show high correlation in red and low correlation in blue, with some missing values in Edmonton and Winnipeg. From the mean ozone concentration plots, strong diurnal cycle can be found in all six stations (top panels in Fig 5.3). Station Vancouver, which has the lowest mean ozone concentration among the six stations, has the peak during late afternoon (local time 4 pm - 6 pm). Forecast models have better correlation scores during the peak time and the 00 UTC initiated models also work well from 1 to 12 forecast hours. UMOS lost to other methods in Vancouver, whereas the seasonally updated MLP NN method has competitive performance against the daily updated linear methods.

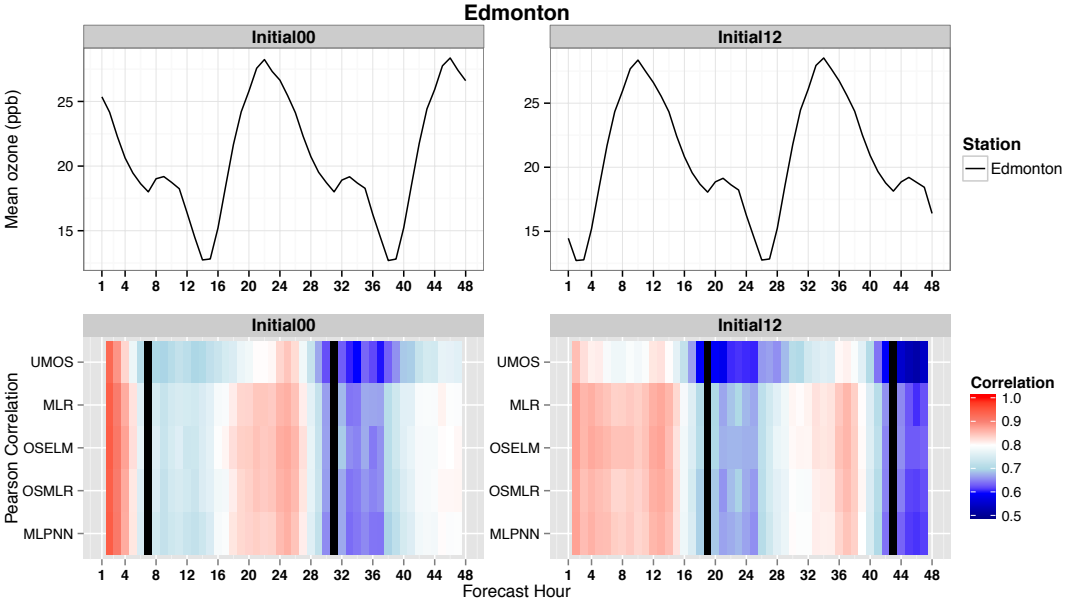
Edmonton has average ozone concentration above 25 ppb in local late afternoon. Correlations above 0.8 occur in 1-4 and 16-27 hour forecast lead times of the 00 UTC initiated models and 1-16 and 30-38 hour lead times of the 12 UTC initiated models. However, models show poor performance when ozone is low at night (12 am - 6 am) relative to other forecast hours as well as other stations. Winnipeg has highest ozone concentration among the six stations, with over 35 ppb mean value during peak time. Models have good performance for first 24 hours while initiated at 00 UTC but only works well in first 12 forecast hours of the 12 UTC initiated models.

In station Toronto, correlation scores of the 2-3, 16-26 and 42-48 hour lead time forecasts initiated at 00 UTC are above 0.8, while for models initiated at 12 UTC, better performance occurs in the 1-14 and 30-37 hour lead times. These forecast lead times all correspond to the high ozone concentration period in the diurnal cycle. We can conclude that models initiated at 00 UTC and 12 UTC both work well to predict ozone concentration in local time (LST) 12 pm - 9 pm and the 12 UTC initiated model also has good behavior during morning. Ozone in Montreal has average concentration over 30 ppb in late afternoon. Models initiated at 00 UTC have higher correlation scores during 1-10, 19-24 and 44-46 hour lead times, while the 12 UTC initiated models do well for 1-12, 32-35 hour lead times, indicating good model skills during daytime (9 am - 8 pm LST).

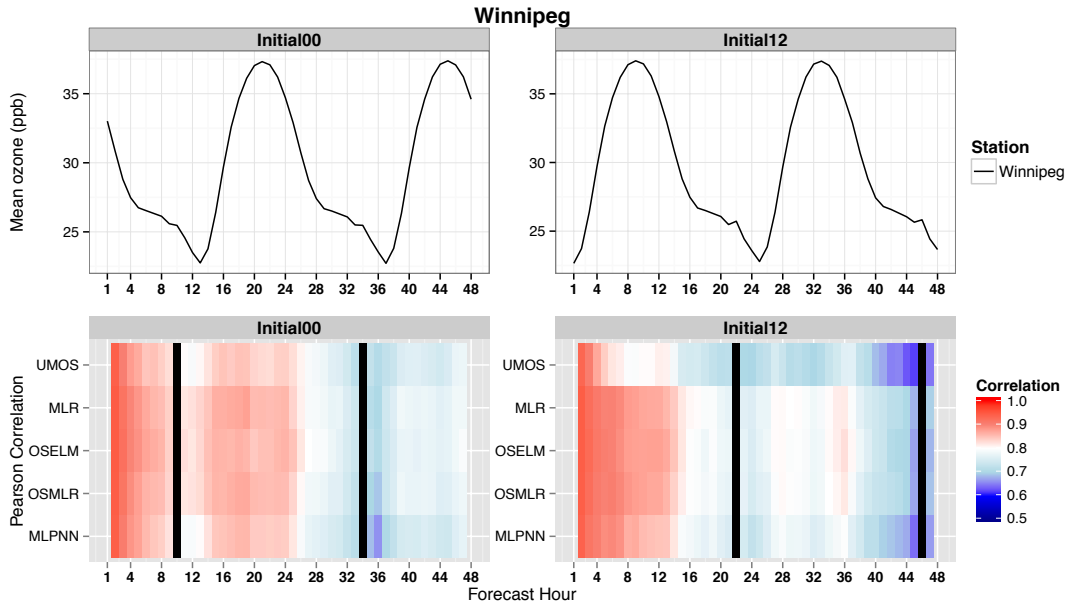
While comparing different models, OSELM tends to outperform the others by making accurate predictions over a wider range of forecast hours. Comparing stations, models generally have lowest accuracy in Halifax.


Vancouver
Initial00
Initial12
Mean ozone (ppb)
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Forecast Hour
Station
Vancouver
Correlation

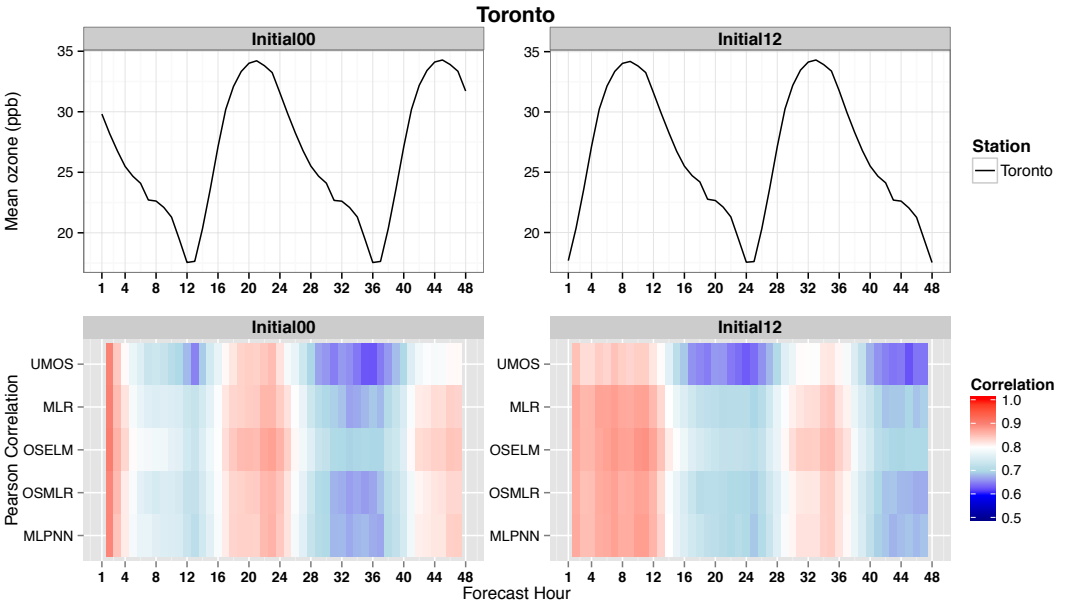
(a)


Edmonton
Initial00
Initial12
Mean ozone (ppb)
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Forecast Hour
Station
Edmonton
Correlation

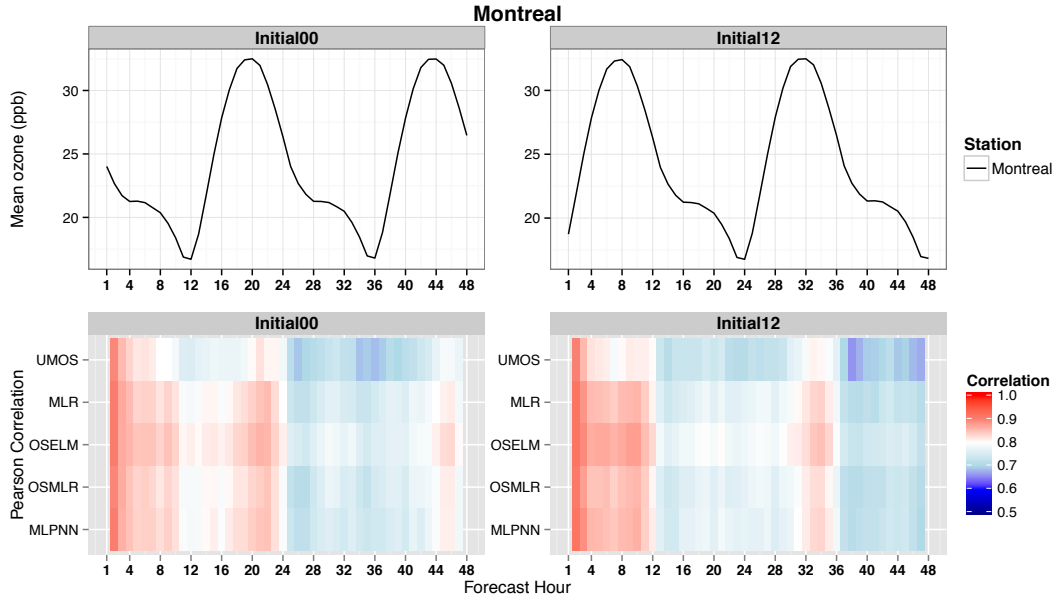
(b)


Winnipeg
Initial00
Initial12
Mean ozone (ppb)
35
30
25
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Winnipeg
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.9
0.8
0.7
0.6
0.5
Forecast Hour

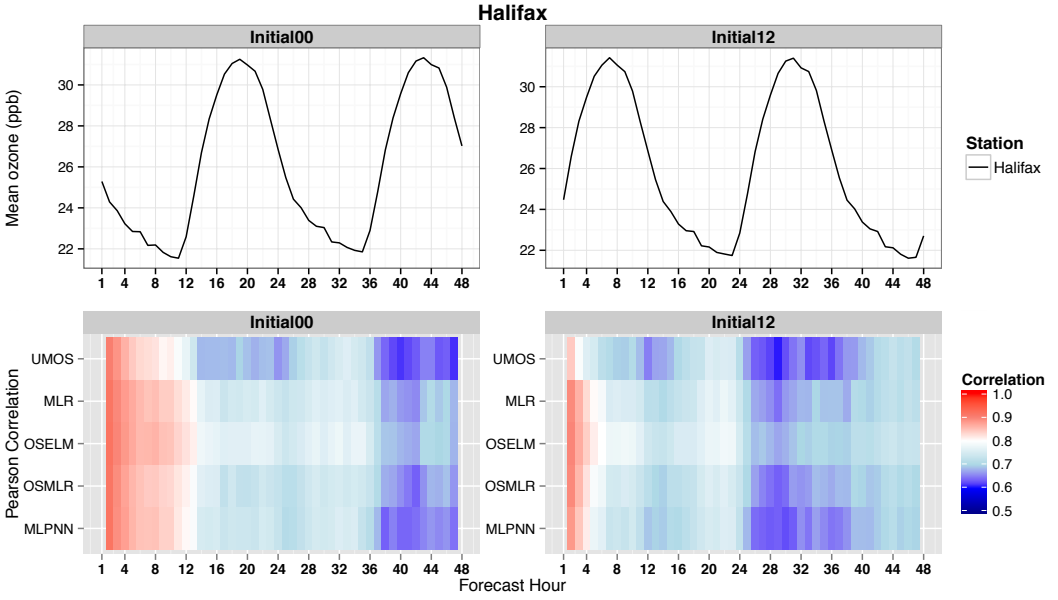
(c)


Toronto
Initial00
Initial12
Mean ozone (ppb)
35
30
25
20
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Toronto
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.9
0.8
0.7
0.6
0.5
Forecast Hour

(d)




(e)




(f)

Figure 5.3: Forecast correlation score as a function of the forecast lead time (1-48hr) from the five models displayed in a heat map (bottom panels) for forecasts initiated at 00 UTC (left) and 12 UTC (right) at six stations. Black vertical stripes indicate "missing values", i.e. fewer than 100 data points were available for model training during 2009/07-2011/07. The mean diurnal ozone cycle is displayed in the top panels. 00 UTC corresponds to local time (daylight saving time) of 5 pm, 6 pm, 7 pm, 8 pm, 8 pm and 9 pm at Vancouver, Edmonton, Winnipeg, Toronto, Montreal and Halifax, respectively.

To evaluate the performance of models, forecast skill scores were calculated relative to UMOS reference model based on the MAE. Fig 5.4 shows the skill scores for different stations and forecast lead times, with scores above zero indicating that the models have smaller MAE than UMOS. All four methods (stepwise MLR, OSELM, OSMLR and MLP NN) have positive skill scores for most forecast lead times at all six stations, especially during the high ozone concentration period in the diurnal cycle. OSELM shows the best performance in stations Vancouver, Edmonton, Toronto, Montreal and Halifax, often outperforming UMOS by more than 10%. Linear methods (MLR and OSMLR) also do well in ozone forecasting, showing higher skill score than OSELM in Winnipeg. The seasonally updating MLP NN method slightly underperformed the other three methods and is the only method underperforming UMOS in Winnipeg when initiated at 00 UTC.

To compare model performance in different seasons, prediction and observational data during the testing period (2011/8-2014/7) are broken into a warm season (April, May, June, July, August, September) and a cold season (October, November, December, January, February, March). Table 5.1 shows the mean, standard deviation and maximum value by station and season. It indicates the higher mean and maximum values during warm season, but some extreme events can contribute to maximum value during cold season (e.g. Winnipeg). The ozone residuals (prediction−observation) (Fig 5.5) show a tendency to have a negative median value (i.e. the models have a tendency to underpredict) during the warm season. The UMOS model was developed separately for the warm season and the cold season, whereas MLR, OSELM, OSMLR and MLP NN were developed using two entire years of data, which could be the reason that the latter four models underestimated the ozone concentration during the warm season. For forecast scores in Fig 5.6, although MAE and RMSE are higher during the warm season, models tend to have better performance in the warm season in terms of MAE/MAD and $r$ for all stations except Vancouver. Comparing different models, OSELM tends to have better scores than the other four models in Vancouver, Toronto, Montreal and Halifax for both seasons.

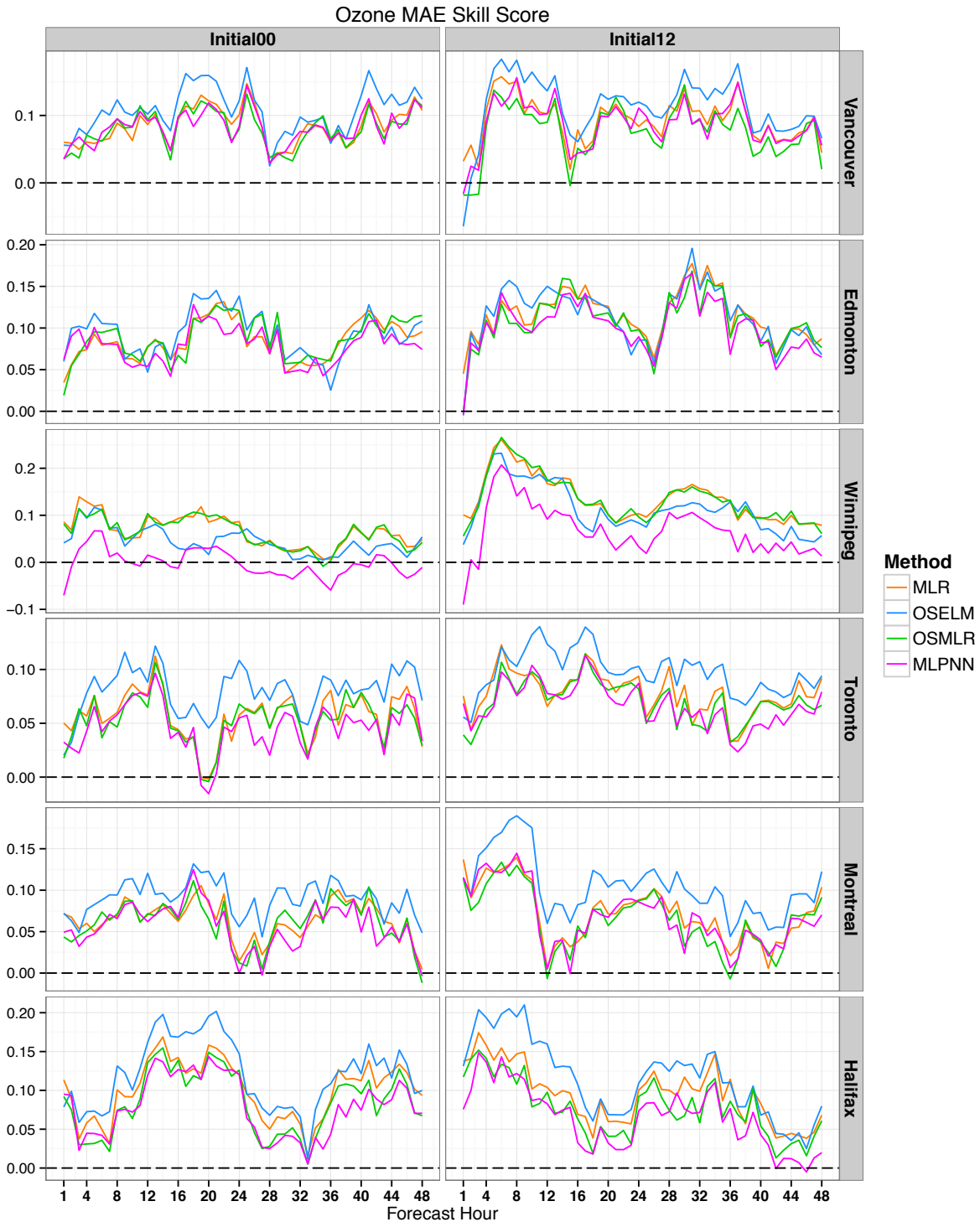



Figure 5.4: Ozone MAE skill score of different models by forecast hour, with forecasts initiated at 00 UTC (left column) and 12 UTC (right column). The panels are arranged in six rows, from Vancouver (top) to Halifax (bottom).

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 19.9 | 11.5 | 52.0 | 13.9 | 12.2 | 47.6 |
| Edmonton | 25.3 | 12.3 | 64.0 | 15.9 | 10.4 | 63.0 |
| Winnipeg | 33.9 | 17.0 | 97.0 | 26.0 | 13.4 | 147.0 |
| Toronto | 31.2 | 14.4 | 90.0 | 21.4 | 10.2 | 72.0 |
| Montreal | 27.3 | 13.0 | 75.9 | 20.9 | 11.2 | 56.5 |
| Halifax | 25.2 | 11.6 | 68.0 | 27.0 | 9.8 | 58.0 |

Table 5.1: Statistical properties of ozone concentration (ppb) by station and season.

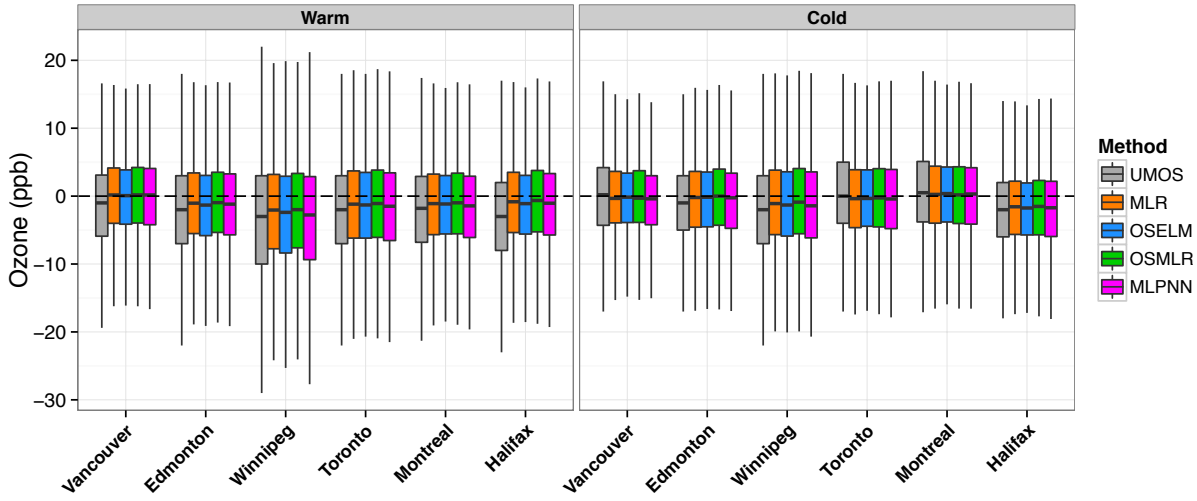



Figure 5.5: Boxplot of the ozone residuals (prediction−observation) by season and station. Outliers are not plotted but can be seen in Fig 5.7.

