University of Alberta

ANN Modeling of Ambient $PM_{2.5}$ in Fort McKay, Alberta

by

Yaming He ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Environmental Engineering

Department of Civil and Environmental Engineering

Edmonton, Alberta

Fall 2004

Your file Votre référence
ISBN: 0-612-95762-4
Our file Notre référence
ISBN: 0-612-95762-4

Acknowledgements

I would first like to thank Dr. Warren Kindzierski, my supervisor for his inspiration, sound guidance and patience on explaining things throughout the process of this research.

I appreciate Dr. Mohamed Gamal El–Din. I learned a lot from his course about artificial neural networks.

I would also like to thank Alex Mackenzie of Alberta Health and wellness in Edmonton for providing funding, Brian Wiens of Environment Canada in Edmonton for providing meteorological data, Wood Buffalo Environmental Association in Fort McMurray for providing air quality data, and Alberta Sustainable Resource Development in Edmonton for providing forest fire data.

I wish to thank my student colleagues in environmental engineering group. They are always helpful and warm-hearted.

Finally, I would like to thank my mother, Shuhuai Zheng, and my father, Junhua He. Their moral support is always an invaluable treasure in my life.

Table of Contents

1.0 INTRODUCTION

1.1 PROJECT BACKGROUND 1

1.2 STUDY OBJECTIVES 4

2.0 LITERATURE REVIEW *6*

2.1 AIRBORNE PARTICULATE MATTER 6

2.1.1 $PM_{2.5}$ 6
2.1.2 Relationship of $PM_{2.5}$ and PM_{10} 8
2.1.3 $PM_{2.5}$ Environmental and Health Effects 9
2.1.4 Guidelines and Standards for $PM_{2.5}$ 12
2.1.5 $PM_{2.5}$ Physical and Chemical Characteristics 14
2.1.6 Distribution and Dispersion 16
2.1.7 Air Quality Monitoring 21

2.2 ARTIFICIAL NEURAL NETWORK (ANN) 22

2.2.1 Basic Model Structure 23
2.2.2 Learning Process 24
2.2.3 Model Development 25
2.2.4 Comparison with Other Models 25
2.2.5 Applications with $PM_{2.5}$ 28
2.2.6 Advantages and Limitations of ANN 29

3.0 ANN MODELING OF $PM_{2.5}$ IN FORT MCKAY *31*

3.1 SOURCE DATA PRE-PROCESSING 31

3.1.1 Statistical Analysis 33
3.1.2 Data Screening and Re-formatting 40

3.2 MODEL DEVELOPMENT 45

3.2.1 Input and Output Parameter Selection 45
3.2.2 Model Evaluation Criteria 49
3.2.3 Model Development Step One 52
3.2.4 Model Development Step two 60

4.0 MODEL CHECK AND APPLICATION *72*

4.1 MODEL STABILITY AND OVERALL FIT 72

4.1.1 Model Stability 72
4.1.2 Overall Model Fit 73

4.2 SENSITIVITY ANALYSIS 78

5.0 FINDINGS AND RECOMMENDATIONS *92*

REFERENCES *97*

APPENDIX A – DISTANCE BETWEEN AMS #1 AND AMS #13 .... *104*

APPENDIX B – SUMMARY STATISTICS FOR PARAMETERS USED IN ANN MODELING *106*

APPENDIX C – TIME SERIES PLOT FOR PARAMETERS USED IN ANN MODELING 110

APPENDIX D –HOURLY $PM_{2.5}$ PREDICTION AFTER $PM_{2.5}$ CONCENTRATION AT AMS #13 INCREASED BY 25%.. 137

List of Tables

Table 1.	*Estimated PM emissions in Canada in 1990*	8
Table 2.	*Site-specific relationships between PM_{10} and $PM_{2.5}$ for samples from year 1984 to 1993.*	10
Table 3.	*Comparison between ANN and multi-regression model for a water treatment application.*	27
Table 4.	*Explanation for flag numbers in original data files for AMS #1 and AMS #13 parameters.*	40
Table 5.	*Number of invalid data with negative value and labeled by flag number at the Fort McKay air monitoring station (AMS #1).*	42
Table 6.	*Number of invalid data with negative value and labeled by flag number at AMS #13*	43
Table 7.	*Number of invalid data (with value "-99999") at Fort McMurray airport meteorological monitoring station*	44
Table 8.	*Computed R^2 values and relative factor weightings for ANN models constructed*	54
Table 9.	*Different bonds of weight update methods and data pattern selection methods for ANN modeling.*	63
Table 10.	*R^2 values for different activation function settings in ANN sub-model using category #1 data set.*	67
Table 11.	*R^2 values for different activation function settings in ANN sub-model using category #2 data set.*	68
Table 12.	*Comparison of R^2 for optimum ANN sub-models after training with original and exchanged production data and testing data sets.*	73