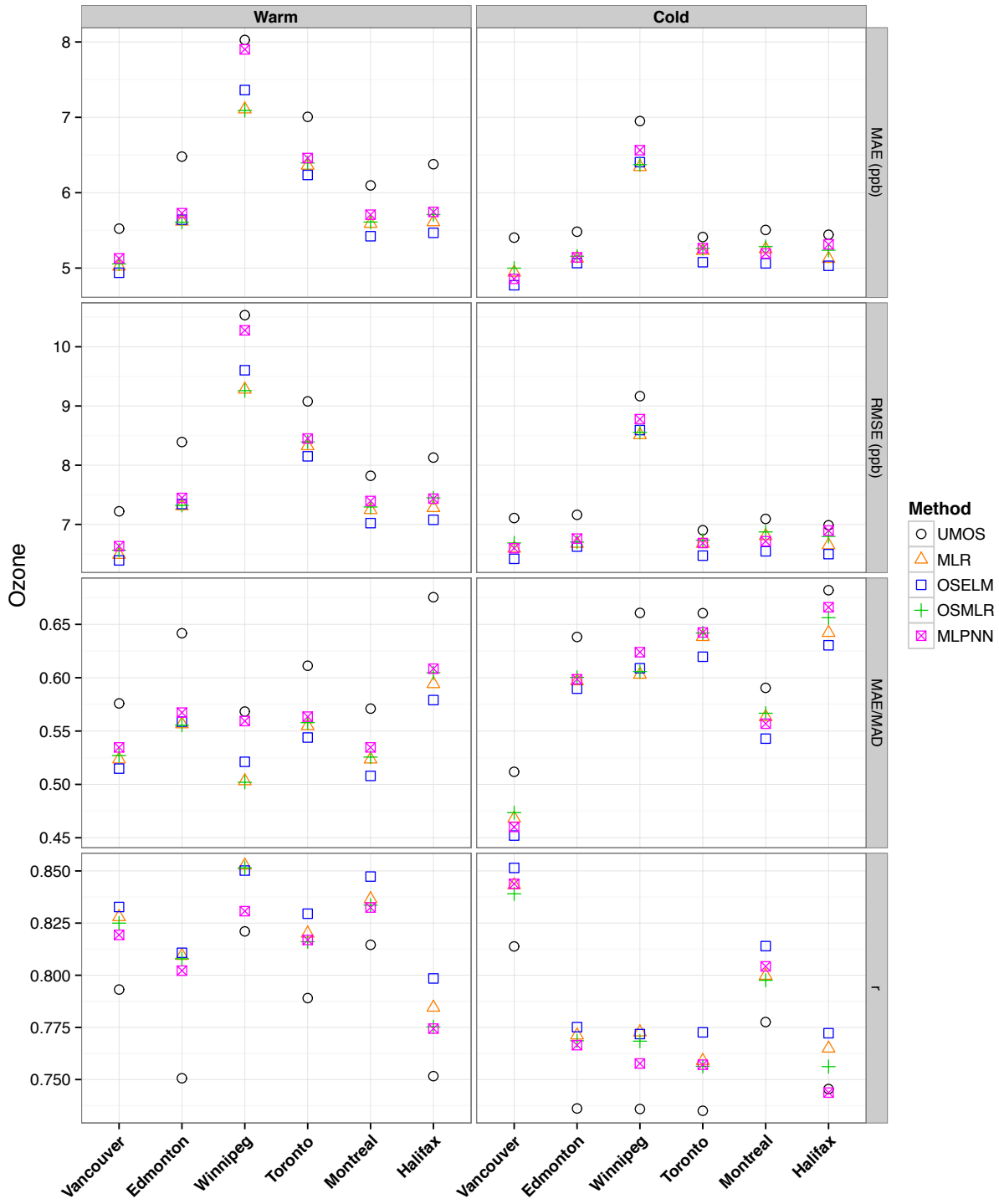



Figure 5.6: Ozone forecast scores (MAE, RMSE, MAE/MAD and $r$) by season and station.

To study how well the models are in forecasting extreme events of ozone concentration, the top $10^{th}$ percentile ozone observational data and the corresponding model predictions are extracted for the two season during 2011/08-2014/07. Table 5.2 provides statistical properties of the top $10^{th}$ percentile ozone concentration observations by season and station, which shows the mean value over 40 ppb during the warm season and over 35 ppb for the cold season. Winnipeg has highest extreme ozone concentration with 64 ppb average during warm season and 147 ppb maximum during the cold season over three years. Fig 5.7 presents the ozone residuals for top $10^{th}$ percentile. As the models produced 48 hours of forecasts, a high ozone day can generate multiple outlier points in the boxplot. The models all have a negative median indicating underprediction of the extreme values, slightly less serious in the cold season than the warm season. The forecast scores for the top $10^{th}$ percentile (Fig 5.8) when compared with the forecast scores for all data (Fig 5.6 or Fig 5.2) revealed that the linear models (OSMLR and MLR) tended to improve relative to the nonlinear models (OSELM and MLP NN) when considering only the top $10^{th}$ percentile, i.e. the linear models tended to perform better than the nonlinear models when forecasting on high ozone days.

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 40.2 | 3.3 | 52.0 | 37.0 | 3.2 | 47.6 |
| Edmonton | 46.6 | 4.3 | 64.0 | 35.3 | 5.3 | 63.0 |
| Winnipeg | 64.0 | 6.3 | 97.0 | 48.8 | 6.8 | 147.0 |
| Toronto | 58.1 | 7.3 | 90.0 | 38.8 | 4.8 | 72.0 |
| Montreal | 50.1 | 5.2 | 75.9 | 39.3 | 3.8 | 56.5 |
| Halifax | 45.0 | 3.9 | 68.0 | 42.5 | 2.5 | 58.0 |

Table 5.2: Statistical properties of top $10^{th}$ percentile ozone concentration by station and season.

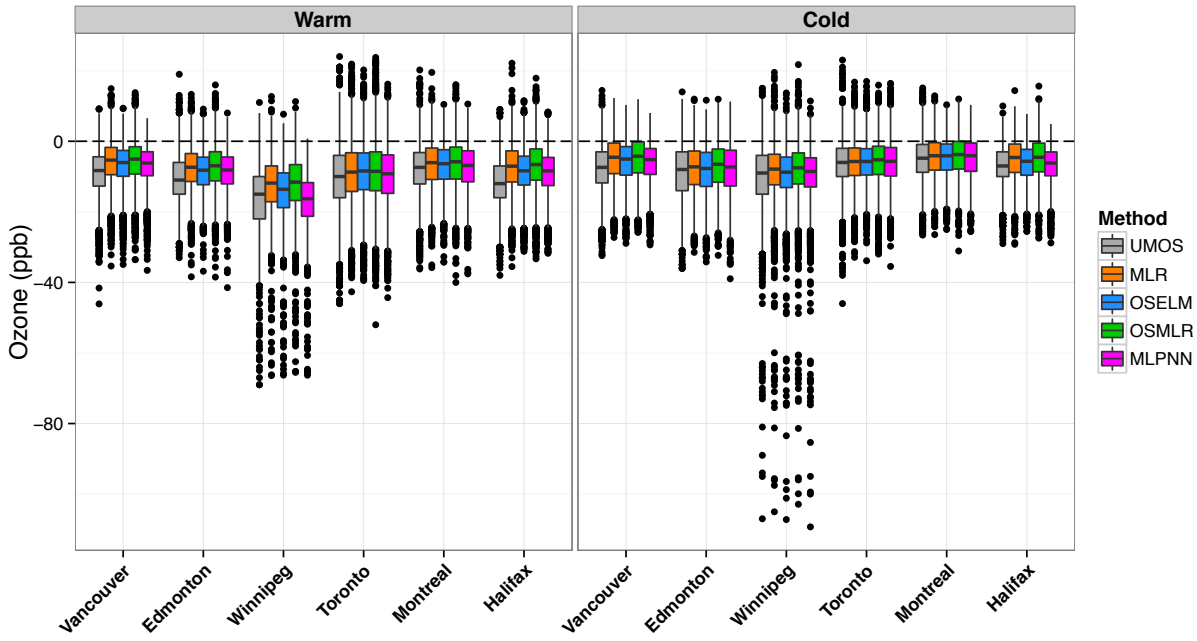



Figure 5.7: Boxplot of ozone residuals (prediction−observation) (over all leadtimes) from the top $10^{th}$ percentile by season and station.

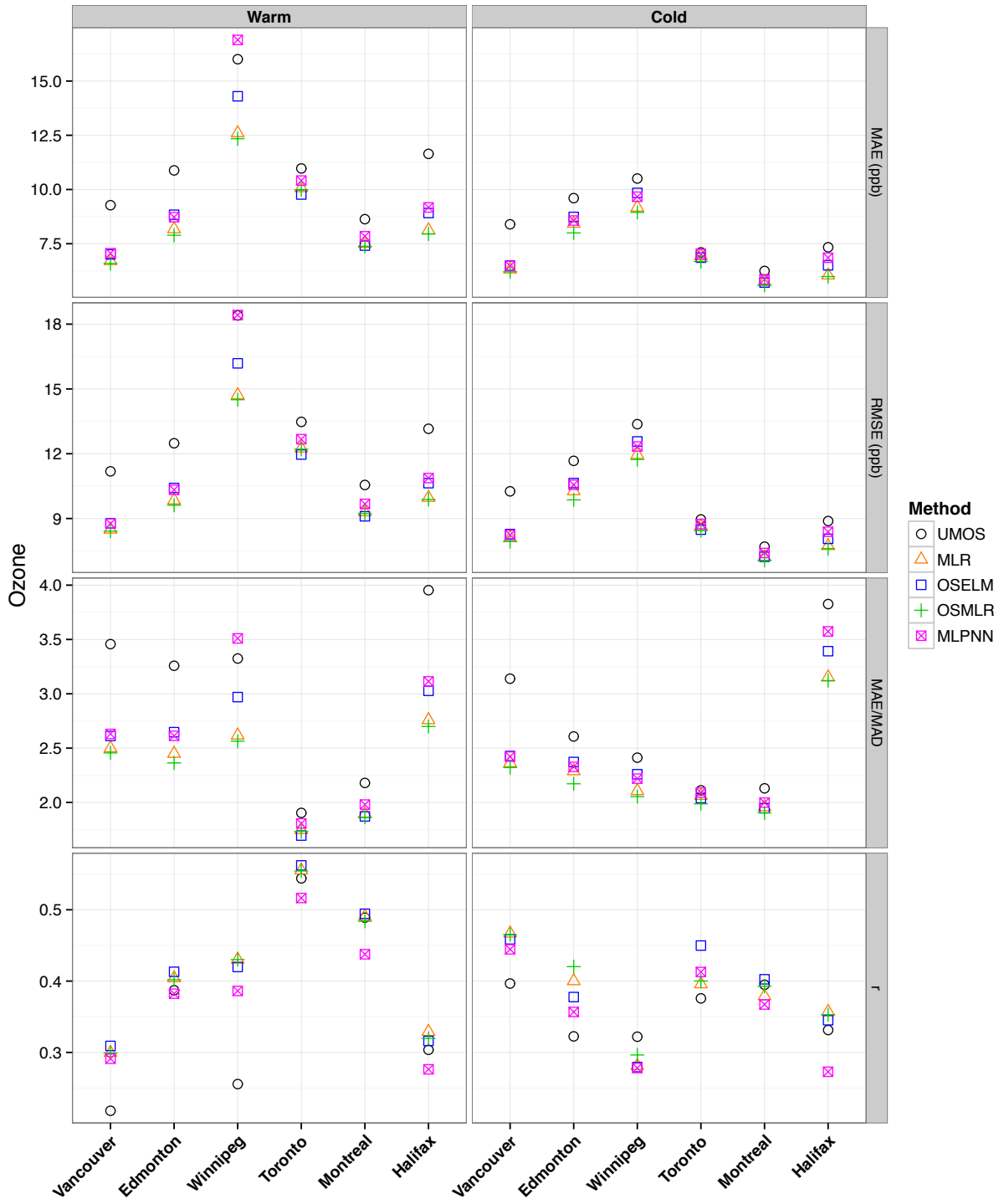



Figure 5.8: Ozone forecast scores of top $10^{th}$ percentile by season and station.

## 5.2 $PM_{2.5}$

$PM_{2.5}$ observations and predictions by station are shown in Fig 5.9. Edmonton and Montreal have the highest median $PM_{2.5}$ concentration (over 8 $\mu g/m^3$) among the six stations while Vancouver has the lowest median (5 $\mu g/m^3$). From the boxplot, in terms of the median, UMOS tends to agree with the observed $PM_{2.5}$ values better than the other methods, which tend to over-predict in Edmonton, Winnipeg, Montreal and Halifax. Outliers in Edmonton, Winnipeg and Montreal are more common than in the other three stations, indicating more extreme $PM_{2.5}$ values.

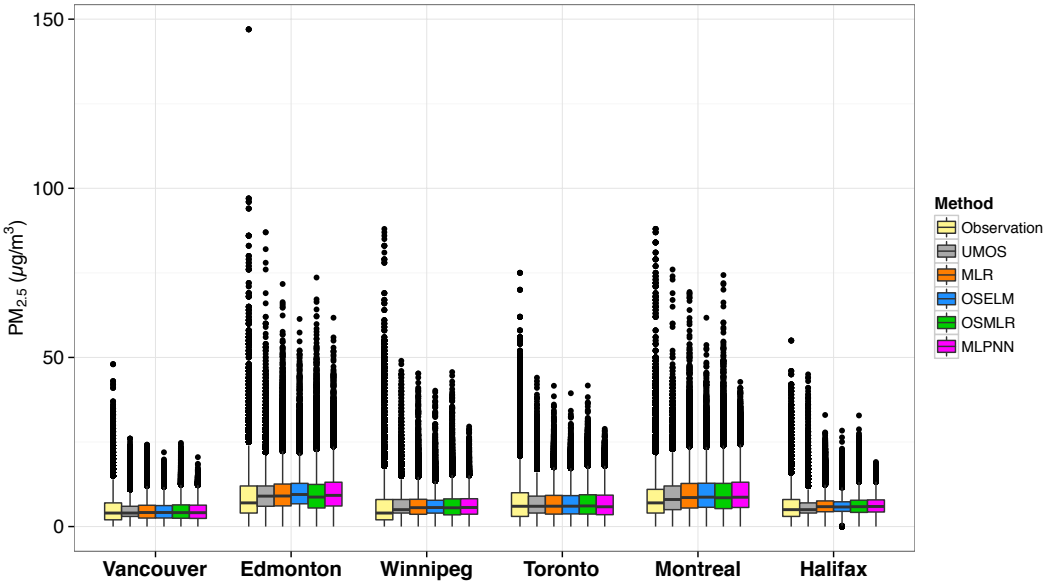



Figure 5.9: Boxplot of the $PM_{2.5}$ observations and predictions by different methods at six stations.

Fig 5.10 illustrates the $PM_{2.5}$ model forecast scores during the testing period by station. Models have poorer scores (MAE/MAD and $r$) for $PM_{2.5}$ than for ozone (Fig 5.2). All models have similar MAE and RMSE for $PM_{2.5}$. The correlation score ($r$) ranges from 0.4 to 0.7, with higher $r$ found in Vancouver, Toronto and Montreal, and lower $r$ in Winnipeg and Halifax. Comparing different methods, from the relative error (MAE/MAD) plot we found that UMOS had the lowest MAE/MAD in Edmonton, Winnipeg and Montreal. In Fig 5.9, these three stations have more outliers, and MLR, OSELM, OSMLR and MLP NN are all over predicting, as their median values are above

the observed median. OSELM slightly outperformed the other methods in Vancouver, Toronto and Halifax, while OSMLR tended to have the highest relative error in most stations.

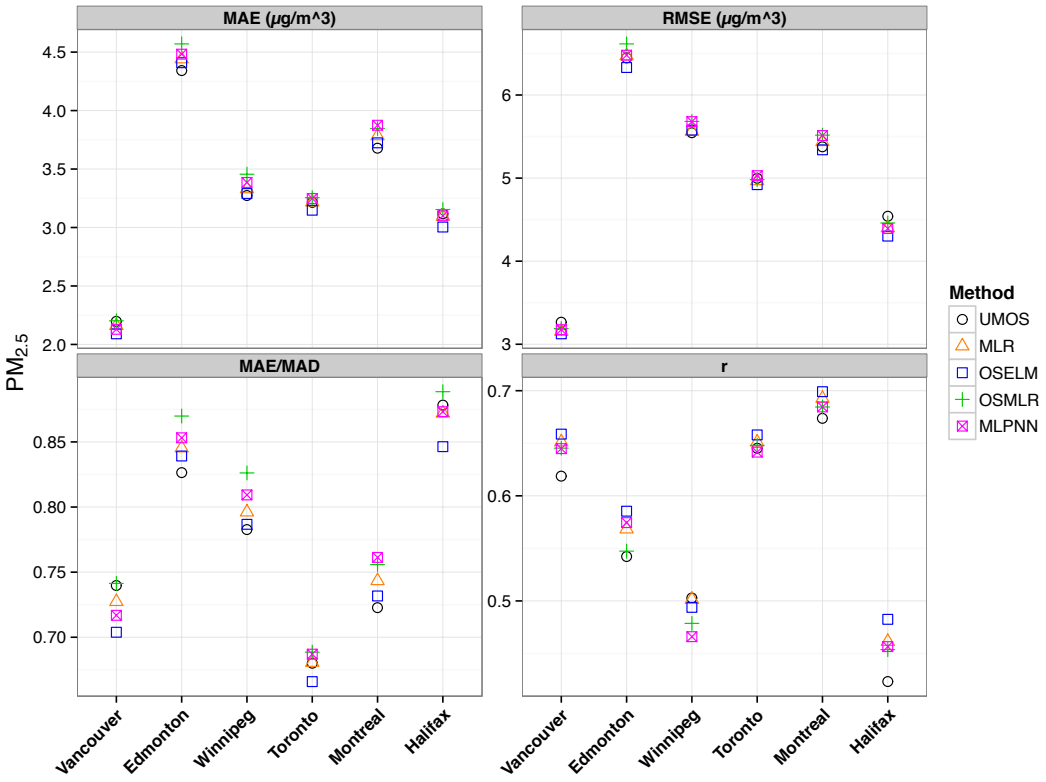


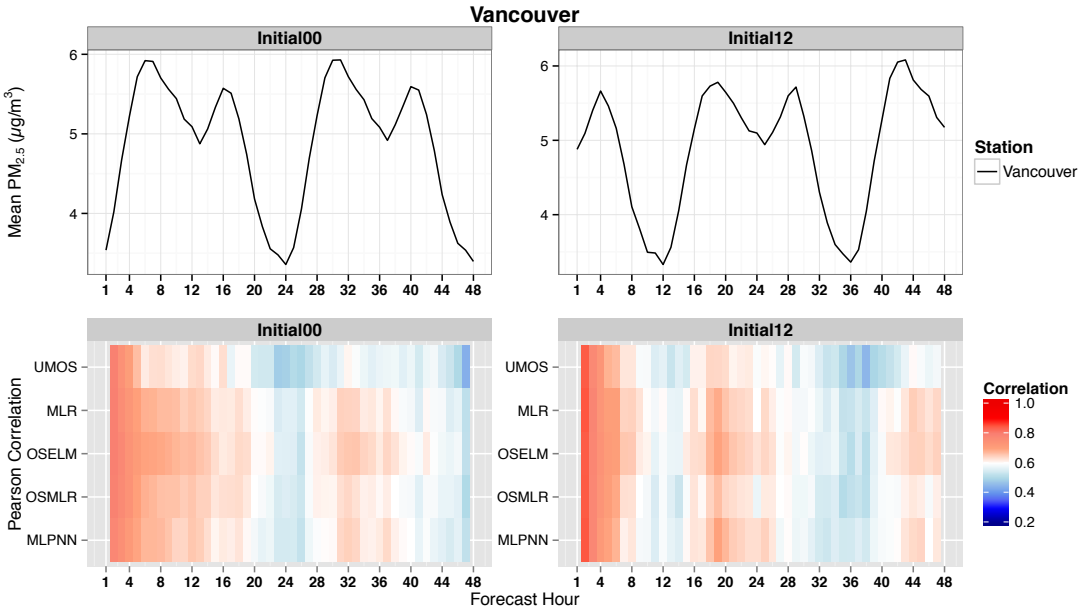

Figure 5.10: $PM_{2.5}$ forecast scores of different methods at the six test stations.

The mean diurnal cycle in the $PM_{2.5}$ concentration and the forecast correlation score for different forecast lead times and initial hours are shown in Fig 5.11. The heat map shows $r > 0.6$ in red, $r = 0.6$ in white, $r < 0.6$ in blue, and missing values in black. Station Vancouver, which has the lowest mean $PM_{2.5}$ concentration, shows two peak periods at night (11 pm LST) and in the morning (9 am LST) and the trough in late afternoon (5 pm LST). Forecast correlation scores vary according to the $PM_{2.5}$ diurnal cycle, as models have better performance around 9 am and 11 pm LST. The diurnal cycle in Edmonton is similar to that of Vancouver, with the highest concentration happening in the morning (10 am LST) and at night(10 pm LST). However, models only have $r > 0.6$ for 1-8hr forecast lead time for both 00 UTC and 12 UTC initiation time. OSELM and MLP NN slightly outperformed other methods for 12-20hr forecast lead

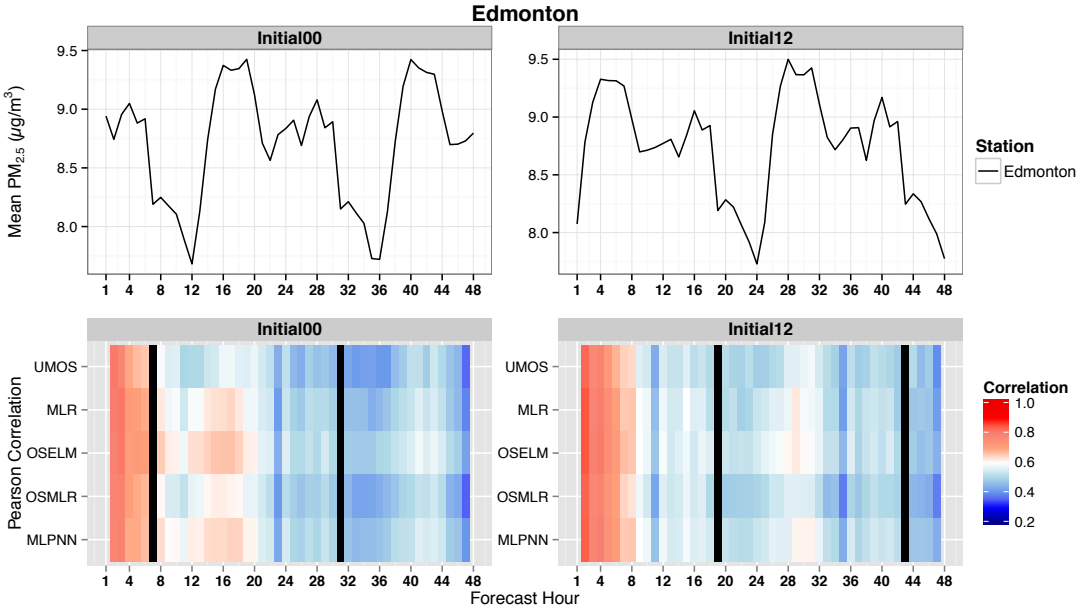
time at 00 UTC initiation. Models in Winnipeg performed worse with $r < 0.6$ for most forecast hour, especially for long lead time.

Correlation scores in Toronto and Montreal are the best among the six stations. $PM_{2.5}$ concentration peaks in the morning (9 am LST) with a secondary peak at night (10 pm LST) for both station. In Toronto, models have $r > 0.6$ during 1-24 forecast hours for 00 UTC initiation and 1-36 forecast hours for 12 UTC initiation. The correlation heat map of Montreal shows $r > 0.7$ for 1-24 forecast hours and the models performed well even during the low concentration period in the diurnal cycle (32-36 forecast lead times of the 12 UTC models). Halifax has a similar mean $PM_{2.5}$ value as Vancouver, but the models' scores are much worse. Most of the correlation scores are below 0.6, during both high and low concentration periods. UMOS loses to other methods for the 20-24 and 40-48 hour forecasts initiated at 00 UTC, whereas OSELM tends to slightly outperform the other models.

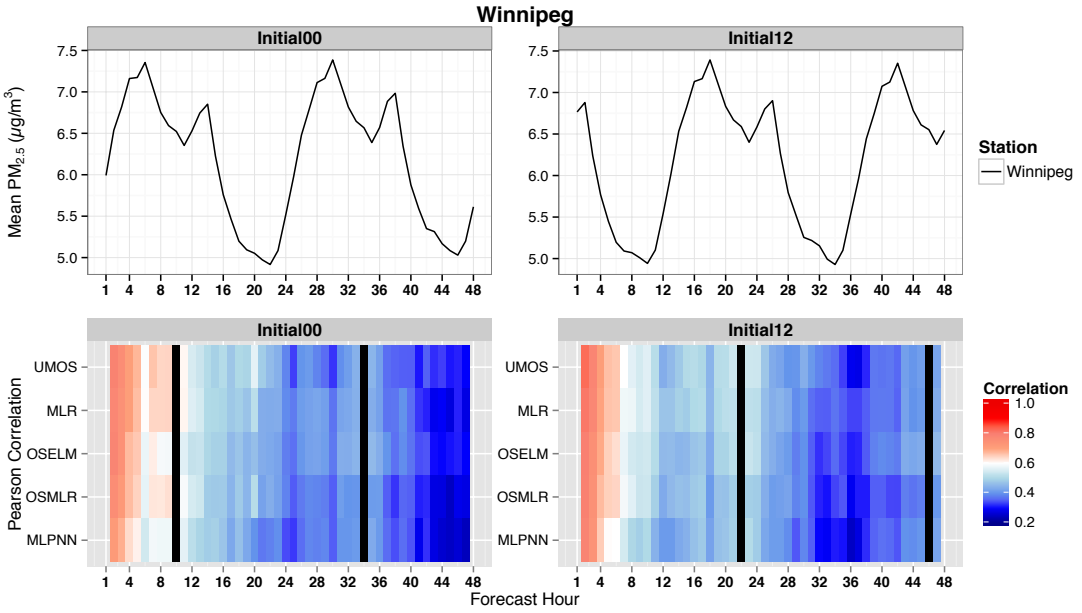
$PM_{2.5}$ model MAE skill scores (relative to UMOS) are plotted in Fig 5.12. For Vancouver, positive skill scores are found for most forecast lead times for the 00 UTC initiated models, while only OSMLR loses to UMOS in the 12 UTC initiated models. OSELM and MLP NN tend to have better scores than the two linear daily updating methods in Vancouver. For Edmonton, Winnipeg and Montreal, all of the models developed underperform UMOS (by about 5%), especially for the MLP NN and OSMLR models. OSELM is the only one surpassing UMOS for most forecast lead times in Toronto and Halifax.


Vancouver
Initial00
Initial12
Mean PM2.5 (μg/m3)
Station
Vancouver
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
Forecast Hour

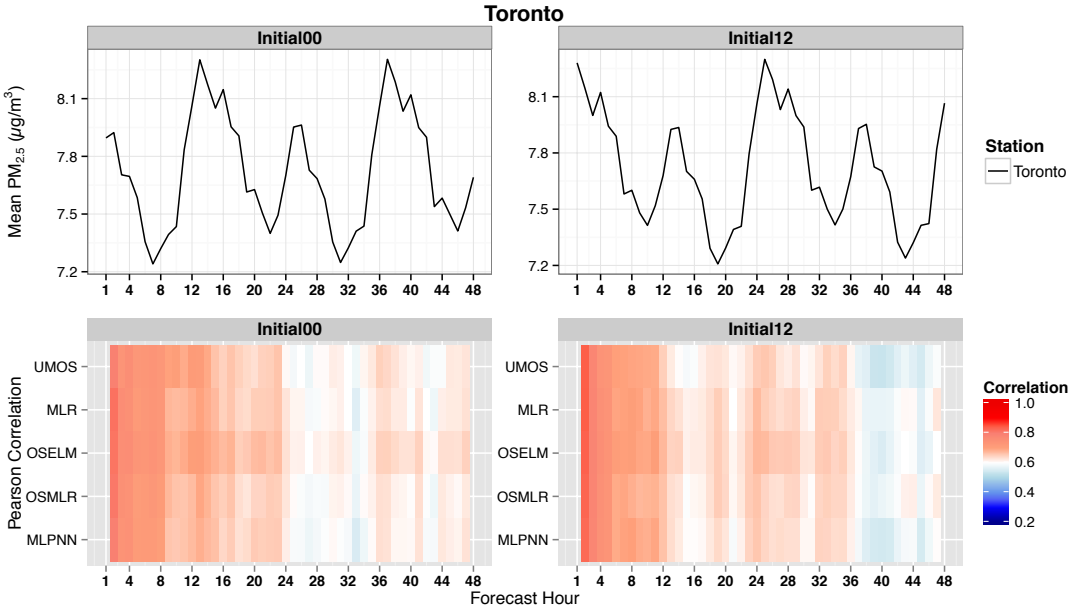
(a)


Edmonton
Initial00
Initial12
Mean PM2.5 (μg/m3)
Station
Edmonton
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
Forecast Hour

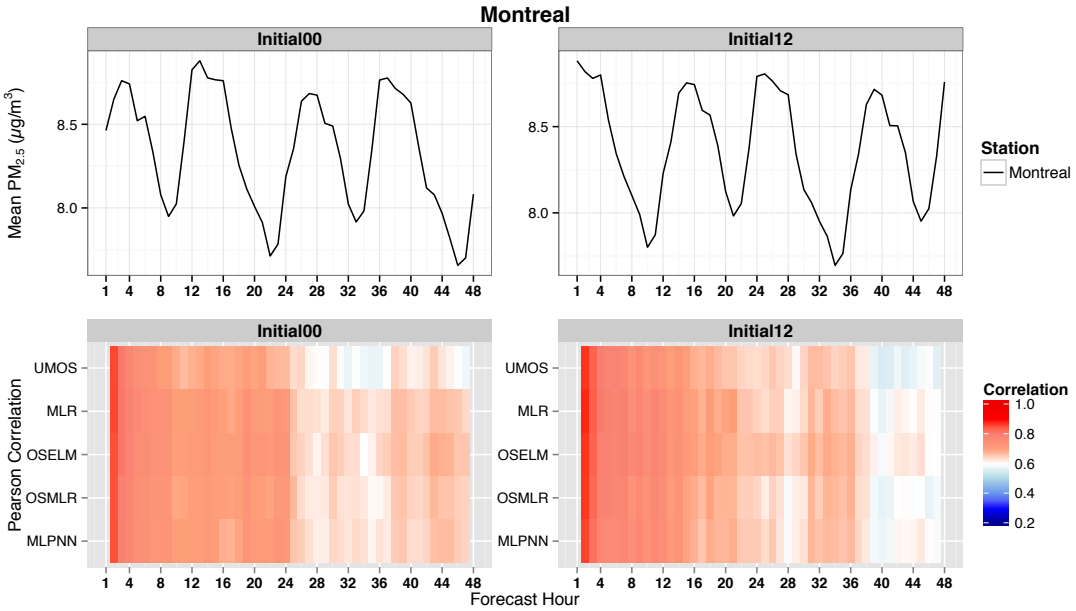
(b)


Winnipeg
Initial00
Initial12
Mean PM2.5 (μg/m3)
7.5
7.0
6.5
6.0
5.5
5.0
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Winnipeg
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
Forecast Hour

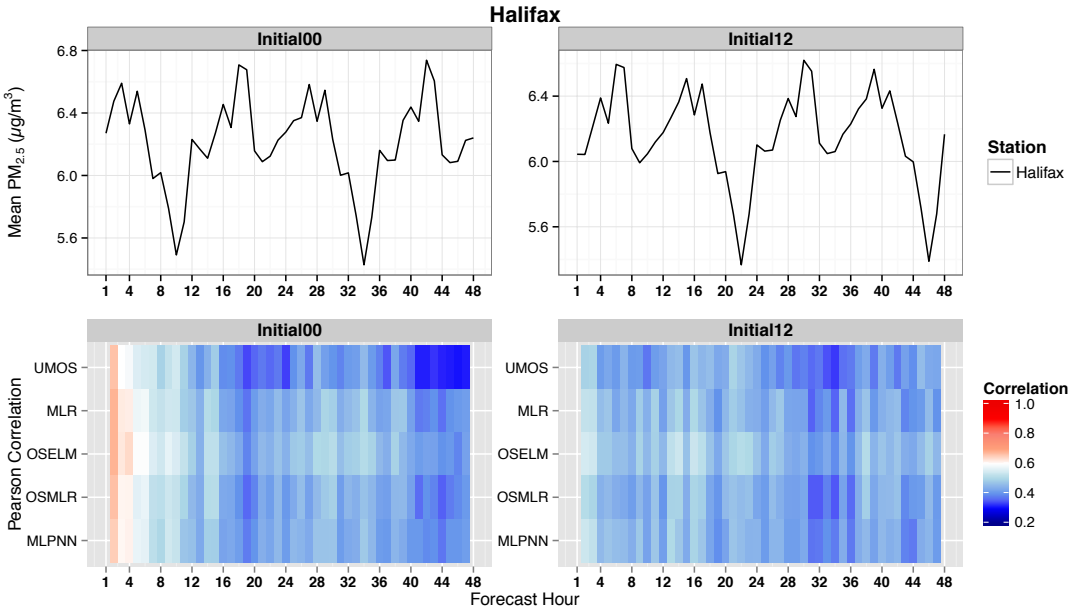
(c)


Toronto
Initial00
Initial12
Mean PM2.5 (μg/m3)
8.1
7.8
7.5
7.2
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Toronto
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
Forecast Hour

(d)




(e)




(f)

Figure 5.11: Mean diurnal $PM_{2.5}$ concentration and heat map of the correlation score by model and station for forecast lead time 1-48hr and forecasts initiated at 00 UTC (left) and 12 UTC (right).

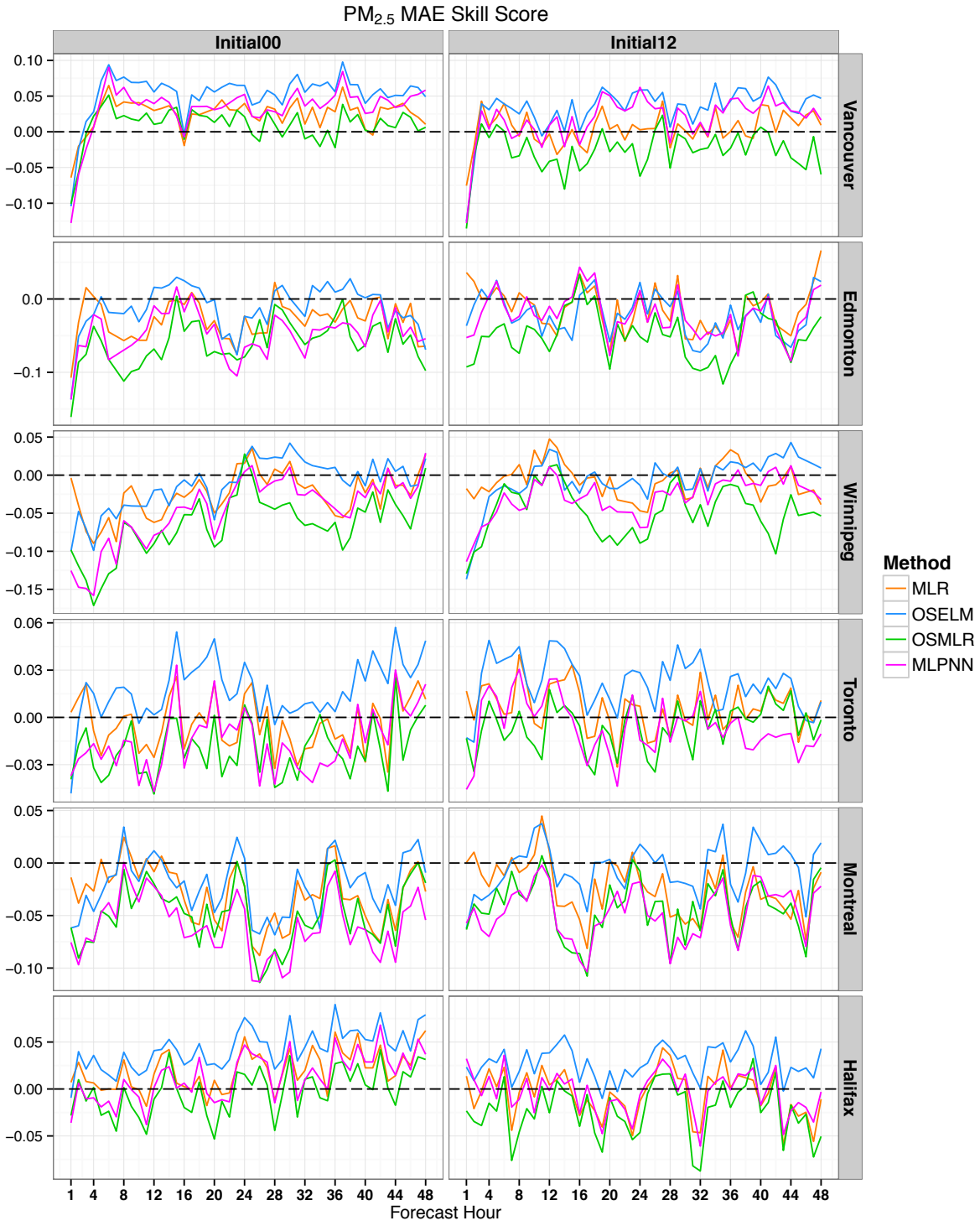



Figure 5.12: $PM_{2.5}$ MAE skill score of different models by forecast hour for the six stations.

Table 5.3 presents statistical properties of $PM_{2.5}$ concentration by station and season. High and low $PM_{2.5}$ values do not correspond to different seasons. Vancouver, Edmonton, Montreal and Halifax have lower mean values during the warm season, though the outliers value of 147 $\mu g/m^3$ (Fig 5.9) was found in the warm season in Edmonton. In contrast, Winnipeg and Toronto have higher mean $PM_{2.5}$ concentration during the warm season. Fig 5.13 and Fig 5.14 display model performance by season according to the residuals and forecast scores. Model residuals are relative smaller in magnitude in the warm season according to the first and third quartiles in Fig 5.13. According to the median of the residuals (Fig 5.13), overprediction mainly occur in Edmonton, Winnipeg and Montreal for both seasons, which is consistent with the models performing worst relative to UMOS at these three station (Fig 5.12).

In Fig 5.14, there is little difference in the forecast scores among the models, however there are differences in the scores between the warm and cold seasons - e.g. all the models perform better in the warm season in Toronto and Halifax, and in the cold season in Edmonton and Montreal. However, the much poorer MAE/MAD and $r$ scores in the warm season relative to the cold season for Edmonton could be caused by the outlier of 147 $\mu g/m^3$ in the warm season data as noted earlier. For the correlation score, OSELM was marginally ahead of all the other models in the cold season for all stations, and in the warm season for all stations except Edmonton and Winnipeg.

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 4.6 | 3.0 | 24.0 | 5.2 | 5.1 | 48.0 |
| Edmonton | 8.1 | 6.9 | 147.0 | 9.3 | 8.1 | 97.0 |
| Winnipeg | 6.7 | 6.9 | 83.0 | 5.7 | 5.9 | 88.0 |
| Toronto | 8.1 | 6.2 | 62.0 | 7.3 | 6.7 | 75.0 |
| Montreal | 7.9 | 6.6 | 88.0 | 8.9 | 7.7 | 75.0 |
| Halifax | 6.1 | 4.9 | 46.0 | 6.3 | 5.0 | 55.0 |

Table 5.3: Statistical properties of $PM_{2.5}$ concentration ($\mu g/m^3$) by station and season.

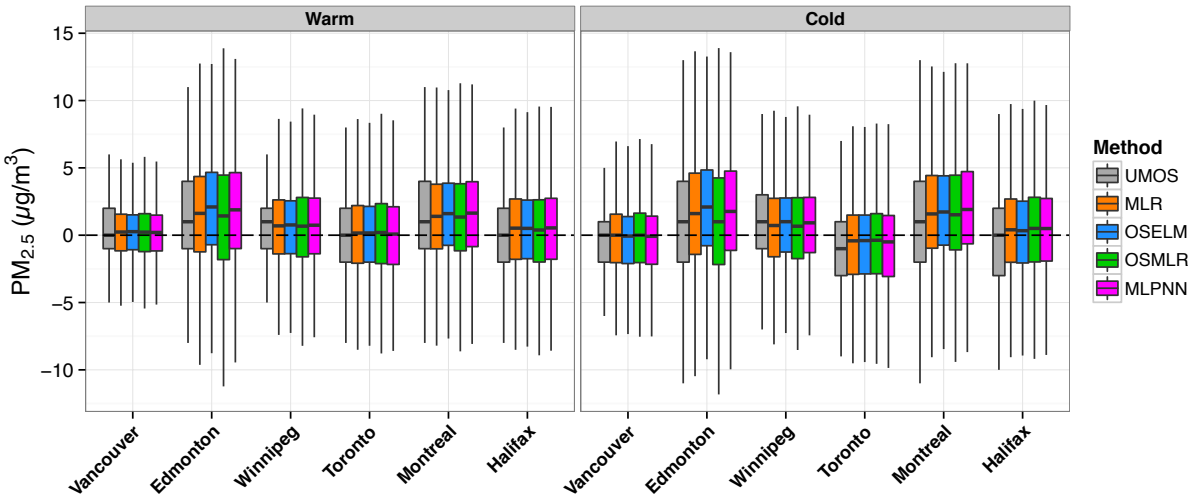



Figure 5.13: Boxplot of $PM_{2.5}$ residuals (prediction−observation) by season and station. Outliers are not plotted but can be seen in Fig 5.15.