List of Figures

Figure 1. *Location of Fort McKay, Alberta.* ... 3
Figure 2. *PM_{10} and $PM_{2.5}$ size distribution for ambient particles.* ... 9
Figure 3. *Chemical compound distribution corresponding with particle size.* ... 15
Figure 4. *Relationships between lapse rate and air stability.* ... 17
Figure 5. *Gaussian dispersion of a point source located at (0, 0, H) in a wind-oriented coordinate system.* ... 20
Figure 6. *Basic structure for a multilayer perceptron ANN.* ... 24
Figure 7. *Result from the ANN model - Actual vs. Prediction for a water treatment application.* ... 27
Figure 8. *Result from the linear regression model- Actual vs. Prediction for a water treatment application.* ... 28
Figure 9. *Position of WBEA air quality monitoring stations* ... 34
Figure 10. *Forest fire generating Particulate matter and various transport pathways.* ... 44
Figure 11. *R^2 values versus data pattern selection and weight update method settings for ANN sub-model using category #1 data set.* ... 64
Figure 12. *R^2 values versus data pattern selection and weight update method settings for ANN sub-model using category #2 data set.* ... 64
Figure 13. *R^2 values versus number of hidden layer neurons for ANN sub-model using category #1 data set.* ... 69
Figure 14. *R^2 values versus number of hidden layer neurons for ANN sub-model using category #2 data set.* ... 69
Figure 15. *R^2 values versus learning epochs for ANN sub-model using category #1 data set.* ... 70
Figure 16. *R^2 values versus learning epochs for ANN sub-model using category #2 data set.* ... 70
Figure 17. *Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [$\mu g/m^3$] for ANN sub-model using category #1 data (205 data patterns) corresponding to typical range of fluctuation observed in $PM_{2.5}$ levels.* ... 74
Figure 18 *Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [$\mu g/m^3$] for ANN sub-model using category #1 data (205 data patterns) corresponding to a range encompassing peak concentration events.* ... 75
Figure 19 *Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [$\mu g/m^3$] for ANN model using category #2 data (205 data patterns) corresponding to typical range of fluctuation observed in $PM_{2.5}$ levels.* ... 76
Figure 20 *Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [$\mu g/m^3$] for ANN model using category #2 data (205 data patterns) corresponding to a range encompassing peak concentration events.* ... 77
Figure 21 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 25% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.* ... 80
Figure 22 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.* ... 80

Figure 23 *$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.* 81
Figure 24 *$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.* 82
Figure 25 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 25% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.* 83
Figure 26 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.* 83
Figure 27 *$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.* 84
Figure 28 *$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.* 85
Figure 29 *WBEA Fort McKay Station showing 60° angles used for examination of wind direction and corresponding $PM_{2.5}$ concentrations.* 86
Figure 30 *Location of oil sand mining leasehold areas around Fort McKay* 88
Figure 31 *Cumulative distribution of hourly average $PM_{2.5}$ concentrations at Fort McKay station.* 89
Figure 32 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 25% and 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13 and for winds originating from three 60°-angles ranges.* 90
Figure 33 *Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 25% and 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13 and for winds originating from three 60°-angles ranges.* 91

List of Abbreviations

AEP	Alberta Environmental Protection
AMS	Air Monitoring Station
ANN	Artificial Neural Network
CASA	Clean Air Strategic Alliance
CCME	Canadian Council of Ministers of the Environment
CEPA	Canadian Environmental Protection Act
CWS	Canada-Wide Standard
EC	Element Carbon
FPAC	Federal-Provincial Advisory Committee
FRM	Federal Reference Method
MLP	Multi-Layer Perceptron
MMD	Maximum Mixing Depth
PM	Particulate Matter
$PM_{2.5}$	Particulate Matter with an aerodynamic diameter smaller than 2.5 microns
PM_{10}	Particulate Matter with an aerodynamic diameter smaller than 10 microns
RERI	Real Estate Research Institute
SAGD	Steam Assisted Gravity Drainage
SOM	Self-Organization Map
TEOM	Tapered Element Oscillating Microbalance
US EPA	United States Environmental Protection Agency
WBEA	Wood Buffalo Environmental Association

1.0 INTRODUCTION

This thesis describes a modeling study of airborne particulate matter (PM) using an Artificial Neural Network (ANN). This first section of the thesis introduces the study and provides background information and research objectives. The second section presents a brief literature review of particulate matter with an aerodynamic diameter less than 2.5 microns ($PM_{2.5}$) and ANN. The third section presents details about ANN model construction and prediction of ambient $PM_{2.5}$ concentrations in the community of Fort McKay, Alberta. The fourth section describes methods for identifying the best ANN model architecture to use in a sensitivity analysis. The final section presents conclusions and recommendations of the study.

1.1 PROJECT BACKGROUND

Good air quality is considered important in society. $PM_{2.5}$ is an urban air pollutant that has been linked to health problems in humans and an ability to impair ecological systems. The makeup of $PM_{2.5}$ in urban air can be very complex. There are multiple point and non-point emission sources in cities that can contribute to airborne PM concentrations. Dispersion and movement of $PM_{2.5}$ in an airshed is affected by meteorology, objects such as buildings, and terrain. Therefore, modeling of ambient $PM_{2.5}$ can be difficult using conventional Gaussian dispersion models that depend on mass balance principles and require site-specific information on emissions, meteorology, and terrain. Typically, the presence of objects and/or buildings and irregular terrain that may impede air movement are not handled well in these models.

Artificial Neural Network (ANN) is not constrained by these types of limitations and offers an opportunity for modeling complex pollutant behaviours – such as those associated with $PM_{2.5}$. ANN is applicable for solving non-linear and complex problems, which are difficult and expensive to solve using conventional modeling methods. ANN is considered a semi-black box method and does not require knowledge of relationships and/or equations describing emissions, dispersion, meteorology, terrain, etc. Only historical data on input and output parameters and overall knowledge about air pollutant behaviours in an airshed are required by an ANN system. ANN will learn patterns from these historical data and construct a relationship between the input and output parameters.

The target location for the ANN study was Fort McKay, Alberta. Fort McKay is a small community on the west bank of Athabasca River, located about 55 km north of Fort McMurray and 450 km north of Edmonton (Figure 1). The population of the community is around 330 people residing in 130 residences. Housing in the community consists mainly of single-family dwellings including a substantial number of mobile homes. Industrial activity occurs all around the community and consists mainly of surface mining and subsurface extraction of oil sands, and oil sand bitumen refining.

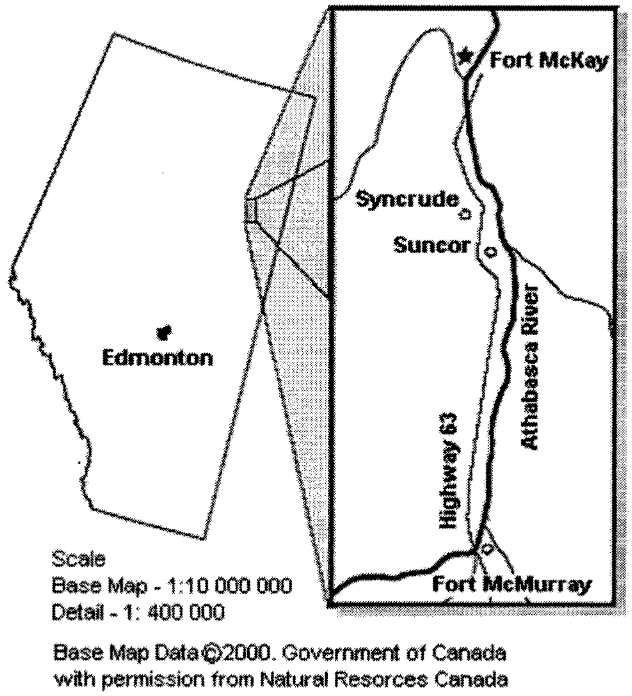


Figure 1. Location of Fort McKay, Alberta.

The ANN study used historical air quality and meteorological data for the period 1 October 2002 to 31 September 2003 from two Wood Buffalo Environmental Association (WBEA) air monitoring stations and an Environment Canada atmospheric weather station in the airshed. One WBEA air monitoring station (AMS #1) is situated near the northwest corner of the Fort McKay Water Treatment Plant property, just outside of the community. The second WBEA air monitoring station (AMS #13) is located between the community of Fort McKay and the Syncrude Canada Ltd. Mildred Lake surface mine, approximately 4.4 km directly south of AMS #1 on Syncrude Canada Ltd. property (Appendix A). The Environmental Canada atmospheric weather station is located about 8 km south of Fort McMurray at Fort McMurray airport (about 59 km south of Fort McKay).

1.2 STUDY OBJECTIVES

Air quality in the community of Fort McKay is concern to its residents in relation to current and future oil sands activities surrounding the community. In particular, surface mining and other related PM-generation activities at the Syncrude Canada Ltd. site are perceived as having a direct influence on resulting ambient $PM_{2.5}$ concentrations in the community and at the Fort McKay air monitoring station. An ANN modeling study was undertaken to examine whether and to what extent this influence is occurring. The two objectives of this study are described below:

1. The first objective was to evaluate the feasibility of using ANN to predict ambient $PM_{2.5}$ concentrations at the Fort McKay air monitoring station (AMS #1). In the process of developing a suitable ANN model to predict ambient $PM_{2.5}$ concentrations at the Fort McKay air monitoring station, a series of ANN models with different input data were evaluated. Results from these models were further examined in order to identify an optimum ANN model structure to predict ambient $PM_{2.5}$ concentrations. Input factors that were shown to be important contributors to prediction of ambient $PM_{2.5}$ concentrations at the Fort McKay air monitoring station were identified. Air quality parameters measured at AMS #13 were used as surrogate input information to represent “near-field” air quality resulting from PM-generation activities at the Syncrude Canada Ltd. site.

2. Given the optimum ANN model structure, a second objective was to conduct a sensitivity analysis. The sensitivity analysis examined whether the final model would respond to simulated increases in one of the important input factors. This input factor consisted of $PM_{2.5}$ concentration at AMS #13 that represented the "near-field" influence of PM-generation activities at the Syncrude Canada Ltd. site. If increases in "near-field" $PM_{2.5}$ concentration (i.e. at AMS #13) occur as a result of increases in PM-generation activities at the Syncrude Canada Ltd. site, the issue of whether similar responses would occur at the Fort McKay air monitoring station was examined.

2.0 LITERATURE REVIEW

2.1 AIRBORNE PARTICULATE MATTER

2.1.1 $PM_{2.5}$

$PM_{2.5}$ is particulate matter with an aerodynamic diameter less than 2.5 μm (Tucker, 2000). "Aerodynamic diameter" is a term used for identifying non-spherical particles (US EPA, 2003a). It is the diameter of a spherical particle with a unit density and has the same aerodynamic properties in air as the particle of interest. Particles having identical aerodynamic diameters may have different shape and density.