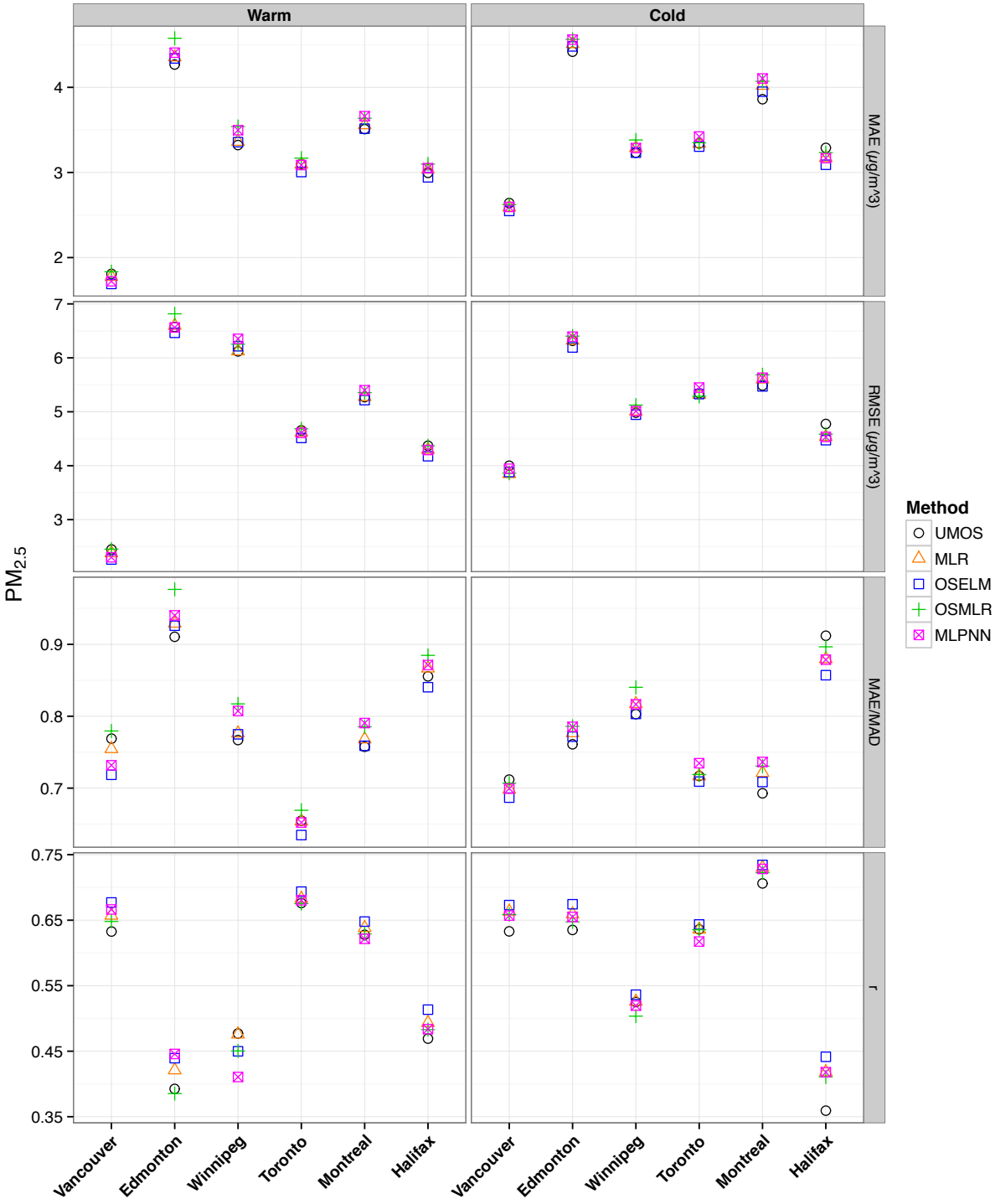



Figure 5.14: $PM_{2.5}$ forecast scores by season and station.

The top $10^{th}$ percentile of the observational data and the corresponding model predictions are extracted by season and station. Table 5.4 shows that the mean extreme value in the cold season is higher than that in the warm season for all stations except Winnipeg, and the maximum $PM_{2.5}$ concentrations are also higher for the cold season in Vancouver, Winnipeg, Toronto and Halifax. The boxplot of residuals (Fig 5.15) shows that all five models were unable to capture the maximum $PM_{2.5}$ values (147 $\mu g/m^3$) in Edmonton, which would have a major impact on the forecast scores. Fig 5.16 presents the error and correlation scores for different models by season. All of the MAE/MAD values are above 1 and most correlations are below 0.3, which indicate weak model performance. For extremes, the nonlinear models in general did not improve on the linear models (UMOS, MLR and OSMLR).

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 11.0 | 2.4 | 24.0 | 17.0 | 5.1 | 48.0 |
| Edmonton | 22.6 | 9.7 | 147.0 | 27.0 | 9.5 | 97.0 |
| Winnipeg | 20.7 | 12.4 | 83.0 | 19.2 | 7.2 | 88.0 |
| Toronto | 21.7 | 5.5 | 62.0 | 22.6 | 7.5 | 75.0 |
| Montreal | 21.5 | 9.1 | 88.0 | 26.2 | 7.6 | 75.0 |
| Halifax | 16.7 | 5.0 | 46.0 | 17.1 | 5.3 | 55.0 |

Table 5.4: Statistical properties of top $10^{th}$ percentile $PM_{2.5}$ concentration ($\mu g/m^3$) by station and season.

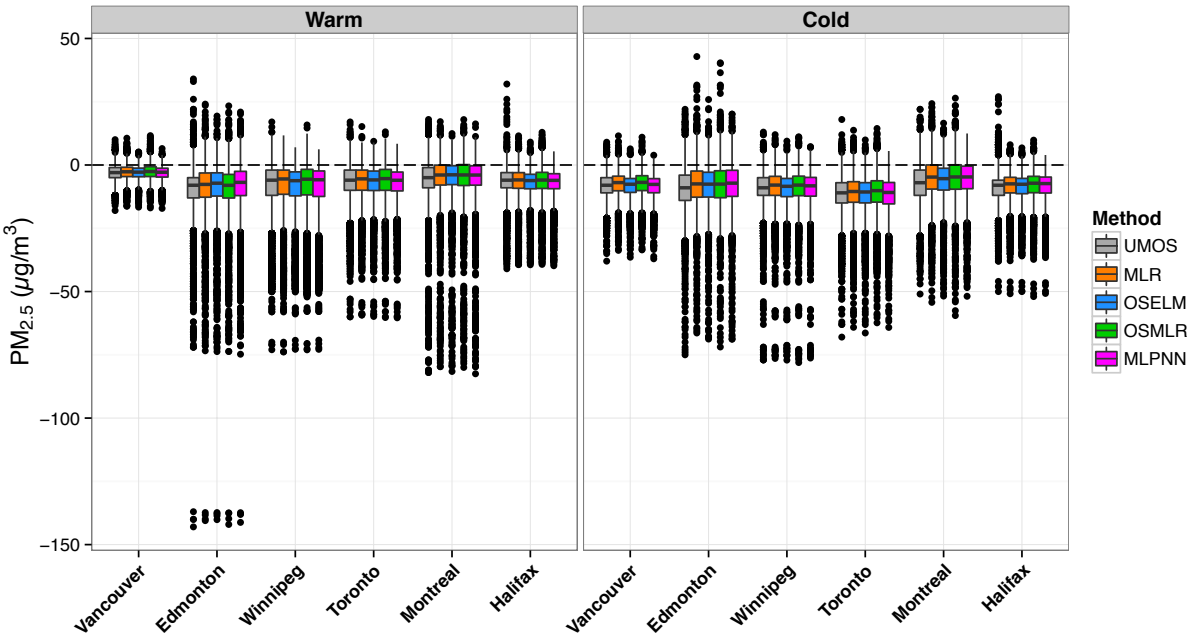



Figure 5.15: Boxplot of $PM_{2.5}$ residuals (prediction−observation) (over all lead times) from the top $10^{th}$ percentile by season and station.

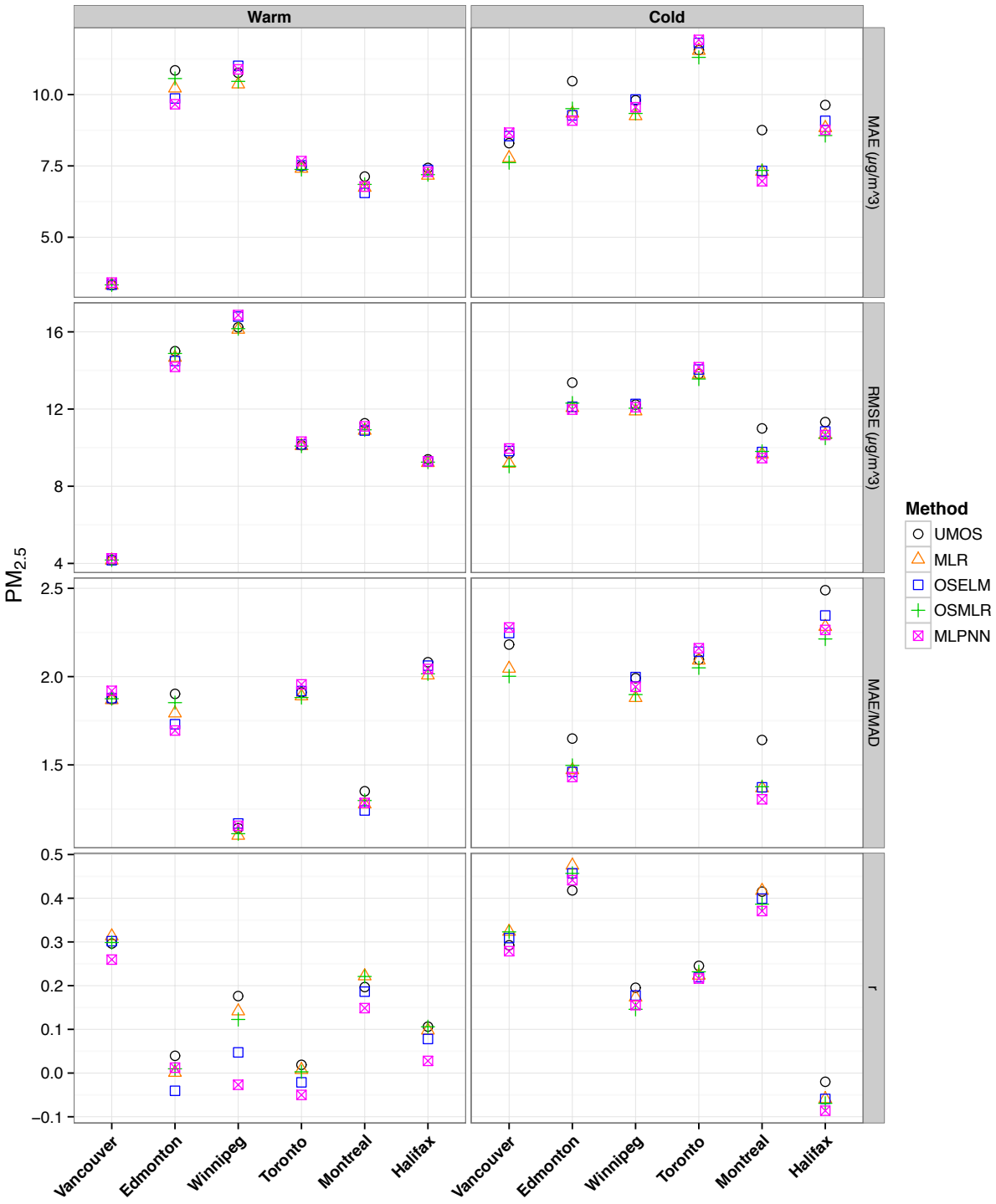



Figure 5.16: $PM_{2.5}$ forecast scores of top $10^{th}$ percentile by season and station.

## 5.3 $NO_2$

Boxplot of $NO_2$ observations and predictions by station are shown in Fig 5.17. Edmonton has highest median $NO_2$ concentration (16 ppb) and the most extreme events, whereas Halifax's median concentration is only 1.4 ppb. The station in Halifax is located near Lake Major, which may explain the low $NO_2$ concentration as automobile emission is the main source of $NO_2$. Fig 5.17 also indicates the model medians to lie above the observed median for all stations. The forecast scores in Fig 5.18 show all five methods to have similar performance, and except for Halifax, the relative errors are generally below 0.7, with OSELM being marginally better than the other methods. In Halifax, UMOS has lowest relative error, but it is still greater than 1. From the correlation score, OSELM is slightly ahead and UMOS is slightly behind all other methods at all five stations.

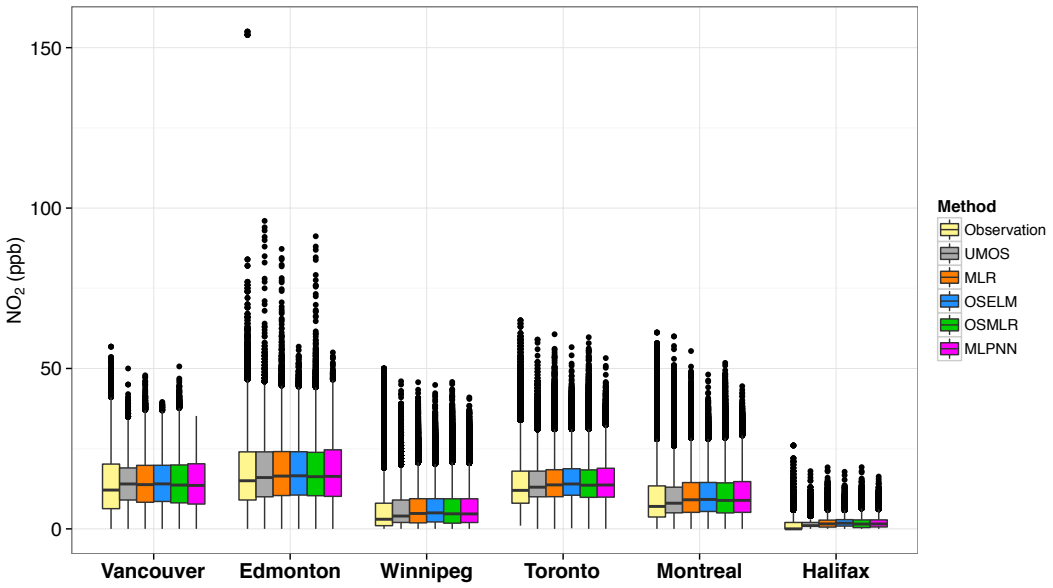



Figure 5.17: Boxplot of the observed $NO_2$ values and the predicted values from five models (over all forecast lead times ) at six stations.

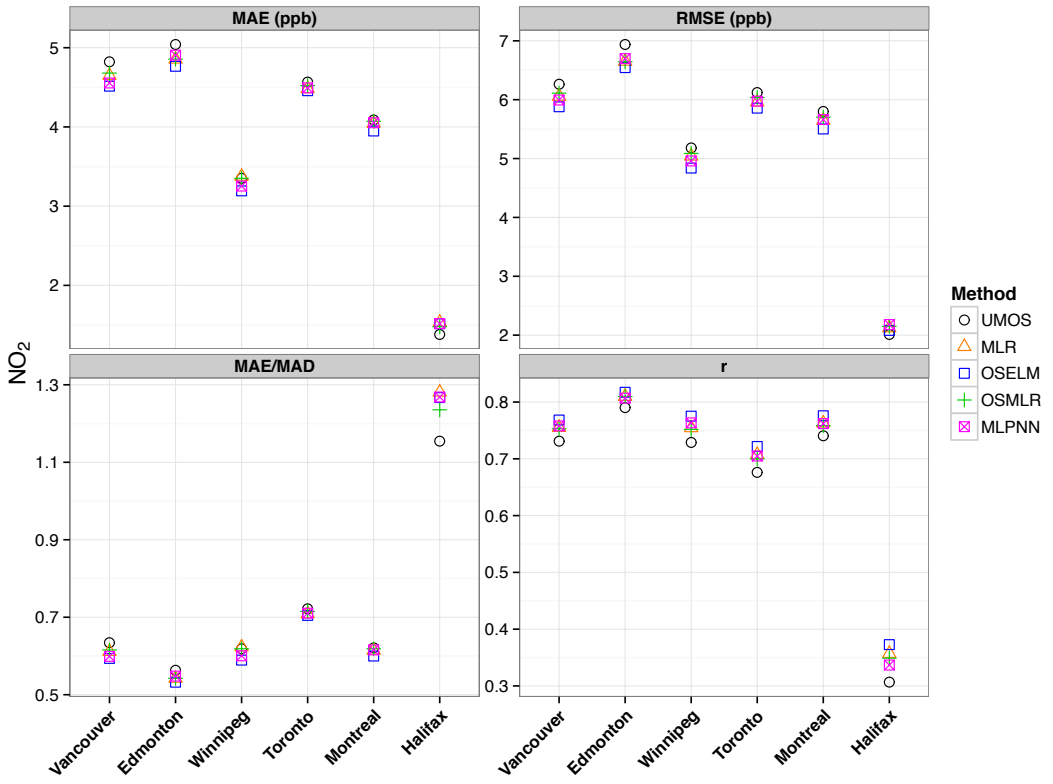
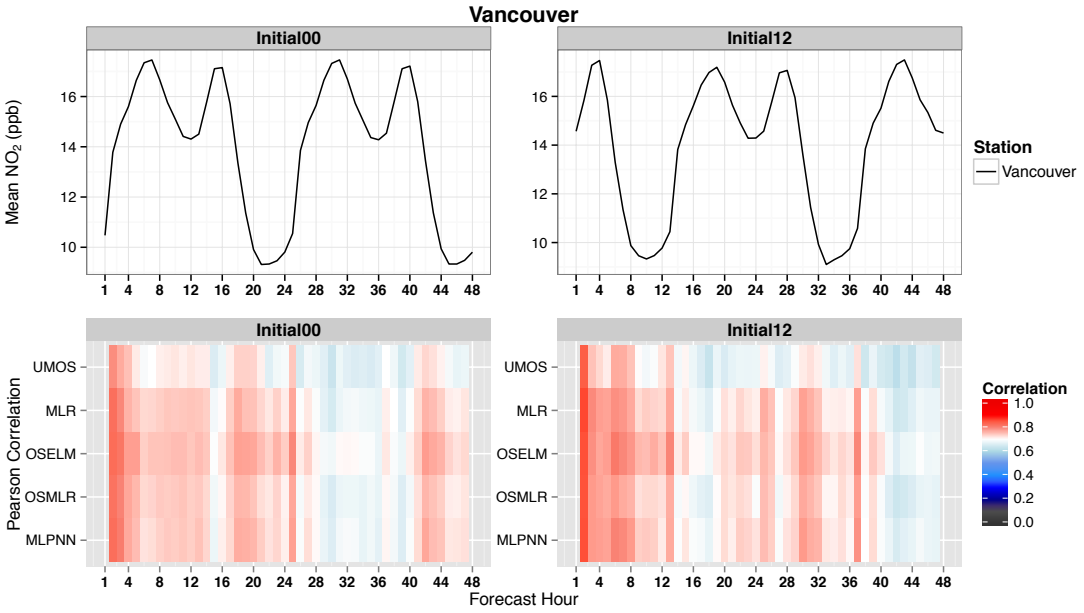



Figure 5.18: $NO_2$ forecast scores of different methods in the six stations.

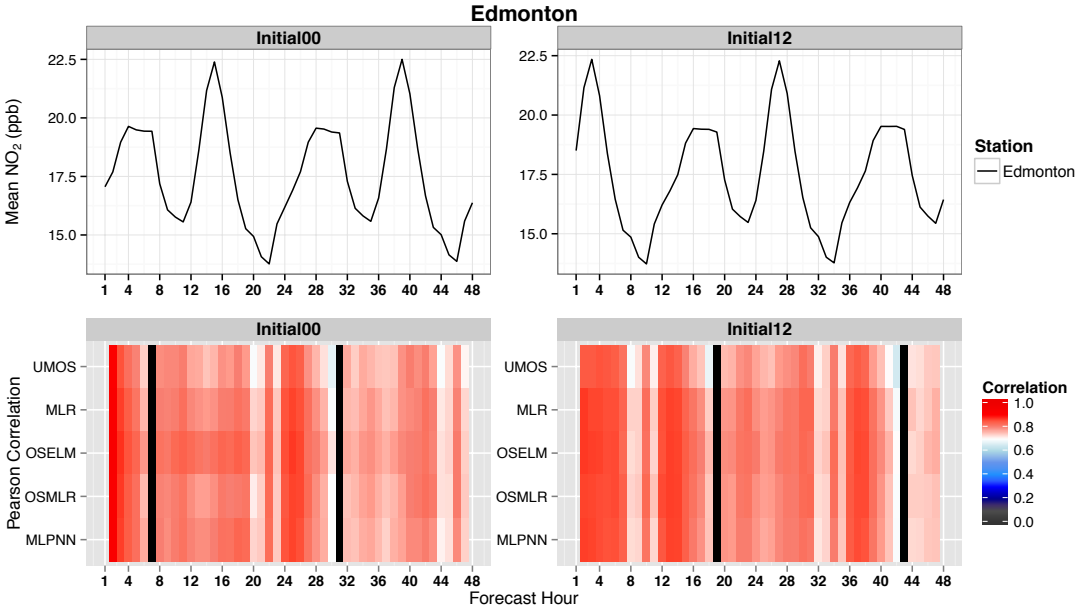
In Fig 5.19, the heat map displays $r > 0.7$ in red and $r < 0.7$ in blue and missing data at Edmonton and Winnipeg in black. $NO_2$ often has the highest diurnal value during the morning (9 am LST), with a second peak during the night (10-11 pm LST) and becoming lower in the afternoon (2-4 pm LST). The main except occurs in Vancouver where the second peak is slightly higher than the first, and in Halifax where the highest peak occurs at 3-4 am LST. Models in Vancouver have good performance during 1-25 hour forecasts when initiated at 00 UTC and UMOS underperforms other method for 20-40 hour forecasts when initiated at 12 UTC. Edmonton, which has the highest mean concentration among the six stations, has the best model behavior among the stations. The difference in $r$ between the five methods is small, though OSELM and MLR are slightly stronger and UMOS slightly weaker than others. In Winnipeg, OSELM is slightly stronger and UMOS slightly weaker than the others, with all models forecasting poorer during the trough in the diurnal cycle.