There are three major forms of $PM_{2.5}$ (Pacific Environmental Services, Inc., 1999): primary, condensable, and secondary particulate matter. Primary particulate matter is emitted directly in the solid phase. Condensable particulate matter emitted at high temperature in the gas phase. This form will condense into the solid phase upon dilution and cooling. Secondary particulate matter forms through atmospheric reactions of gaseous SO_2 and NO_x. The formation of these secondary components involves complex chemical and physical interactions.

Primary solid particulate matter includes soil-type particles and organic/elemental carbon-type particles (US EPA, 1997). Many anthropogenic sources, like road dust, dust from construction, dust from ore processing and refining, and dust from agriculture can contribute to the soil-type fine particles. Sources of carbon-type particles are diesel vehicles, prescribed or open burning, wood stoves and fireplaces, and boilers. Primary condensable particulates consist mainly of semi-volatile organic compounds, which condense at ambient temperature to form aerosol sources (Pacific Environmental Services, Inc., 1999). This component may represent a significant fraction of the $PM_{2.5}$ emitted from some sources.

Industrial activities, forest fires, non-industrial fuel combustion and transportation were reported as main sources of primary $PM_{2.5}$ emitted in Canada (Deslauriers, 1996). Other activities like incineration only contributed a small part to primary $PM_{2.5}$ in Canada. For example, total primary $PM_{2.5}$ emissions in Canada were estimated at 786,700 tons in 1990 (Table 1). Industrial activities, forest fires, non-industrial fuel combustion and transportation contributed about 762,000 tons, whereas incineration contributed only 13,683 tons (Table 1). Data in table 1 are outdated, however they do show the relative contribution of each category.

Secondary $PM_{2.5}$ is formed through heterogeneous chemical reactions that transform gaseous pollutants into very small particles (Pacific Environmental Services, Inc., 1999). It can be a predominant part of airborne particulate matter. For instance, Erishman and Schaap (2004) indicated that secondary PM may comprise 50% or more of ambient $PM_{2.5}$. In most urban sites, secondary $PM_{2.5}$ is made up of by sulfur and nitrogen species. However, secondary organic aerosols can also significantly contribute to secondary $PM_{2.5}$ in other locations.

Table 1. Estimated PM emissions in Canada in 1990 (adapted from Deslauriers, 1996).

Category/Sector	PM (t)	$PM_{2.5}$ (t)	$PM_{2.5\text{-}10}$ (t)	PM_{10} (t)
Industrial Sector	810,366	270,182	178,509	448,691
Non-industrial Fuel Combustion	272,842	126,513	31,041	157,554
Transportation	133,489	101,493	13,896	115,389
Incineration	34,248	13,683	5,128	18,811
Forest Fires	293,123	263,811	26,381	290,192
Miscellaneous	30,430	11,026	7,105	18,131
TOTAL	1,574,498	786,708	262,060	1,048,768

Note: Open sources and secondary particles not included.
PM = particulate matter.
$PM_{2.5}$ = particulate matter with a diameter less than 2.5 μm.
$PM_{2.5\text{-}10}$ = particulate matter with a diameter less than 10 μm and greater than 2.5μm.
t = metric ton.

2.1.2 Relationship of $PM_{2.5}$ and PM_{10}

$PM_{2.5}$ is the fine component of particles with a diameter smaller than 10 microns. Most of the particles in coarse mode of PM_{10} (diameter >2.5μm and ≤10μm) originate from mechanical processes. A typical size distribution of particles in ambient air is shown in Figure 2 after Wilson and Suh (1996).

Generally there is a strong site-specific correlation between PM_{10} and $PM_{2.5}$ for Canadian cities (CEPA/FPAC, 1999). However, this is not the case for two prairie cities, Winnipeg and Edmonton. As reported by CEPA/FPAC (1999), R^2 values are low for the relationship between PM_{10} and $PM_{2.5}$ as described by a linear equation for these prairie cities. As shown in Table 2, R^2 values were 0.42 and 0.46 for Winnipeg and Edmonton, respectively. Low R^2 values indicate that linear equations are not suitable for modeling the relationship between PM_{10} and

$PM_{2.5}$ at these cities (e.g. <0.45).

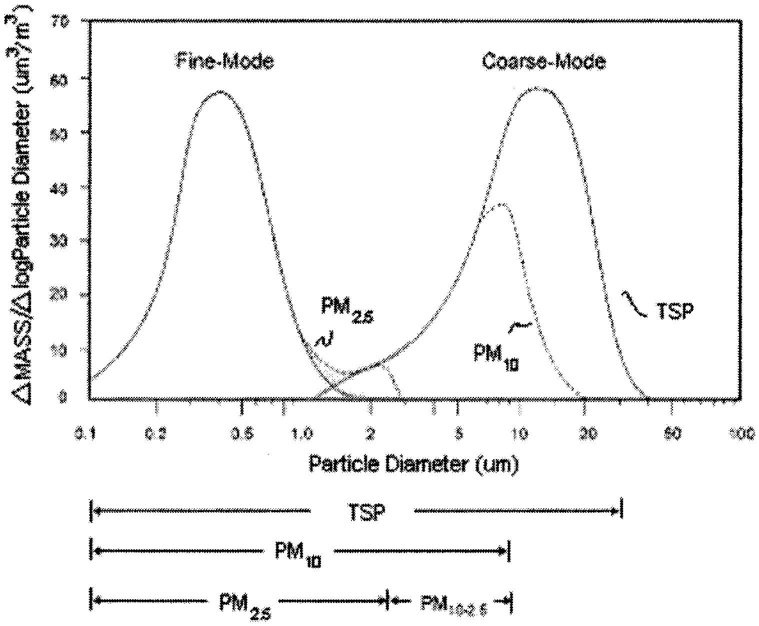


Figure 2. PM_{10} and $PM_{2.5}$ size distribution for ambient particles (adapted from Wilson and Suh, 1996).

2.1.3 $PM_{2.5}$ Environmental and Health Effects

<u>Corrosion and Discoloration Effect on Materials</u>

Santachiara et al. (1989) reported that $PM_{2.5}$ could perform like a catalyst for converting SO_2 and NO_x into sulfuric acid and nitric acid. After deposition on a material surface, these acidic particles can degrade material extensively. Many materials like metal, paint, stone and electronics, can be damaged by the corrosive/erosion effect of $PM_{2.5}$ (CEPA/FPAC, 1999).

Table 2. Site-specific relationships between PM_{10} and $PM_{2.5}$ for samples from year 1984 to 1993 (adapted from CEPA/FPAC, 1999).

Cities	Number of Samplers	Mean Fine Mass (µg/m^3)	Mean $PM_{2.5}$ to Mean PM_{10} Ratio	Linear Equation: $PM_{2.5} = a + b*PM_{10}$		
				b	a	R^2
Saint John	292	10.1	0.58	0.58	0.09	0.86
Halifax	304	14.1	0.55	0.55	0.06	0.74
Kejimkujik	277	7.2	0.63	0.69	-0.62	0.86
Montreal	577	15.9	0.57	0.65	-2.12	0.81
Montreal-Duncan/Decarie	314	20.9	0.47	0.46	0.37	0.7
Quebec City	221	11.9	0.5	0.54	-0.72	0.7
Sutton	136	7.7	0.68	0.83	-1.59	0.98
Ottawa	358	12.6	0.56	0.63	-1.66	0.77
Windsor	352	18.1	0.57	0.57	-0.16	0.81
Windsor-College	422	16.8	0.56	0.59	-0.76	0.73
Toronto	586	16.8	0.6	0.64	-1.14	0.84
Walpole	275	17.6	0.59	0.56	1	0.7
Egbert	137	10.4	0.61	0.64	-0.49	0.89
Winnipeg	447	10.3	0.36	0.29	2.02	0.42
Edmonton	380	10.5	0.39	0.34	1.32	0.46
Calgary	504	11.2	0.42	0.41	0.3	0.65
Vancouver	334	15.5	0.63	0.74	-2.71	0.88
Vancouver-W 10th	360	15.6	0.58	0.65	-1.78	0.81
Victoria	393	11.5	0.65	0.81	-2.91	0.91
All Sites	6669	13.8	0.53	0.52	0.32	0.7

Note: PM_{10} = particulate matter with a diameter less than 10 µm.
$PM_{2.5}$ = particulate matter with a diameter less than 2.5 µm.

The elementary carbon (EC) component in $PM_{2.5}$ can stain materials after deposition. Surfaces of historical buildings, monuments, and statues deteriorate from cleaning work caused by this problem. Particles settled in households can stain fabric and furniture. The Real Estate Research Institute (RERI, 1994) estimated that a unit ($\mu g/m^3$) reduction in PM_{10} could produce a $3.13 savings per household in washing and cleaning.

Adverse Environmental Effects

Because of acidic and heavy metal components, particulate matter deposition can lead to acidic conditions in water, soil nutrient depletion, and vegetation damage (US EPA, 2000). Fine particles may have an important effect on distant vegetation, because they can remain airborne, travel long distances, and are more easily captured by impaction (Smith, 1990). Visibility impairment results from scattering and absorption of light by fine particles in the atmosphere (CEPA/FPAC, 1999). $PM_{2.5}$ scatters and absorbs light more efficiently than larger particles (Pacific Environmental Services Inc., 1999).

Adverse Human Health Effects

Particles less than 10 microns ("inhalable" particles) may be inhaled through the lung (Vedal, 1995). Particles less than 2.5 microns ("respirable" or "fine" particle) can get further into the alveoli in the distal parts of the lung (CEPA/FPAC, 1999). Breathing fine particles in urban air has been statistically related to a series of health problems including: aggravated asthma, increases in respiratory symptoms like coughing and difficult or painful breathing, chronic bronchitis, decreased lung function, and premature death (Hornberg et al., 1998; Schwartz et al., 1996; Vedal, 1996). Burnett et al. (1995) reported that ambient inhalable particulate matter concentrations in cities have been linked to mortality, increased respiratory symptoms, and changes in pulmonary mechanical functions. Although scientific studies have linked breathing ambient PM to these health problems, there are still uncertainties associated with this link (Sloss and Smith, 2000). In addition, the actual chemistry and processes that contribute to formation of $PM_{2.5}$ are still poorly understood (Jones, 1996).

2.1.4 Guidelines and Standards for $PM_{2.5}$

US Environmental Protection Agency (EPA) Standard

In 1996, the US EPA acknowledged that ambient PM standards could be associated with serious health effects (US EPA, 1996a). Standards for PM at the time emphasized particles smaller than 10 microns in diameter. However, studies indicated that fine particles (smaller than 2.5 microns) contribute more to adverse effects to human and were more responsible for visibility impairment (US EPA, 1996a). US EPA introduced new standards as a result that focused more on fine particles in the atmosphere.