In Toronto, the mean $NO_2$ concentration in the morning is over 17.5 ppb. Correlation scores for 1-15 forecast hours from 00 UTC and 1-5 forecast hours from 12 UTC are generally higher than other lead times, again with OSELM being slightly stronger and UMOS slightly weaker among the models. In Montreal, models have relatively high correlation scores, but poorer performance can be found during the low concentration period and the second peak period, corresponding to the local afternoon and midnight. OSELM and MLR are slightly stronger and UMOS slightly weaker among the models. Halifax has the worst model scores, with correlation below 0.4 most of the time.

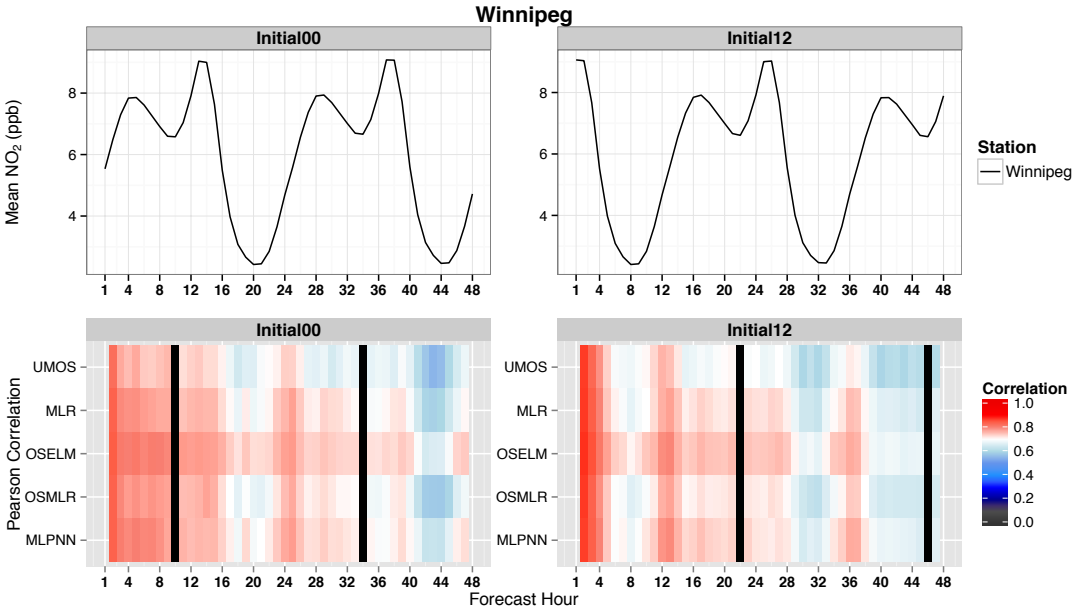
MAE skill scores relative to UMOS (Fig 5.20) illustrates that MLR, OSELM, OSMLR and MLP NN have positive skill scores in Vancouver and Edmonton for most forecast hours, and OSELM slightly outperforms other methods, whereas OSMLR and MLP NN slightly underperform in Vancouver and Edmonton, respectively. For Winnipeg, Toronto and Montreal, negative skill scores occur during the low concentration period in the diurnal cycle, upon comparing Fig 5.20 with Fig 5.19, indicating that UMOS is slightly better at forecasting during low $NO_2$ hours. In Winnipeg and Montreal, OSELM slightly outperforms MLR, OSMLR and MLP NN, but all four models have comparable skills in Toronto. In Halifax, all four methods lose to UMOS in the MAE skill score.


Vancouver
Initial00
Initial12
Mean NO2 (ppb)
16
14
12
10
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Vancouver
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
0.0
Forecast Hour

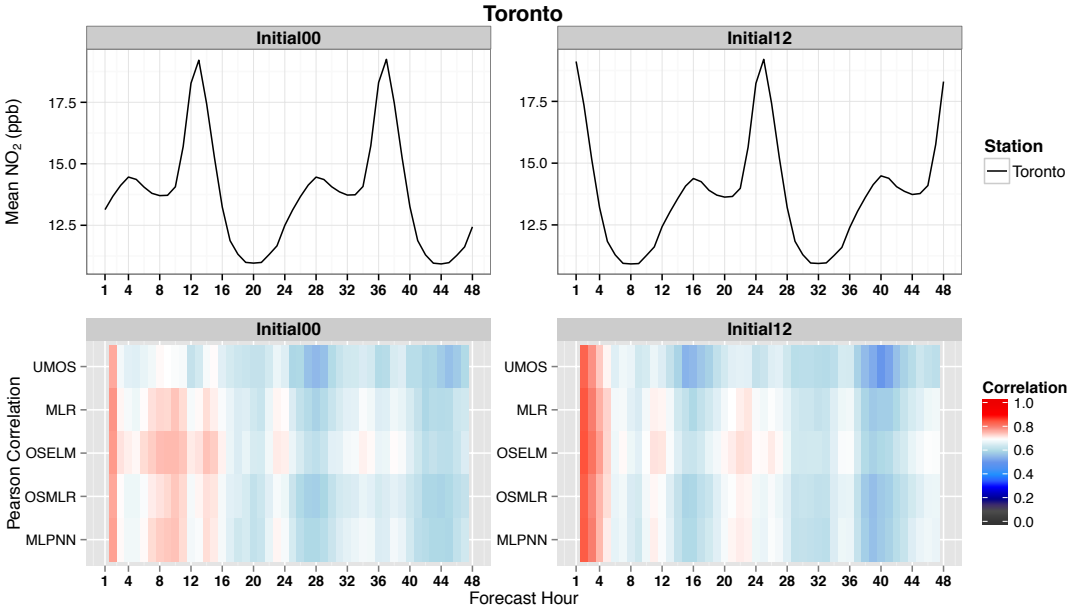
(a)


Edmonton
Initial00
Initial12
Mean NO2 (ppb)
22.5
20.0
17.5
15.0
1 4 8 12 16 20 24 28 32 36 40 44 48
Station
Edmonton
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
1.0
0.8
0.6
0.4
0.2
0.0
Forecast Hour

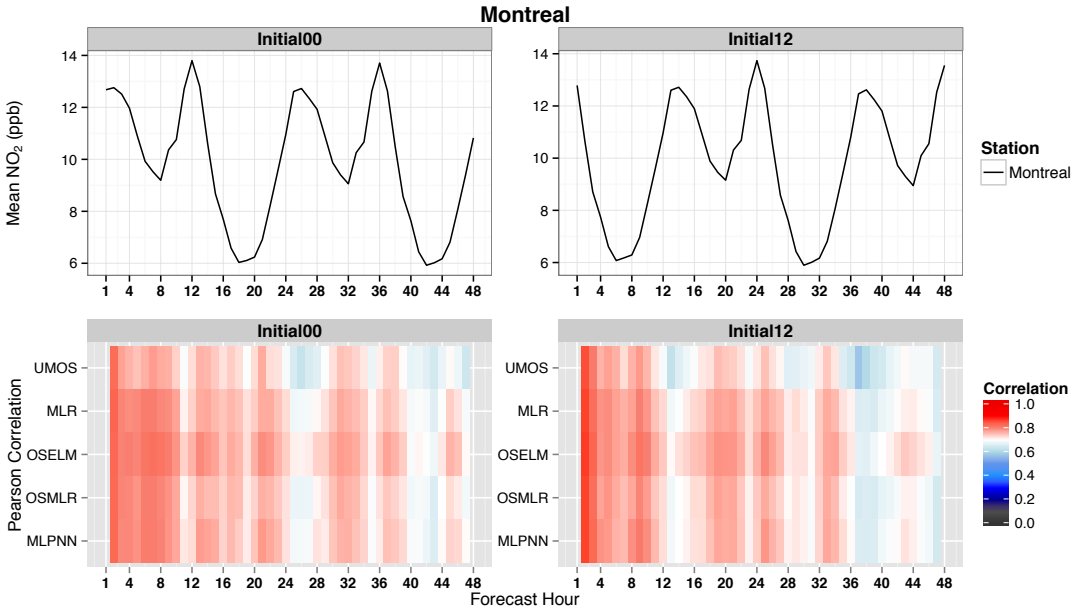
(b)


Winnipeg
Initial00
Initial12
Mean NO2 (ppb)
Station
Winnipeg
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
Forecast Hour

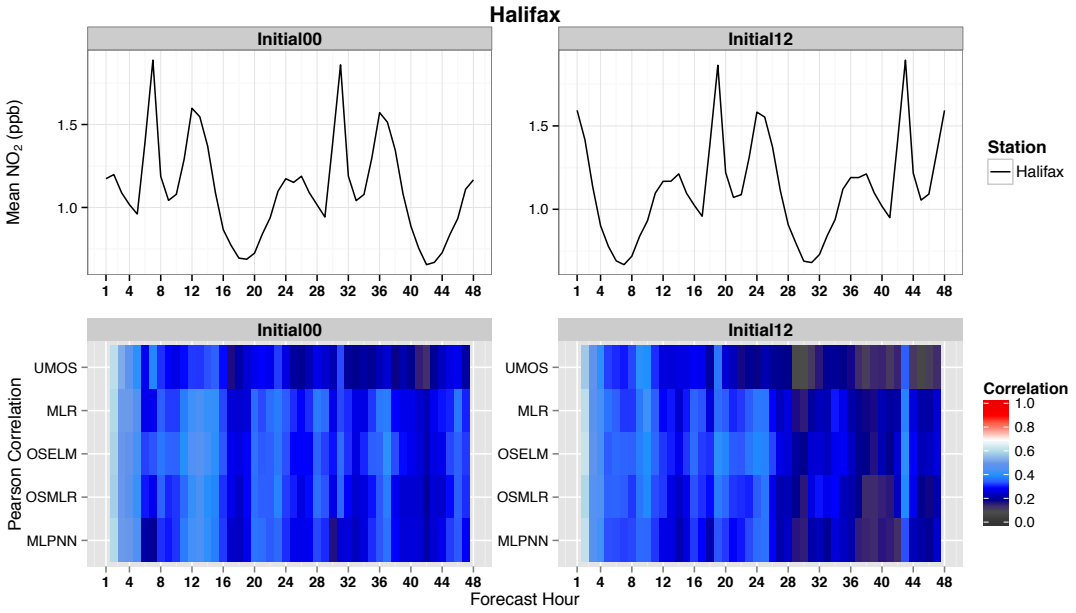
(c)


Toronto
Initial00
Initial12
Mean NO2 (ppb)
Station
Toronto
Pearson Correlation
UMOS
MLR
OSELM
OSMLR
MLPNN
Correlation
Forecast Hour

(d)




(e)




(f)

Figure 5.19: Mean diurnal $NO_2$ concentration and heat map of the correlation score by model and station.

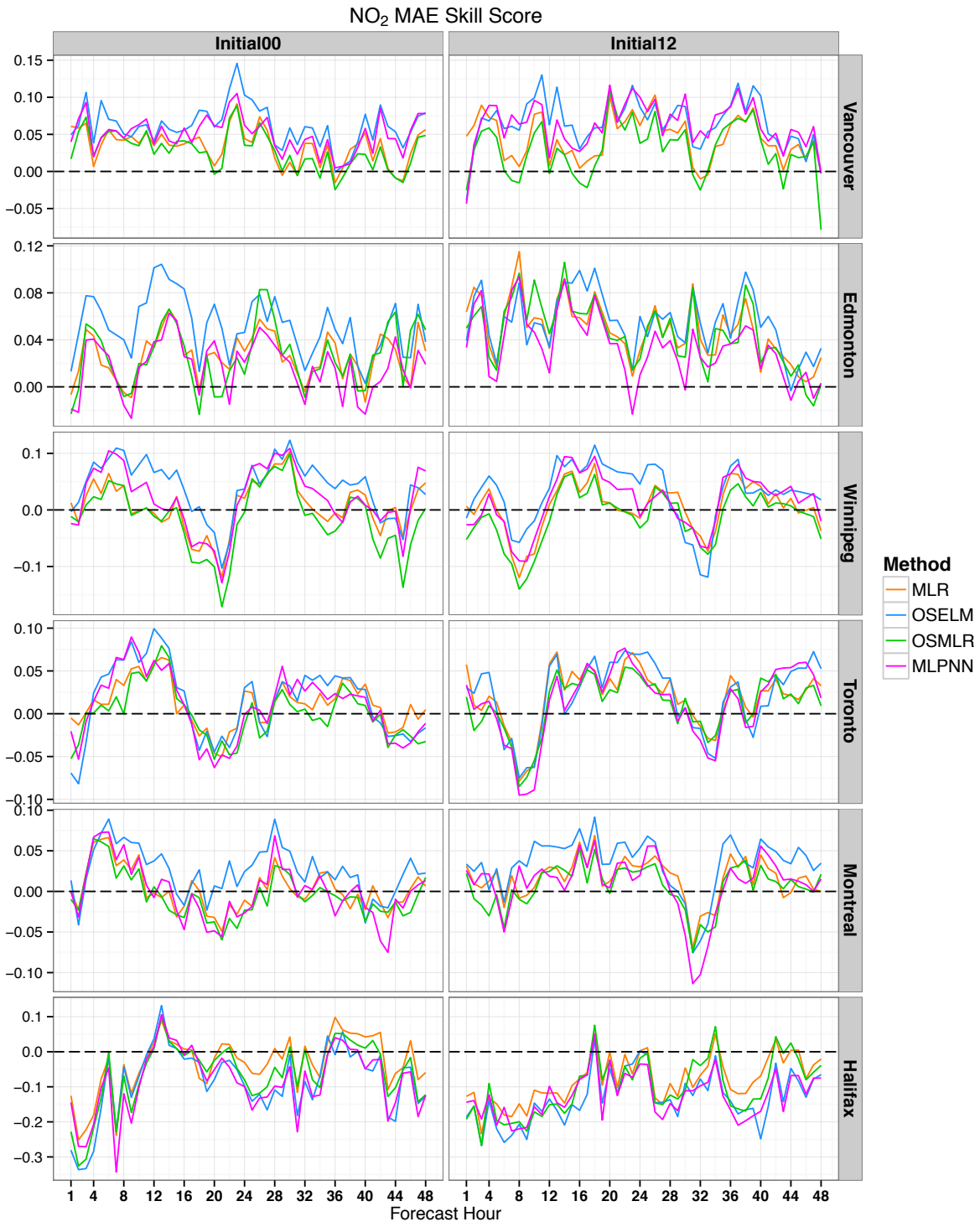



Figure 5.20: $NO_2$ MAE skill score of different models by forecast hour for the six stations.

To analyze the model performance seasonally, Table 5.5 shows the statistical properties of $NO_2$ by season and station, where higher $NO_2$ concentration occurs in the cold season for all six stations. For the warm season (Fig 5.21), linear models (MLR, OSMLR) have median residuals slightly closer to 0 and all our models have median residuals slightly closer to 0 than UMOS for all stations. However, UMOS has median residuals slightly better than most models in the cold season, especially in Halifax. The models have little spread in the errors (MAE and RMSE) (Fig 5.22), however, MLR, OSELM, OSMLR and MLP NN have slightly higher $r$ scores than UMOS at all stations for both the warm and cold seasons, with OSELM slightly ahead of the other methods.

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 11.2 | 7.6 | 46.8 | 17.2 | 9.7 | 56.8 |
| Edmonton | 11.6 | 7.7 | 62.0 | 23.6 | 11.3 | 155.0 |
| Winnipeg | 3.5 | 5.3 | 40.0 | 8.1 | 8.6 | 50.0 |
| Toronto | 12.0 | 7.5 | 60.0 | 15.6 | 8.6 | 65.0 |
| Montreal | 7.4 | 6.2 | 57.9 | 12.5 | 10.0 | 61.2 |
| Halifax | 1.0 | 1.5 | 21.0 | 1.3 | 2.0 | 26.0 |

Table 5.5: Statistical properties of $NO_2$ concentration (ppb) by station and season.

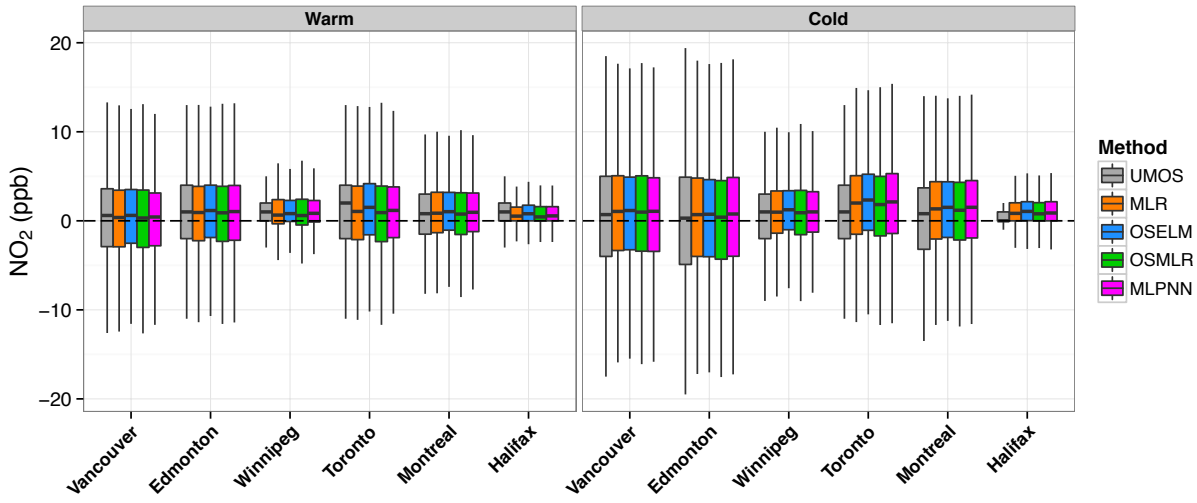



Figure 5.21: Boxplot of $NO_2$ residuals (prediction−observation) by season and station. Outiers are not plotted but can be seen in Fig 5.23.

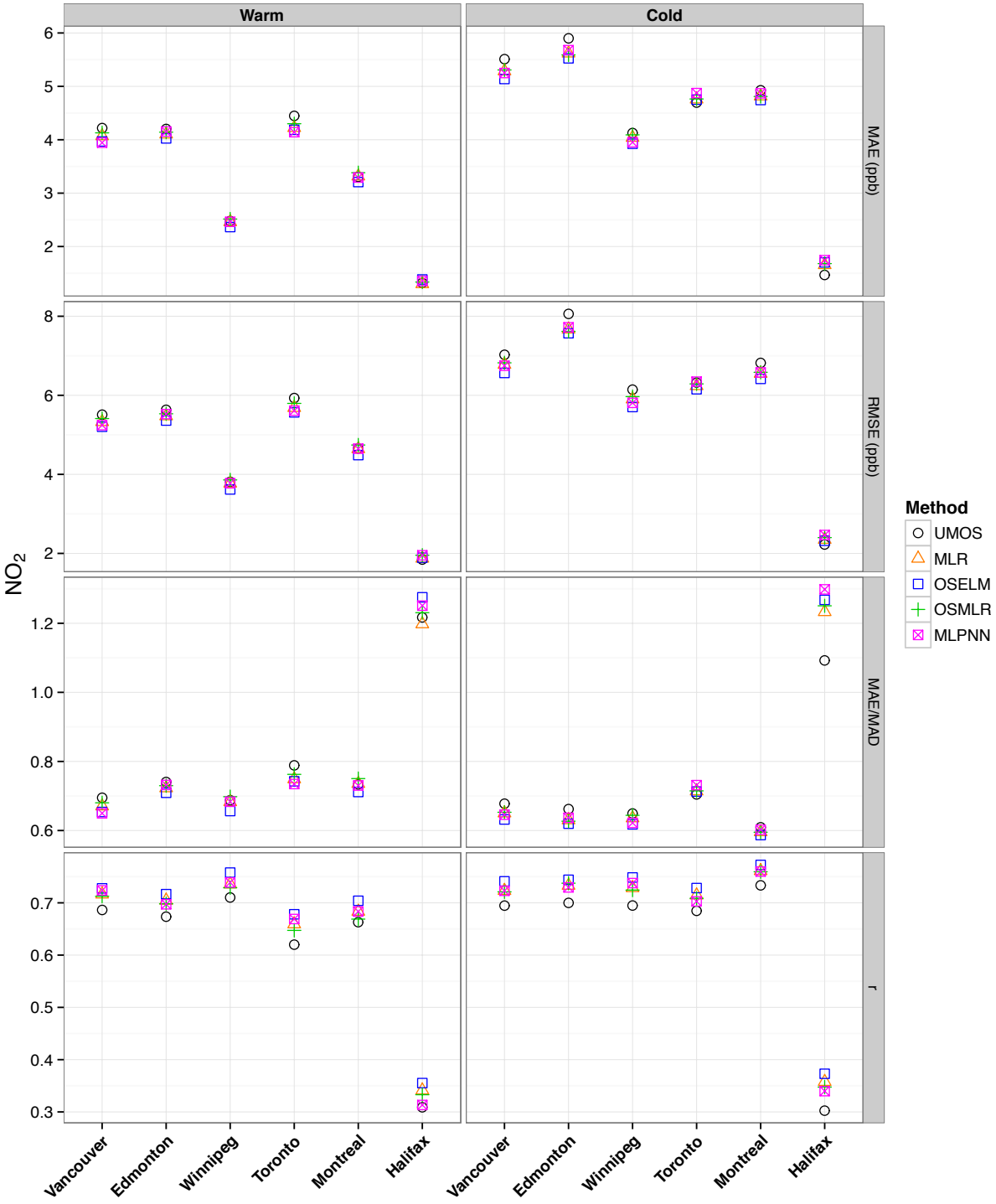



Figure 5.22: $NO_2$ forecast scores by season and station.

For the top $10^{th}$ percentile observations, Table 5.6 shows Edmonton to have the most extreme event with 155 ppb of $NO_2$. From the model residual shown in Fig 5.23, the medians are all below 0, indicating under prediction of the extremes, though our model medians tend to be slightly less negative than those from UMOS. The MAE, RMSE, MAE/MAD and correlation scores of top $10^{th}$ percentile data in Fig 5.24 show that model skills tend to be better in the warm season, and in terms of MAE/MAD our models tend to slightly outperform UMOS at all stations in both the warm and cold seasons.

| Station | Warm Season | | | Cold Season | | |
|---|---|---|---|---|---|---|
| | Mean | Std.Dev. | Max | Mean | Std.Dev. | Max |
| Vancouver | 27.3 | 4.08 | 46.8 | 34.83 | 3.83 | 56.8 |
| Edmonton | 27.78 | 6.73 | 62 | 45.36 | 8.62 | 155 |
| Winnipeg | 16.34 | 5.85 | 40 | 28.10 | 6.68 | 50 |
| Toronto | 28.83 | 5.91 | 60 | 34.58 | 6.12 | 65 |
| Montreal | 21.63 | 6.07 | 57.9 | 34.08 | 5.52 | 61.2 |
| Halifax | 4.49 | 1.94 | 21 | 5.8 | 2.67 | 26 |

Table 5.6: Statistical properties of top $10^{th}$ percentile $NO_2$ concentration (ppb) by station and season.

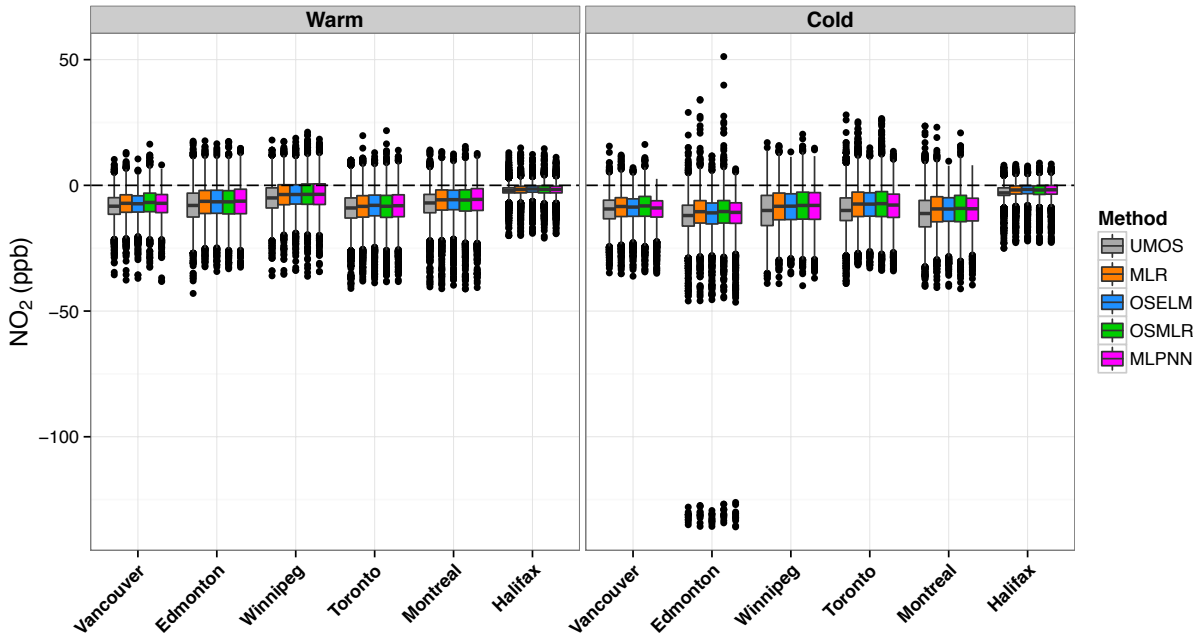



Figure 5.23: Boxplot of $NO_2$ residuals (prediction−observation) (over all lead times) from the top $10^{th}$ percentile by season and station.

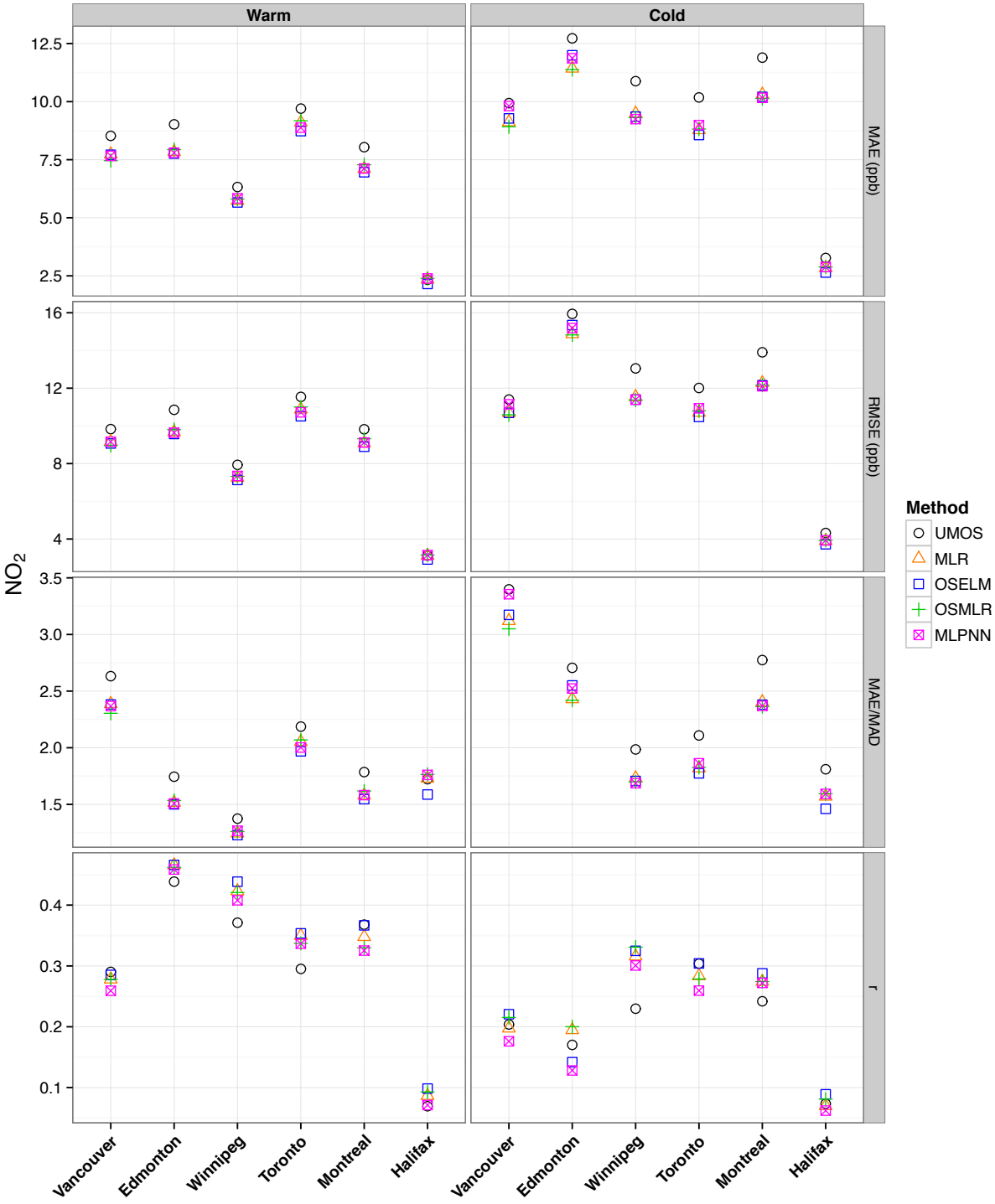



Figure 5.24: $NO_2$ forecast scores of top $10^{th}$ percentile by season and station.

## 5.4 Model Results with Antecedent Predictors

Nine antecedent predictors, i.e. the pollutant concentration 24 hours prior to the forecast time, the maximum and minimum pollutant concentration observed within the 24 hour period prior to the model initialization for ozone, $PM_{2.5}$ and $NO_2$, were added to the OSELM-A and OSMLR-A models to test whether they would contribute to the model accuracy, with the new model results compared with the original UMOS, OSELM and OSMLR results.

For ozone, models with antecedent predictors performed better than the original models in Winnipeg and Halifax but only marginally at the other four stations, as seen in the MAE, RMSE, MAE/MAD and $r$ (Fig 5.25). In the ozone MAE skill score (Fig 5.26), the models with antecedent predictors appear to improve on the original models mainly in Winnipeg and Halifax. Forecast scores for data in the top $10^{th}$ percentile (Fig 5.27) show that in the warm season, the antecedent predictors mainly improved the scores in Winnipeg and Edmonton, while in the cold season, they mainly improved in Halifax, followed by Winnipeg and Edmonton. We conclude that adding the antecedent predictors tend to improve on the ozone forecasts, especially in predicting extreme values at some stations.

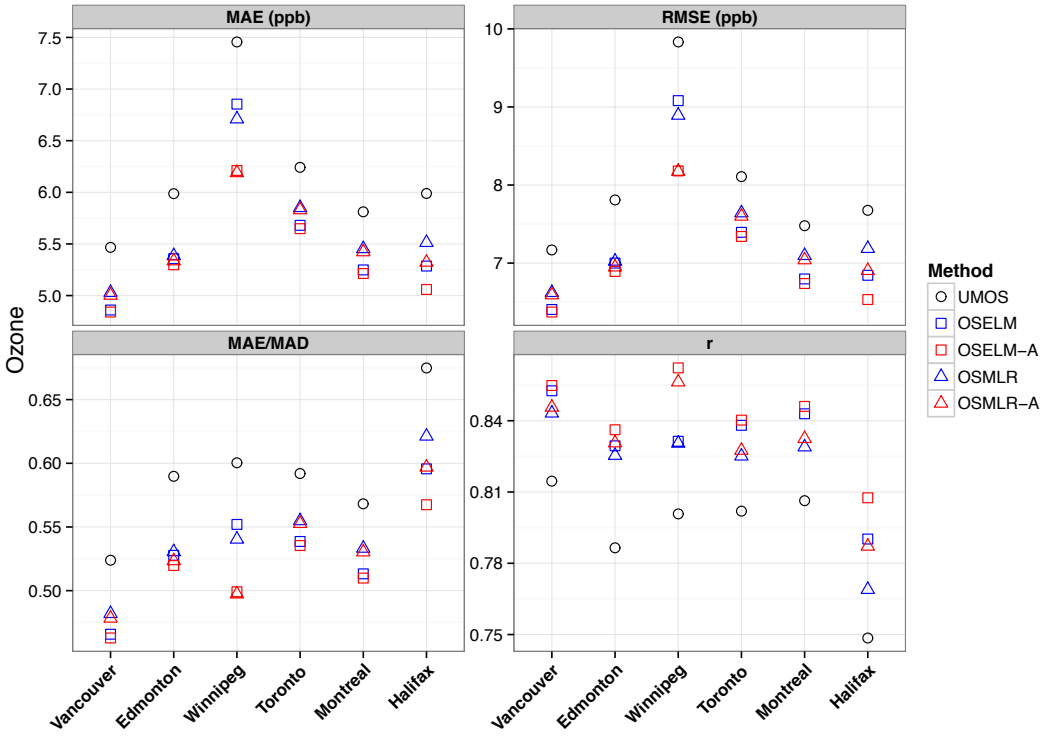



Figure 5.25: Ozone forecast scores from models with and without antecedent predictors in the six stations. Models with antecedent predictors (OSELM-A and OSMLR-A) are in red, the original models without the extra predictors are in blue, and UMOS is in black.

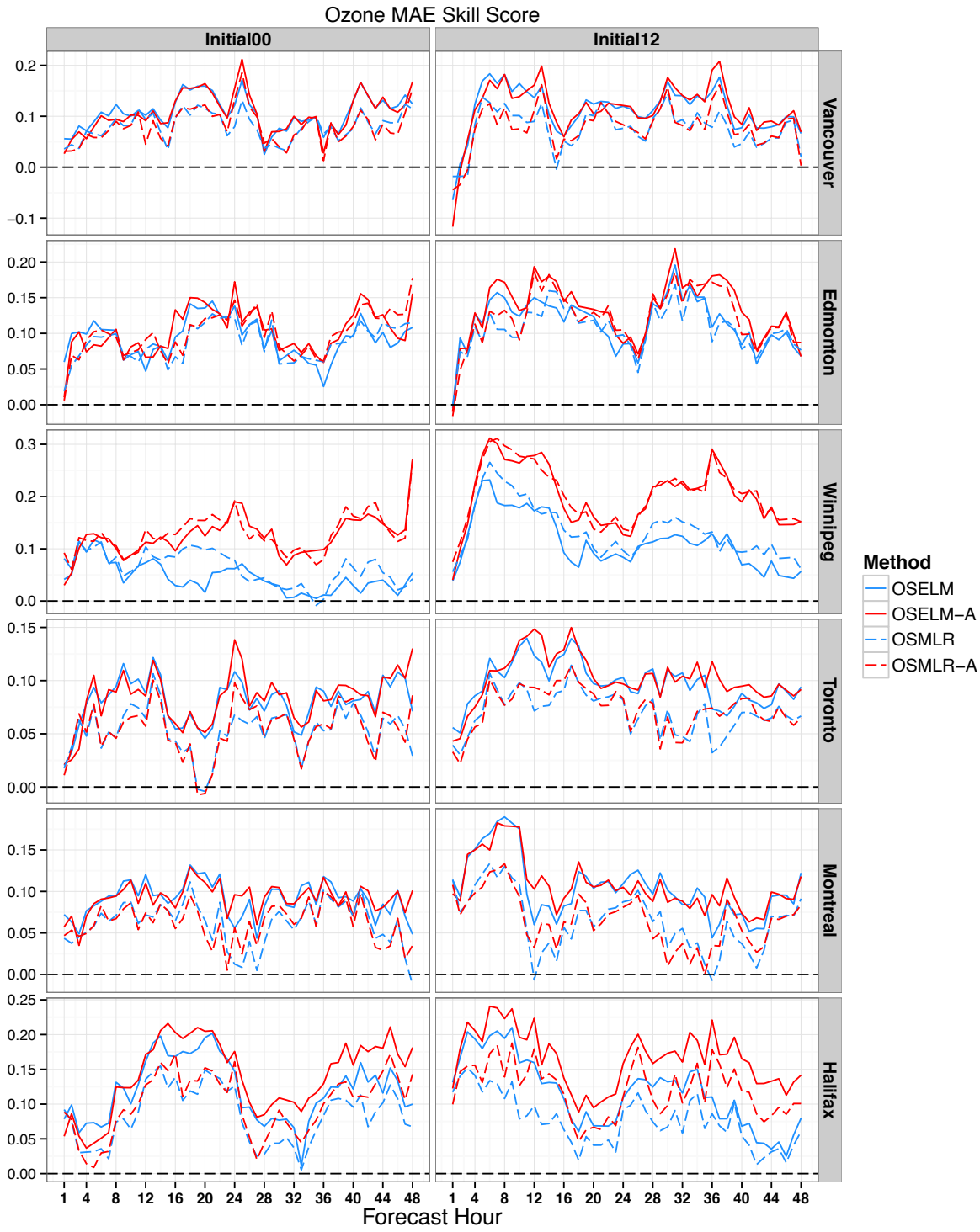



Figure 5.26: Ozone MAE skill score from models with and without antecedent predictors by forecast hour.

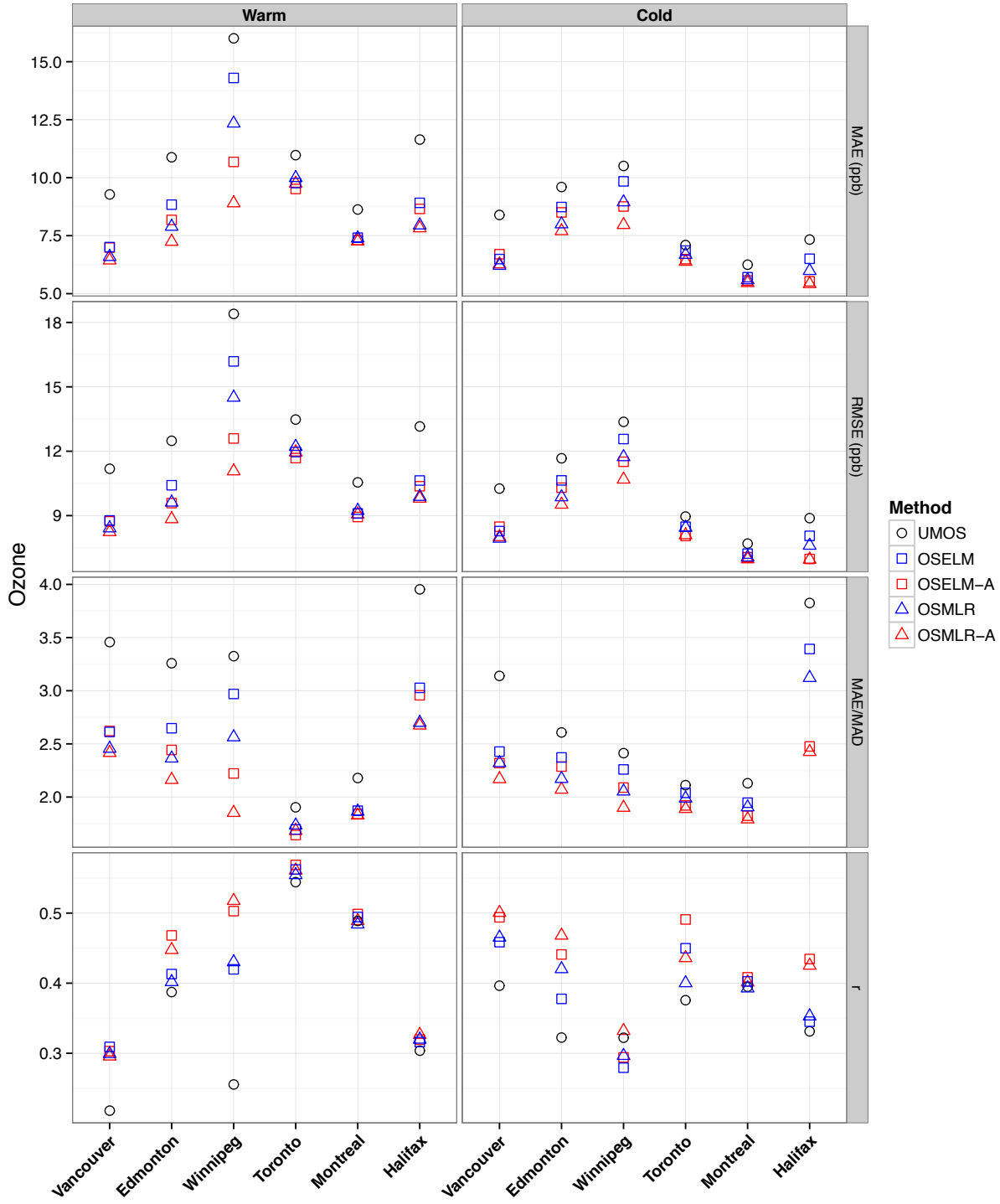



Figure 5.27: Ozone top $10^{th}$ percentile forecast scores from models with and without antecedent predictors by season and station.

For $PM_{2.5}$, Fig 5.28 shows that models with antecedent predictors tended to improve on the original models at all six stations, with smaller MAE/MAD and higher Pearson correlation, mainly in Edmonton, Winnipeg and Halifax. For the MAE skill score (Fig 5.29) at stations Edmonton, Winnipeg and Montreal, all our original models had negative skill scores relative to UMOS for most forecast lead times, but with the antecedent predictors added, OSELM-A is slightly outperforming UMOS for most forecast hours in these three stations, though OSMLR-A is still behind UMOS. In Vancouver, Toronto and Halifax, the new models have similar performance as the original models. For the top $10^{th}$ percentile $PM_{2.5}$ data, Fig 5.30 shows that other than improving the scores in Halifax in the cold season, adding the antecedent predictors brought no clear benefit. In conclusion, models with antecedent predictors added seemed to improve on $PM_{2.5}$ forecasting, but not in forecasting extreme $PM_{2.5}$ concentration.