The US EPA annual $PM_{2.5}$ primary standard is 15 $\mu g/m^3$ (US EPA, 2003b). To attain this standard, the 3-year average of the annual arithmetic mean of 24-hour concentrations from single or multiple population oriented monitors must not exceed 15 μg/m^3. The 24-hour primary standard is 65 μg/m^3 (US EPA, 2003b). To attain this standard, the 98th percentile of the distribution of the 24-hour concentrations for a period of 1 year, averaged over 3 years, must not exceed 65 μg/m^3 at each monitor within an area. The 24-hour and annual secondary standards are same as the primary standards. The US EPA primary standards are based upon protecting public health. The US EPA secondary standards are based upon protecting public welfare, such as visibility, material and ecological systems.

Canada-wide Standard

In June 2000, the Canada-Wide Standards (CWS) for PM and Ozone were introduced to reduce PM and ground level ozone in Canada by 2010 (CCME, 2000). This was reported as a necessary effort to achieve a long-term goal of reducing the risks of these pollutants to human health and the environment. PM and ozone are included in the same CWS because they have common sources and they contribute to smog (CCME, 2000). The CWS for PM and ozone was established to achieve a balance between health and environmental protection and the feasibility and costs of reducing pollutant emissions that contribute to PM and ground level ozone in ambient air.

The CWS for particulate matter focuses on the fine fraction of PM with a diameter less than 2.5 microns (CCME, 2000). The CWS for $PM_{2.5}$ is 30 μg/m^3 averaged over 24 hours, to be achieved by 2010. At the same time, CCME (2000) pointed out that individual jurisdictions should continue to apply their existing air quality objectives and/or guidelines for the coarse fraction of PM to guide management actions.

2.1.5 $PM_{2.5}$ Physical and Chemical Characteristics

Physical Properties

High temperature combustion and secondary particles are main sources of $PM_{2.5}$. Because of its light density and small size, $PM_{2.5}$ has a long residence time and can travel long distances in the atmosphere (CEPA/FPAC, 1999). The source of origin and ways in which it is created (pulverizing, abrasion, condensation, nucleation, crystallization and agglomeration) will determine its shape. Katrinak et al. (1993) indicated that urban particles and combustion particles tend to have an irregular shape. Due to its light density, surface irregularities and internal pores, small particles have a large surface area to mass ratio. These properties will affect the formation, growth, transport, and removal of particles. For small particles (≤1 μm), dry deposition and precipitation scavenging are the predominant removal mechanism instead of sedimentation, which is a main mechanism for larger particles (CEPA/FPAC, 1999).

Chemical Properties

The main $PM_{2.5}$ chemical components are nitrate, sulfate, ammonium, trace elements, and elemental and organic carbon. Particles in this size range are typically acidic while particles in coarse range are basic. Figure 2 displays a typical chemical compound distribution corresponding to particle size.

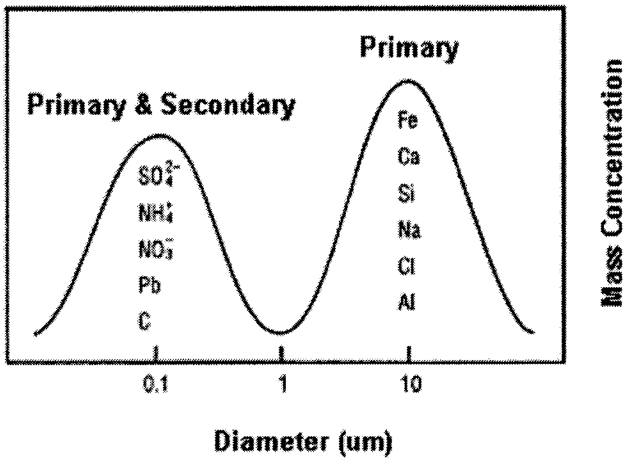


Figure 3. Chemical compound distribution corresponding with particle size (adapted from Seinfeld, 1986).

Urban Particles

Cheng et al. (1998) summarized characteristics of particulate matter in two major cities of Alberta, Edmonton and Calgary. It was mentioned that there was no difference in $PM_{2.5}$ concentrations between Edmonton and Calgary. Monitoring stations in both cities showed slightly higher $PM_{2.5}$ loadings in winter. In addition, a higher fraction of soil was found in the coarse fraction (55% to 65%) than in the fine fraction (7% to 8%). According to element analysis, chemical profiles of fine particulate matter in Edmonton and Calgary were very similar. Sandhu (1998) suggested that particulate matter in Edmonton and Calgary might originate from similar dominant source categories, despite geographical and industrial differences.

Rural Particles

Sandhu (1998) indicated that the background $PM_{2.5}$ concentrations in Alberta were low, in the range of 3 to 6 µg/m^3, whereas background PM_{10} concentrations ranged from 10 to 24 µg/m^3. The background ratio of $PM_{2.5}$ to PM_{10} was about 0.3. Measurements made near significant local sources indicated that local and regional sources could significantly increase $PM_{2.5}$ and PM_{10} concentrations (Sandhu, 1998).