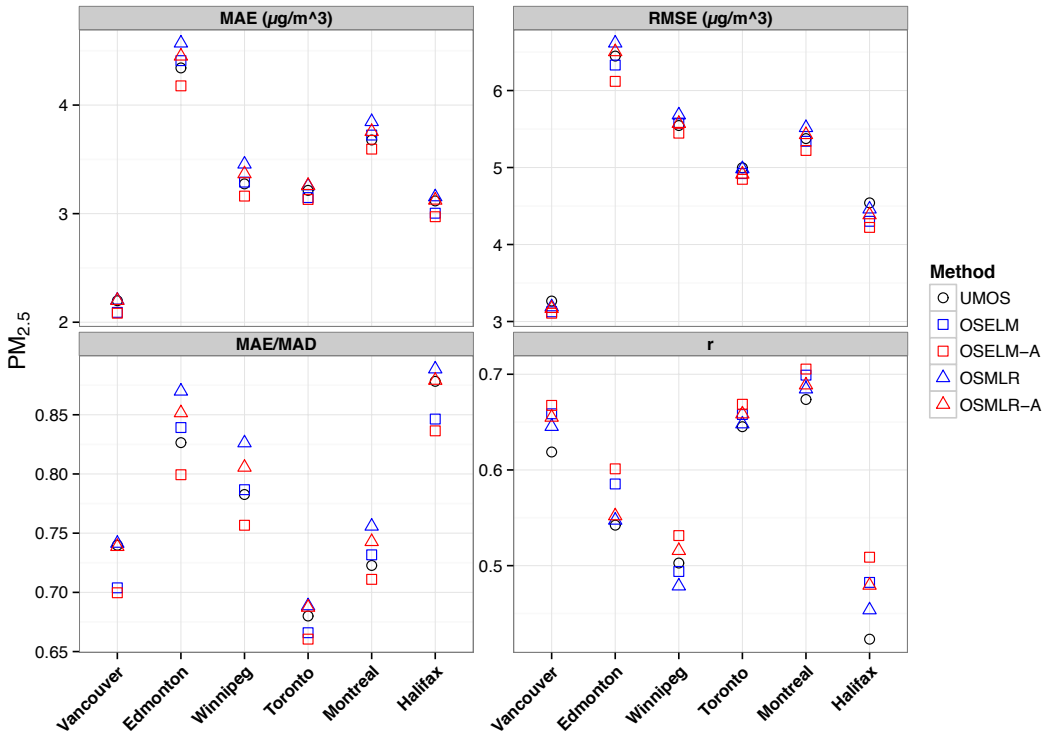



Figure 5.28: $PM_{2.5}$ forecast scores from models with and without antecedent predictors in the six stations.

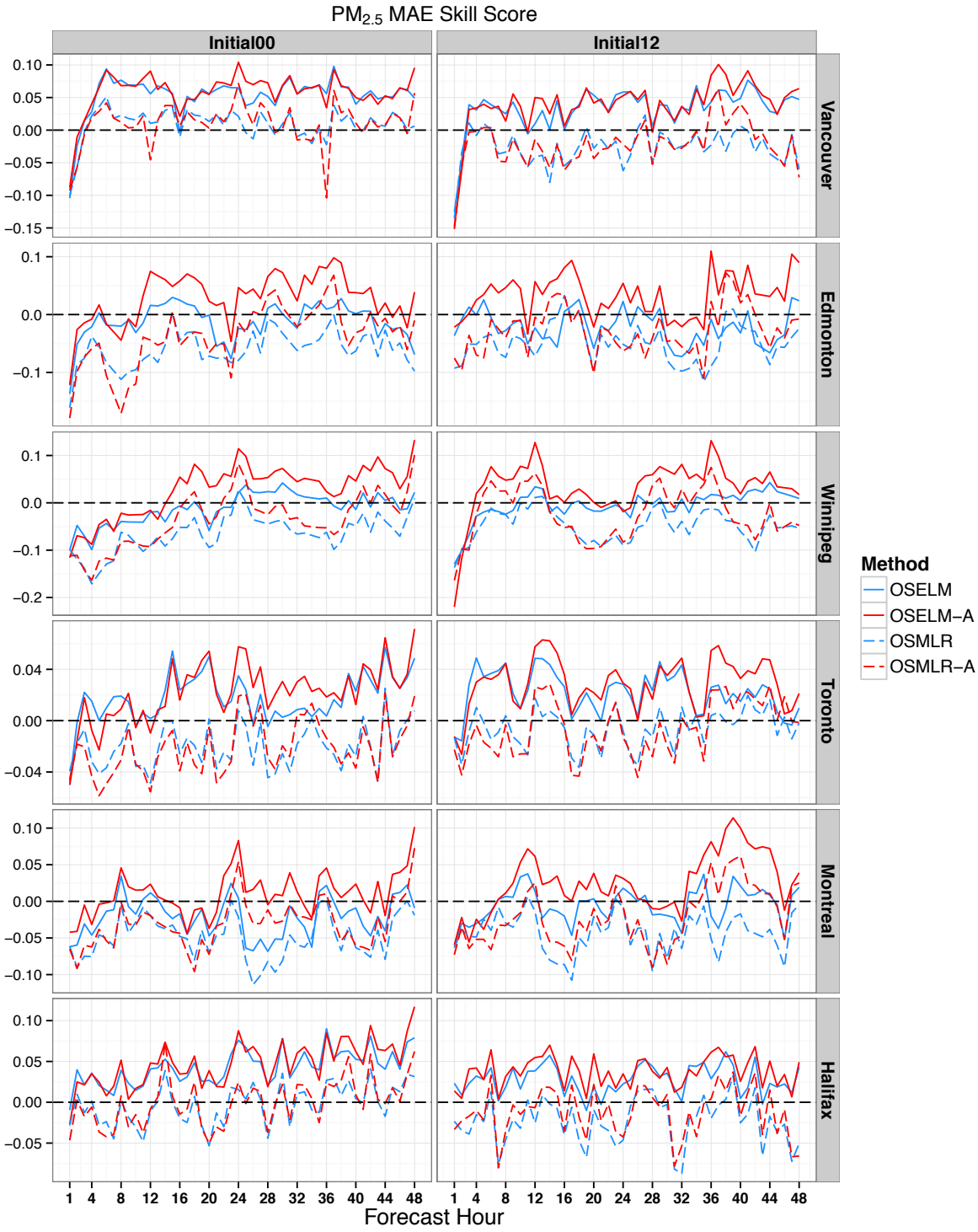



Figure 5.29: $PM_{2.5}$ MAE skill score from models with and without antecedent predictors by forecast hour.

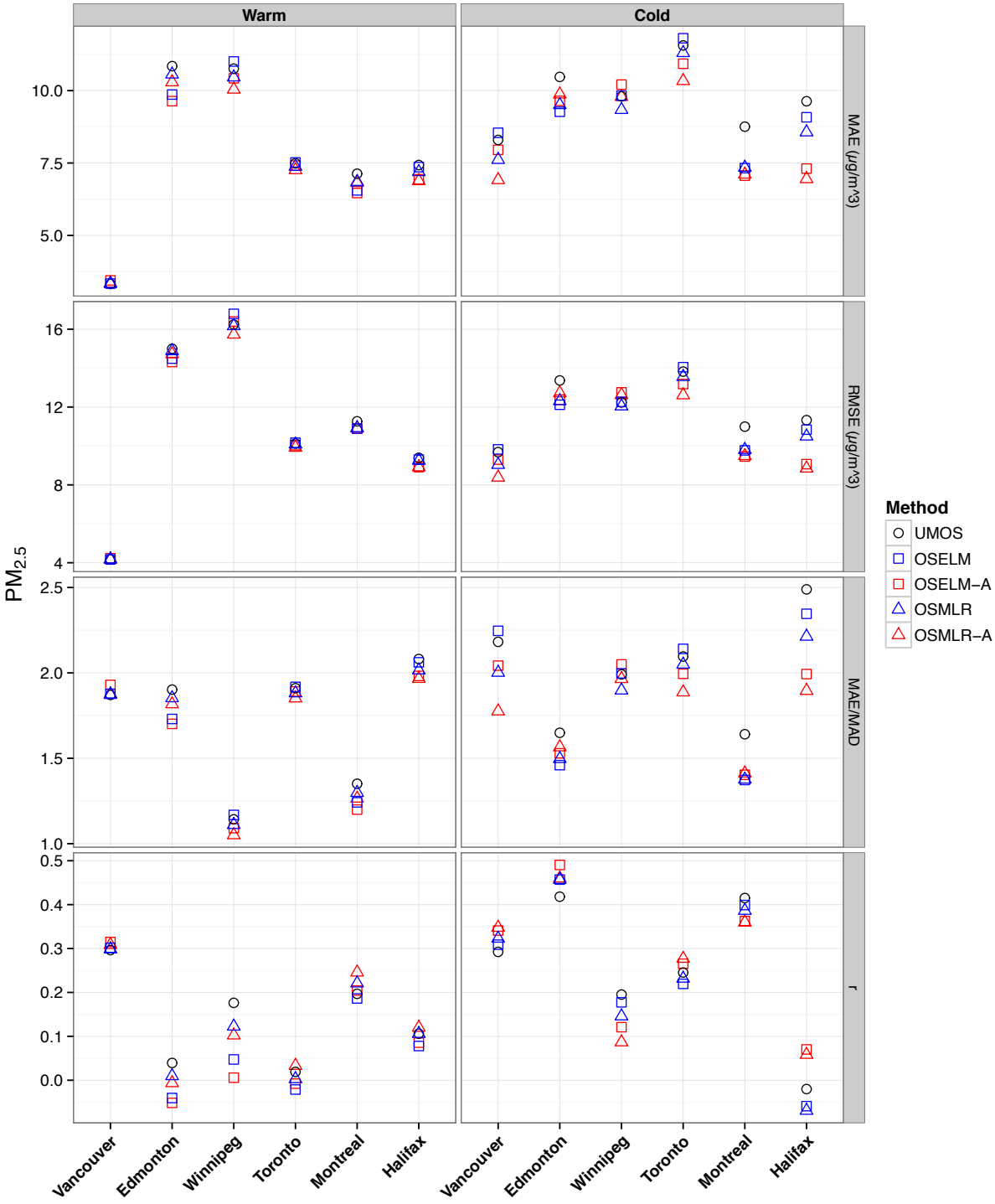



Figure 5.30: $PM_{2.5}$ top $10^{th}$ percentile forecast scores from models with and without antecedent predictors by season and station.

For $NO_2$, Fig 5.31 shows that adding antecedent predictors offered essentially no improvement at all stations except Halifax. The MAE skill score (Fig 5.32) shows clear improvement in Halifax - the skill scores of OSELM and OSMLR were mainly negative over all forecast hours, but have changed to mainly positive in OSELM-A and OSMLR-A. For data in the top $10^{th}$ percentile, adding antecedent predictors mainly helped to improve forecast scores in Halifax, and slightly reduced the errors in the cold season in Vancouver and Montreal.

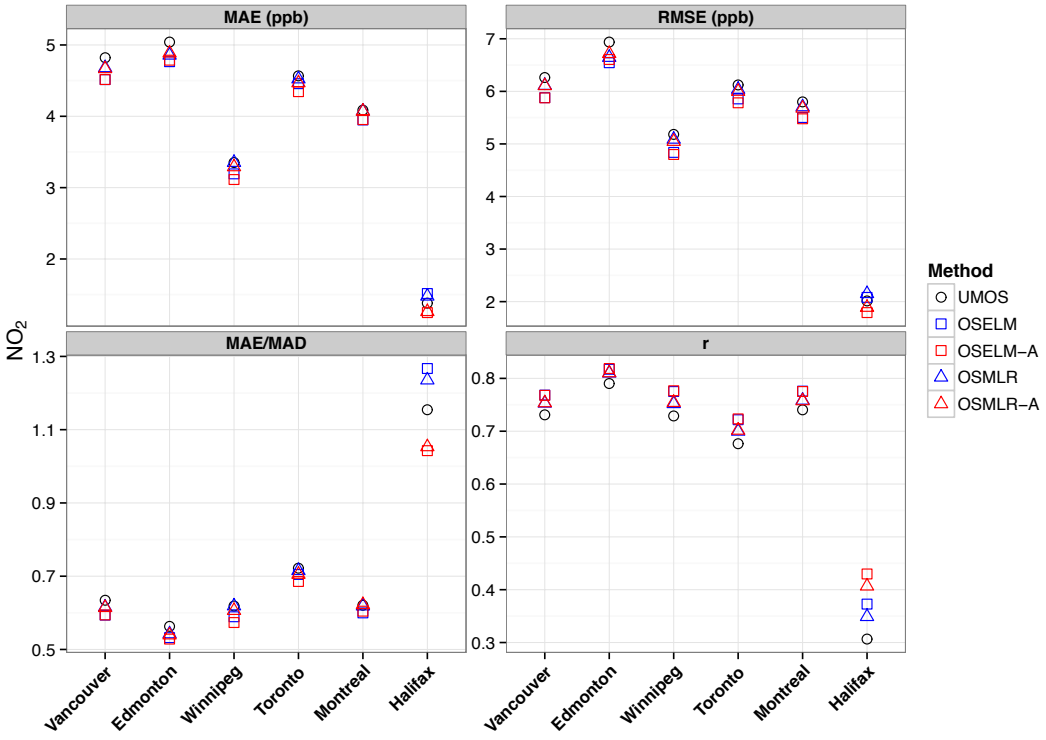



Figure 5.31: $NO_2$ forecast scores from models with and without antecedent predictors in the six stations.

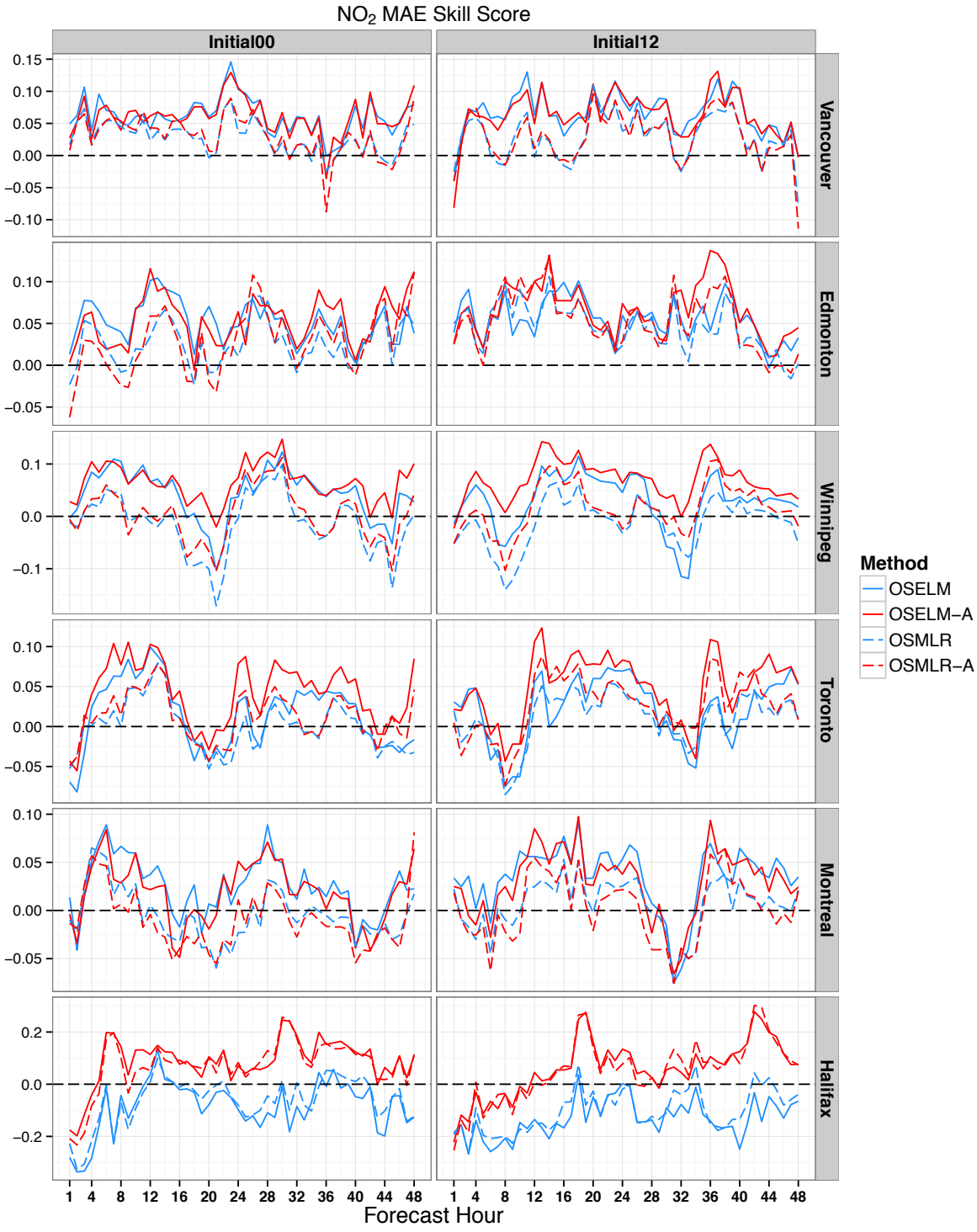



Figure 5.32: $NO_2$ MAE skill score from models with and without antecedent predictors by forecast hour.

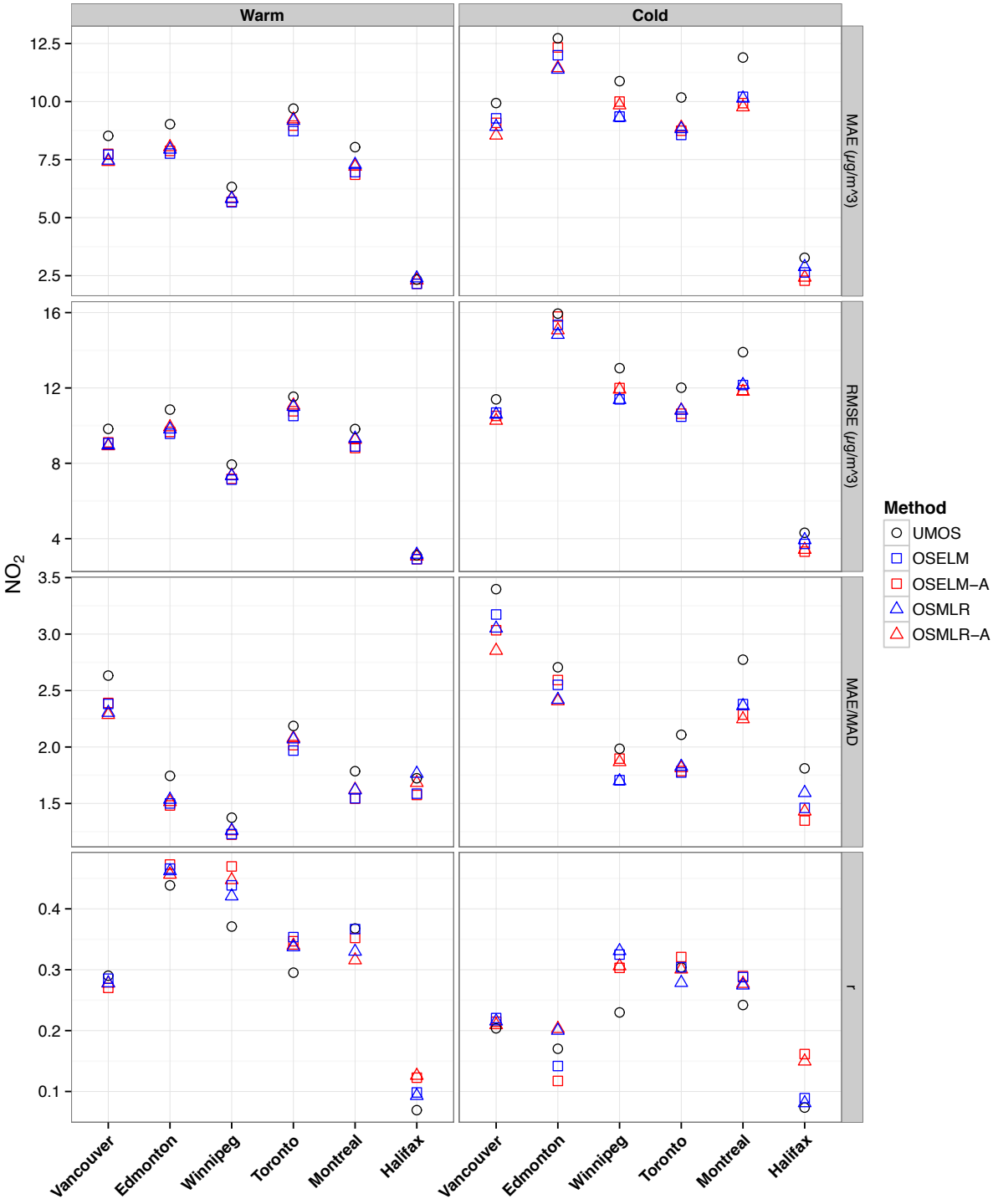



Figure 5.33: $NO_2$ top $10^{th}$ percentile forecast scores from models with and without antecedent predictors by season and station.

# Chapter 6

# Conclusion

## 6.1 Summary

This study focuses on the hourly spot concentration forecasts of ozone ($O_3$), particulate matter 2.5$\mu$m ($PM_{2.5}$) and nitrogen dioxide ($NO_2$) concentration up to 48 hours for six stations across Canada (Vancouver, Edmonton, Winnipeg, Toronto, Montreal and Halifax). In air quality prediction, model accuracy, efficiency and updatability are key considerations. Many current air quality forecasting methods use only linear techniques which would miss nonlinear relationship in the data. In many cases, machine learning methods have been found to outperform linear techniques for air quality prediction. But traditional neural networks have a number of difficulties including computational expense, local minima and over-fitting, which hamper their effectiveness. Consequently, the extreme learning machine (ELM), an updatable machine learning algorithm using randomized neural networks, was applied in air quality forecasting. In this study, air quality forecasting models - the stepwise multiple linear regression (MLR), online-sequential multiple linear regression (OSMLR), multilayer perceptron neural network (MLP NN) and online-sequential extreme learning machine (OSELM) - have been studied using five years of data (2009/07-2014/07). The prediction performances of the MLR, OSMLR, MLP NN and OSELM are evaluated against updatable model output statistics (UMOS) from Environmental Canada.

For ozone, all four models (MLR, OSELM, OSMLR and MLP NN) performed better than UMOS, the benchmark, especially during the high ozone concentration period in the diurnal cycle. OSELM showed the best performance among the various models in all stations except Winnipeg, often outperforming UMOS by more than 10% in the MAE skill score. Linear methods (MLR and OSMLR) also did well in ozone forecasting relative to UMOS, showing the highest MAE skill scores among the models in Winnipeg. For extreme ozone events (top $10^{th}$ percentile) prediction, all models tended to underpredict extreme values, while linear models tended to improve relative to the nonlinear methods when considering only the top $10^{th}$ percentile data.

For $PM_{2.5}$, model correlation scores ranged from 0.5 to 0.7 and all models underperformed UMOS in Edmonton, Winnipeg and Montreal. OSELM was the best model in Vancouver, Toronto and Halifax, surpassing UMOS and the other models for most forecast lead times. For extreme $PM_{2.5}$ events, all of the relative errors (MAE/MAD) were above 1, indicating weak model performance, and nonlinear models still in general did not improve on the linear models.

For $NO_2$, OSELM was marginally better than all the other methods in all stations except Halifax. In Halifax, all four methods (MLR, OSELM, OSMLR and MLP NN) lost to UMOS in the MAE, as UMOS tended to forecast slightly better during the low $NO_2$ hours in the diurnal cycle. For the top $10^{th}$ percentile $NO_2$ observations, OSELM, MLR, OSMLR and MLP NN tended to slightly outperform UMOS at all stations in both warm and cold seasons.

Antecedent predictors have different effects on three different pollutants. Adding the antecedent predictors tended to improve on the ozone forecasts, especially in predicting extreme values. For $PM_{2.5}$, models with antecedent predictors had improved performance compared to the original models, but not in forecasting extreme $PM_{2.5}$ concentration. For $NO_2$, adding antecedent predictors offered improvement only at Halifax, but also slightly reduced the errors for extreme $NO_2$ in the cold season in Vancouver and Montreal.

This study has demonstrated the potential of using nonlinear machine learning methods to improve air quality forecasts. In terms of improving forecast accuracy, OSELM appeared most beneficial in ozone forecast and least in $PM_{2.5}$ forecast. MLP NN when updated with new data only seasonally due to the large computational cost generally underperformed the daily updated linear methods (MLR and OSMLR). In contrast, OSELM, with its low cost updating and nonlinear modeling capability, generally outperformed the linear methods in forecast accuracy. Antecedent predictors could also be added to the models to improve forecast accuracy.

## 6.2 Future Research

A limitation of this research was that models were not developed separately by season because of the limited data record. When more data are available, it could be worth investigating using separate models for different seasons. Another limitation is the model structure of OSELM and MLP NN methods. The number of hidden nodes were decided according to the initial training data and did not change in subsequent model updates because of limiting computational resources, resulting in fixed model complexity. As new data arrive, information on larger term variability, e.g. interannual or interdecadal variability, becomes available, but the fixed model complexity would not have the capacity to learn the additional structure in the data. If model complexity can be changed during the updating process, this may enhance the prediction skills.

# Bibliography

Anselmo, D., Moran, M. D., Ménard, S., Bouchet, V. S., Makar, P. A., Gong, W., Kallaur, A., Beaulieu, P., Landry, H., Stroud, C., Huang, P., Gong, S., and Talbot, D. (2010). A new Canadian air quality forecast model: GEM-MACH15. In *12th Conference on Atmospheric Chemistry*, Atlanta, GA. American Meteorological Society, Boston, 6 pp.

Antonopoulos, S., Bourgouin, P., Montpetit, J., , and Croteau, G. (2012). Forecasting O3, PM25 and NO2 hourly spot concentrations using an updatable mos methodology. In Steyn, D. G. and Castelli, S. T., editors, *Air Pollution Modeling and its Application XXI*, chapter 53. NATO Science for Peace and Security Series C: Environmental Security.

Ballester, E. B., Valls, G. C., Carrasco-Rodriguez, J. L., Olivas, E. S., and Valle-Tascon, S. (2002). Effective 1-day ahead prediction of hourly surface ozone concentrations in eastern Spain using linear models and neural networks. *Ecological Modeling*, 156(1):27–41.

Banerjee, T. and Srivastava, R. K. (2009). Evaluation of environmental impacts of IIE-Pantnagar through application of air and water quality indices. *Environmental Monitoring and Assessment*, 172(1-4):547–560.

Basheer, A. and Najjar, M. (1996). Predicting dynamic response of adsorption columns with neural nets. *Journal of Computing in Civil Engineering*, 10(1):211–220.

Benson, P. E. (1984). CALINE4—a dispersion model for predicting air pollutant concentrations near roadways. Technical Report FHWA/CA/TL-84/15, FHWA User Guide, Trinity Consultants Inc., USA.

Bloomfield, P., Royle, J. A., Steinberg, L. J., and Yang, Q. (1996). Accounting for meteorological effects in measuring urban ozone levels and trends. *Atmospheric Environment*, 30:3067–3077.

Boznar, M., Lesjak, M., and Mlakar, P. (1993). A neural-network-based method for short-term predictions of ambient SO2 concentration in highly polluted industrial-areas of complex terrain. *Atmospheric Environment Part B-Urban Atmosphere*, 27(2):221–230.

Brauer, M. and Brook, J. R. (1997). Ozone personal exposures and health effects for selected groups residing in the Fraser Valley. *Atmospheric Environment*, 31(14):2113–2121.

Breiman, L. (1996). Bagging predictions. *Machine Learning*, 24:123–140.

Cannon, A. J. (2012a). monmlp: Monotone multi-layer perceptron neural network. R package version 1.1.2. https://cran.r-project.org/web/packages/monmlp/index.html.

Cannon, A. J. (2012b). Neural networks for probabilistic environmental prediction: Conditional density estimation network creation and evaluation (CaDENCE) in R. *Computers and Geosciences*, 41:126–135.

Cavazos, T. (1997). Downscaling large-scale circulation to local winter rainfall in north-eastern Mexico. *International Journal of Climatology*, 17(10):1069–1082.

Chaloulakou, A., Assimakopoulos, D., and Kekkas, T. (1999). Forecasting daily maximum ozone concentrations in the Athens basin. *Environmental Monitoring and Assessment*, 56:97–112.

Chen, H. and Copes, R. (2013). *Review of Air Quality Index and Air Quality Health Index*. Ontario Agency for Health Protection and Promotion (Public Health Ontario), Toronto, ON: Queen's Printer for Ontario.

Chen, S. T., Yu, P. S., and Tang, Y. H. (2010). Statistical downscaling of daily precipitation using support vector machines and multivariate analysis. *Journal of Hydrology*, 385:13–22.

Chen, T. M., Gokhale, J., Shofer, S., and Kuschner, W. G. (2007). Outdoor air pollution: ozone health effects. *The American journal of the medical sciences*, 333(4):244–248.

Colvile, R., Woodfield, N., Carruthers, D., Fisher, B., Rickard, A., Neville, S., and Hughes, A. (2002). Uncertainty in dispersion modeling and urban air quality mapping. *Environmental Science and Policy*, 5(3):207–220.

Comrie, A. (1997). Comparing neural networks and regression models for ozone forecasting. *Journal of Air and Waste Management*, 47:653–663.

Demuzere, M., Trigo, R. M., Arellano, J. Vila-Guerau de, and van Lipzig, N. P. M. (2009). The impact of weather and atmospheric circulation on O3 and PM10 levels at a rural mid-latitude site. *Chemical Physics*, 9(8):2695–2714.

Dennis, R. L., Byun, D. W., Novak, J. H., Galluppi, K. J., Coats, C. J., and Vouk, M. A. (1996). The next generation of integrated air quality models: EPA's models-3. *Atmospheric Environment*, 30(12):1925–1938.

Dominick, D., Latif, M. T., Juahir, H., Aris, A. Z., and Zain, S. M. (2012). An assessment of influence of meteorological factors on PM10 and NO2 at selected stations in Malaysia. *Sustainable Environment Research*, 22(5):305–315.

EPA (2005). Six common air pollutants. U. S. environmental protection agency. http://www.epa.gov/air/urbanair/6 poll.html.

Feng, G., Huang, G. B., Lin, Q., and R., G. (2009). Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Trans Neural Netw*, 20(8):1352–1357.

Flemming, J., Reimer, E., and Stem, R. (2001). Long term evaluation of the ozone forecast by an Eulerian model. *Physics and Chemistry of the Earth*, 26:775–779.

Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron). *Atmospheric Environment*, 32:2627–2636.

Gardner, M. W. and Dorling, S. R. (1999). Neural network modeling and prediction of hourly NOx and NO2 concentrations in urban air in london. *Atmospheric Environment*, 33:709–719.

Gardner, M. W. and Dorling, S. R. (2000). Statistical surface ozone models: an improved methodology to account for non-linear behavior. *Atmospheric Environment*, 34(1):21–34.

Golub, G. H., Heath, M., and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223.

Haber, E. and Oldenburg, D. (2000). A GCV based method for nonlinear ill-posed problems. *Computational Geoscience*, 4:41–63.

Hadjiiski, L. and Hopke, P. K. (2000). Application of artificial neural network to modeling and prediction of ambient ozone concentrations. *Journal of Air and Waste Management Association*, 50:894–901.

Handoko, S. D., Keong, K. C., Soon, O. Y., Zhang, G. L., and Brusic, V. (2006). Extreme learning machine for predicting HLA-Peptide binding. *Lecture Notes in Computer Science*, 3973:716–721.

Hansen, D. A., Dennis, R. L., Ebel, A., Hanna, S. R., Kaye, J., and Thuillier, R. (28). The quest for an advanced regional air quality model. *Environmental Science and Technology*, 2:71–77.

Heeswijk, M. V., Miche, Y., Lindh-Knuutila, T., Hilbers, P. A., Honkela, T., Oja, E., and Lendasse, A. (2009). Adaptive ensemble models of extreme learning machines for time series prediction. *Lecture Notes in Computer Science*, 5769:305–314.

Hewitson, B. C. and Crane, R. G. (1996). Climate downscaling: Techniques and application. *Climate Research*, 7:85–95.

Holmes, N. S. and Morawska, L. (2006). A review of dispersion modeling and its application to the dispersion of particles: An overview of different dispersion models available. *Atmospheric Environment*, 40(30):5902–5928.

Hsieh, W. W. (2009). *Machine Learning Methods in the Environmental Sciences: Neural Networks and Kernels*. Cambridge University Press.

Hsieh, W. W. and Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, 79:1855–1870.

Huang, G. B. and Chen, L. (2007). Convex incremental extreme learning machine. *Neurocomputing*, 70:3056–3062.

Huang, G. B. and Chen, L. (2008). Enhanced random search based incremental extreme learning machine. *Neurocomputing*, 71(3460-3468).

Huang, G. B., Chen, L., and Siew, C. K. (2006a). Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Transactions on Neural Networks*, 17(4):879–892.

Huang, G. B., Wang, D., and Lan, Y. (2011). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, 2:107–122.

Huang, G. B., Zhu, Q. Y., and Siew, C. K. (2006b). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1-3):489–501.

Kim, K.-H., Jahan, S. A., and Kabir, E. (2013). A review on human health perspective of air pollution with respect to allergies and asthma. *Environment International*, 59:41–52.

Kolehmainen, M., Martikainen, H., and Ruuskanen, J. (2001). Neural networks and periodic components used in air quality forecasting. *Atmospheric Environment*, 35(5):815–825.

Kukkonen, J., Partanen, L., Karppinen, A., Ruuskanen, J., Junninen, H., Kolehmainen, M., Niska, H., Dorling, S., 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 modeling system and measurements in central Helsinki. *Atmospheric Environment*, 37(32):4549–4550.

Kumar, A. and Goyal, P. (2013). Forecasting of air quality in Delhi using principal component regression technique. *Pure and Applied Geophysics*, 170:711–722.

Kurt, A. and Oktay, A, B. (2010). Forecasting air pollutant indicator levels with geographic models 3 days in advance using neural networks. *Expert systems with Application*, 37:7986–7992.

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

Liang, N. Y., Huang, G. B., Saratchandran, P., and Sundararajan, N. (2006). A fast and accurate on-line sequential learning algorithm for feedforward networks. *IEEE Transactions on Neural Networks*, 17(6):1411–1423.

Lima, A. R., Cannon, A. J., and Hsieh, W. W. (2015). Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. *Environmental Modelling and Software*, 73:175–188.

Lu, W. Z. and Wang, W. J. (2005). Potential assessment of the 'support vector' machine method in forecasting ambient air pollution trends. *Chemosphere*, 59(5):693–701.

Luecken, D. J., Hutzell, W. T., and Gipson, G. L. (2006). Development and analysis of air quality modeling simulations for hazardous air pollutants. *Atmospheric Environment*, 40(26):5087–5096.

Mahfouf, J.-F., Brasnett, B., and Gagnon, S. (2007). A Canadian precipitation analysis (CaPA) project: description and preliminary results. *Atmosphere-Ocean*, 45:1–17.

Maier, H. R., Jain, A., Dandy, G. C., and Sudheer, K. (2010). Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environmental Modeling and Software*, 25:891–909.

Marzban, C. and Stumpf, G. (1996). A neural network for tornado prediction based on Doppler radar-derived attributes. *Journal of Applied Meteorology*, 35:617–626.

McGranahan, G. and Murray, F., editors (2003). *Air Pollution and Health in Rapidly Developing Countries*. Earthscan Publications Ltd.

Milford, J. B., Russell, A. G., and McRae, G. J. (1989). A new approach to photochemical pollution control: Implications of spatial patterns in pollutant responses to reductions in nitrogen oxides and reactive organic gas emissions. *Environmental Science and Technology*, 23(10):1290–1301.

Moran, M. D., S. Ménard, R. P., Anselmo, D., Antonopoulos, S., Makar, P. A., Gong, W., Gravel, S., Stroud, C., Zhang, J., Zheng, Q., Robichaud, A., Landry, H., Beaulieu, P., Gilbert, S., Chen, J., and Kallaur, A. (2014). Recent advances in canada's national operational aq forecasting system. In Steyn, D. G., Builtjes, P. J. H., and Timmermans, R. M. A., editors, *Air Pollution Modeling and its Application XXII*, chapter 37. NATO Science for Peace and Security Series C: Environmental Security.

Nejadkoorki, F. and Baroutian, S. (2012). Forecasting extreme PM10 concentrations using artificial neural networks. *International Journal of Environmental Research*, 6:277–284.

Nguyen, D. (2014). A brief review of air quality models and their applications. *Open Journal of Atmospheric and Climate Change*, 1(2):60–80.

Nordiska, M., editor (2008). *Interaction between climate change, air pollution and related impacts*. Nordic Council of Ministers' publishing house.

Oke, T. R., Cleugh, H. A., Grimmond, S., Schmid, H. P., and Roth, M. (1989). Evaluation of spatially averaged fluxes of heat, mass and momentum in the urban boundary layer. *Weather Climate*, 9:14–21.

Parviainen, E. and Riihimaki, J. (2013). A connection between extreme learning machine and neural network kernel. *Knowledge Discovery, Knowledge Engineering and Knowledge Management Communications in Computer and Information Science*, 272:122–135.

Perez, P. (2001). Prediction of sulfur dioxide concentrations at a site near downtown Santiago, Chile. *Atmospheric Environment*, 35:4929–4935.

Perez, P., Trier, A., and Reyes, J. (2000). Prediction of PM2.5 concentrations several hours in advance using neural networks in Santiago, Chile. *Atmospheric Environment*, 34:1189–1196.

Prybutok, V. R., Yi, J., and Mitchell, D. (2000). Comparison of neural network model with ARIMA and regression models for prediction of Houston's daily maximum ozone concentrations. *European Journal of Operational Research*, 122:31–40.

R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0.

Ramanathan, V. and Feng, Y. (2009). Air pollution, greenhouse gases and climate change: Global and regional perspectives. *Atmospheric Environment*, 43:37–50.

Rasouli, K., Hsieh, W. W., and Cannon, A. J. (2012). Daily stream flow forecasting by machine learning methods with weather and climate inputs. *Journal of Hydrology*, 414:284–293.

Reich, S. L., Gomez, D. R., and Dawidowski, L. E. (1999). Artificial neural network for the identification of unknown air pollution sources. *Atmospheric Environment*, 33(18):3045–3052.

Revlett, G. (1978). Ozone forecasting using empirical modeling. *Journal of Air Pollution Control Association*, 28:338–343.

Roadknight, C. M., Balls, G. R., Mills, G. E., and Palmer, B. D. (1997). Modeling complex environmental data. *IEEE Transactions on Neural Networks*, 8(4):852–861.

Romero, E. and Alquẽzar, R. (2012). Comparing error minimized extreme learning machines and support vector sequential feed-forward neural networks. *Neural Networks*, 25:122–129.

Rong, H. J., Ong, Y. S., Tan, A. H., and Zhu, Z. (2008). A fast pruned-extreme learning machine for classification problem. *Neurocomputing*, 72:359–366.

Roth, M. (2000). Review of atmospheric turbulence over cities. *Quarterly Journal of the Royal Meteorological Society*, 126(564):942–990.

Russell, A. and Dennis, R. (2000). Narsto critical review of photochemical models and modeling. *Atmospheric Environment*, 34(12):2283–2324.

Schmidt, W. F., Kraaijveld, M. A., and Duin, R. P. W. (1992). Feed forward neural networks with random weights. In *11th IAPR International Conference on Pattern Recognition, Proceedings, Vol Ii: Conference B: Pattern Recognition Methodology and Systems*, pages 1–4.

Seinfeld, J. H. and Pandis, S. N. (1997). *Atmospheric Chemistry and Physics from Air Pollution to Climate Change*. Wiley-Interscience.

Song, X. H. and Hopke, P. K. (1996). Solving the chemical mass balance problem using an artificial neural network. *Environmental Science and Technology*, 30(2):531–535.

Sun, Z. L., Choi, T. M., Au, K. F., and Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1):411–419.

Tai, P. K. A. P. K. (2012). *Impact of Climate Change on Fine Particulate Matter Air Quality*. PhD thesis, Harvard University.

Thimm, G. and Fiesler, E. (1997). High-order and multilayer perceptron initialization. *IEEE Transactions on Neural Networks*, 8:349–359.

Thirumalaiah, K. and Deo, M. C. (1998). River stage forecasting using artificial neural networks. *Journal of Hydrologic Engineering*, 3(1):26–32.

Tripathi, S., Srinivas, V., and Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, 330:621–640.

Wallace, J. and Kanaroglou, P. (2008). Modeling NOx and NO2 emissions from mobile sources: A case study for Hamilton, Ontario, Canada. *Transportation Research Part D: Transport and Environment*, 13(5):323–333.

Walter, A., Denhard, M., and Schonwiese, C.-D. (1998). Simulation of global and hemispheric temperature variations and signal detection studies using neural networks. *Meteorologische Zeitschrift*, N.F.7:171–180.

Wang, W., Xu, Z., and Lu, J. W. (2003). Three improved neural network models for air quality forecasting. *Engineering Computations*, 20(2):192–210.

Wilson, L. J. and Vallée, M. (2002). The Canadian updateable model output statistics (UMOS) system: design and development tests. *Weather Forecast*, 17:206–222.

Wilson, L. J. and Vallée, M. (2003). The Canadian updateable model output statistics (UMOS) system: validation against perfect prog. *Weather Forecast*, 18:288–302.

Wolff, G. T. and Lioy, P. J. (1978). An empirical model for forecasting maximum daily ozone levels in the northeastern U.S. *Journal of Air Pollution Control Association*, 28:5087–5096.

Wotawa, G., Stohl, A., and Neininger, B. (1998). The urban plume of vienna: Comparisons between aircraft measurements and photochemical model results. *Atmospheric Environment*, 32:2479–2489.

Yeu, C. T., Lim, M. H., Huang, G. B., Agarwal, A., and Ong, Y. S. (2006). A new machine learning paradigm for terrain reconstruction. *IEEE Geoscience and Remote Sensing Letters*, 3(3):382–386.

Yi, J. and Prybutok, V. R. (2002). A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution*, 92(3):349–357.

Yura, E. A., Kear, T., and Niemeier, D. (2007). Using CALINE dispersion to assess vehicular PM2.5 emissions. *Atmospheric Environment*, 41(38):8747–8757.

Yuval (2000). Neural network training for prediction of climatological time series; regularized by minimization of the generalized cross validation function. *Monthly Weather Review*, 128:1456–1473.

Yuval (2001). Enhancement and error estimation of neural network prediction of El Niño - 3.4 SST anomalies. *Journal of Climate*, 14:2150–2163.