2.1.6 Distribution and Dispersion

There are three key dispersion mechanisms for pollutants in the atmosphere (Wark et al., 1998). The most important one is “mean air motion” that transports pollutants downwind. “Turbulent velocity fluctuations” disperse pollutants in all directions and “mass diffusion” combine to create concentration gradients in the atmosphere. General aerodynamic characteristics of PM (e.g. size, shape, and weight) can affect the settlement rate. Important meteorological parameters for pollutant dispersion are discussed below.

1. **Solar Radiation**

The actual quantity of solar energy received by a unit of surface area on the earth’s surface is determined by factors such as location, season, specific time of a day, and composition of the atmosphere above the surface (Wark et al., 1998). Differential rates of warming create temperature differences between parts of the earth causing air motion (wind).

2. Atmospheric Stability

The stability of the atmosphere is important for estimating dispersion of a pollutant. A stable atmosphere is defined as one without vertical mixing or motion (Wark et al., 1998). Accordingly, temperature gradients and mechanical turbulence due to the shearing action of wind will determine vertical dispersion of pollutants. Mixing caused by thermal conditions can be determined by comparison of the actual temperature gradient (or environmental lapse rate) to the dry adiabatic lapse rate, as illustrated in Figure 4.

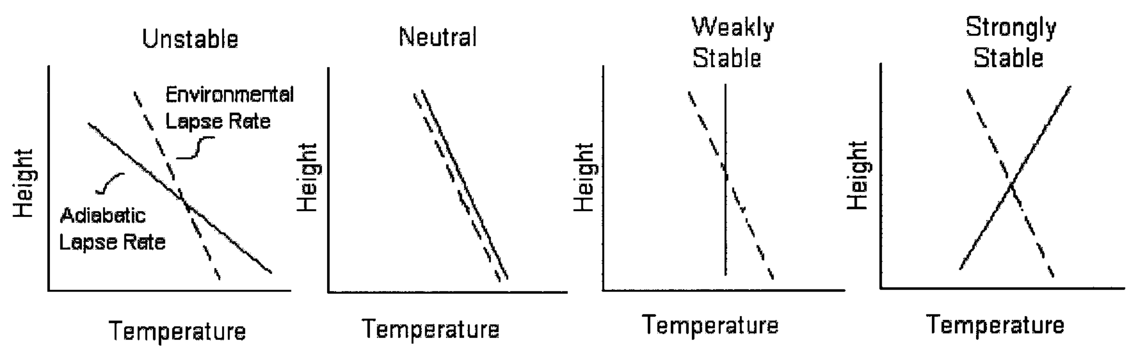


Figure 4. Relationships between lapse rate and air stability (adapted from Wark *et al*, 1998).

The lapse rate is defined as the negative of the temperature gradient in the atmosphere (Wark et al., 1998). Comparison of the dry adiabatic lapse rate to the actual (environmental) lapse rate in the lower atmosphere is used to characterize the stability of the atmosphere. An unstable atmosphere is where buoyancy increases the displacement of a parcel of air that has moved upwards or downwards, e.g. during a windy, gusty day. A stable atmosphere is where buoyancy returns a parcel of air to its original position after it has been displaced upwards or downwards.

3. Wind Velocity Profile

Wind is air movement resulting from pressure and temperature differences in the atmosphere (CASA, 2004a). Different wind profiles originate from different terrain properties such as location and density of trees, and location and size of lakes, rivers, hills, and buildings. The lower air layer – called the planetary boundary layer – can be influenced by friction from a few hundred meters to several kilometers above the surface of the earth (Wark et al., 1998). When there are buildings or trees on the earth surface, the wind speed profile will be gentler because of the friction effect of these obstacles.

4. Maximum Mixing Depth (MMD)

The vertical height of mixing in the atmosphere is usually quantified by mixing height or mixing depth (Beychok, 1995). This is the height at which vertical mixing takes place. The vertical temperature profile is closely related to the forecasting of mixing height.

The height at which the dry adiabatic line intersects the environmental profile line is called the maximum mixing depth - MMD (Wark et al., 1998). In unstable air the MMD is higher and in stable air the MMD is lower. Pollutants in unstable air will be dispersed over a longer vertical distance. This means that the resulting pollutant concentration will be lower in an unstable atmosphere. Generally, the MMD will be lower at night and higher at daytime. MMD can range from <100 m under a severe inversion at night to 3000 m at daytime. In addition, it will be at a lower level in winter than in the early summer.

5. Turbulence

Atmosphere turbulence is caused by thermal and mechanical conditions (CASA, 2004b). Thermal disturbances occur during a sunny day with a negative temperature gradient. Mechanical turbulence is a result of terrain roughness effects.

PM Dispersion Modeling

1. Gaussian Dispersion Model

Most dispersion models for point sources are based on the Gaussian dispersion equation (Dobbins, 1976). This relationship assumes that the concentration of an air pollutant will distribute in a normal/bell shape in the atmosphere:

$$C(x,y,z) = \frac{Q}{2\pi u \sigma_y \sigma_z} \exp[\frac{-(y-y_0)^2}{2\sigma_y^2}] \exp[\frac{-(z-z_0)^2}{2\sigma_z^2}] \quad (1)$$

Q is the emission strength of the source; u is the average wind speed; C is the pollutant concentration; σ is the standard deviation, which will determine the shape of the Gaussian dispersion; x, y and z are the coordinates. This is illustrated in Figure 5.

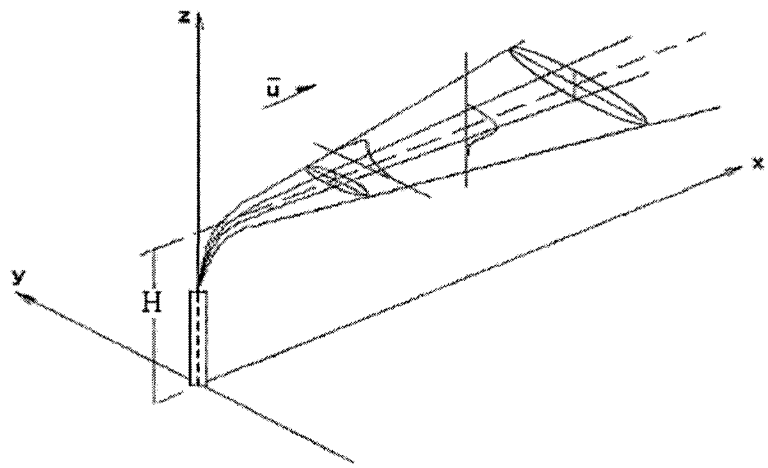


Figure 5. Gaussian dispersion of a point source located at (0, 0, H) in a wind-oriented coordinate system (adapted from Dobbins, 1976).

2. Dispersion Models Used in Alberta

The aim of dispersion models used by Alberta Environment is to provide a method for calculating ambient ground-level concentrations of an emitted substance and to provide information about emissions and the nature of the atmosphere (Idriss, 2003). Alberta Air Quality Modeling Guidelines (Idriss, 2003) sort these models into three levels. First level models are screening models used to determine a specific event or likelihood of a specific event. Second level models are refined models that are named because of their higher level of sophistication. Third level models are advanced models that treat specific dispersion processes in greater detail.

One screening model and five refined models are recommended in the Alberta Air Quality Modeling Guidelines for use in Alberta. All regulatory models except for the CALPUFF model are for short-range applications. These models can only predict air quality within ~25 km from sources. The CALPUFF model can be used for distances up to 200 km. Additional details about these air quality models can be found at http://www.epa.gov/ttn/scram/.

2.1.7 Air Quality Monitoring

Monitoring methods for particulate matter have been summarized by the US EPA (1996b). US Federal Reference Methods for PM_{10} and $PM_{2.5}$ are intended to achieve mass measurements within ±10% precision (Wilson et al., 2002). However, there are problems in getting accurate result because of limitations caused by semi-volatile PM and particle-bound water (Chow, 1995; Wilson et al., 2002). Wilson et al. (2002) mentioned that it is expensive to take measurements according to the US FRM methods, which require a microbalance and extensive quality control.

PM Monitoring at Wood Buffalo Environmental Association (WBEA)

Several methods and techniques are using to monitor outdoor airborne particulate matter. Instruments used in these methods can continuously or intermittently measure concentrations of collected particles (CASA, 2004c). Equipment used by the Wood Buffalo Environmental Association (WBEA) measures $PM_{2.5}$ concentrations hourly (continuous) and every sixth day (intermittent).

The equipment for continuously monitoring PM_{10} and $PM_{2.5}$ is a Tapered Element Oscillating Microbalance (TEOM) (AEP, 2001). Air sampled enters an inlet that can aerodynamically separate particles of a specific diameter. The air passes through a filter that is attached to a tapered element. The element's vibration frequency will change because of accumulation of deposited particles on the filter. Particle mass can be obtained by measuring the changes of vibration frequency. The equipment for intermittently monitoring PM_{10} and $PM_{2.5}$ is a dichotomous high volume sampler. More information about air monitoring equipment in WBEA can be found at the WBEA web site (http://www.wbea.org/).

2.2 ARTIFICIAL NEURAL NETWORK (ANN)

There are three kinds of artificial intelligent technology: fuzzy logic systems, expert systems, and artificial neural networks (Harvey and Harvey, 1998). Fuzzy logic systems use uncertain words, like "a bit", "a lot", "fast" or "slow", to describe a target problem instead of using exact numbers. Expert systems use rules like "if-then-else" to solve problems. These rules are from an expert's experience and stored in the knowledge database for use. Artificial neural networks (ANNs) are different from the other two systems. An ANN simulates the thinking process of human brains to solve complicated problems (Jain et al., 1996).

An artificial neural network (ANN) may offer advantages in modeling processes that follow non-linear relationships, such as those that govern pollutant behavior in the atmosphere. ANN modeling uses historical data to “learn” patterns that occur between given inputs and outputs of the model, and then simulates the outputs. The method tends to be a “semi-black box” method, where equations describing complex situations are not known. Therefore, ANN may be suitable for ambient air pollution modeling for situations where the governing relationships of irregular micrometeorology and terrain are too complex to solve or expensive to derive.

2.2.1 Basic Model Structure

Basic multi-layer perceptron (MLP) ANNs, which are mainly used, include three layers: an input layer, a hidden layer and an output layer. There are neurons that simulate neural cells of the human brain in these layers. Each neuron in the output layer and hidden layer is connected with every neuron in the previous layer. The basic layout of a MLP is displayed in Figure 6.

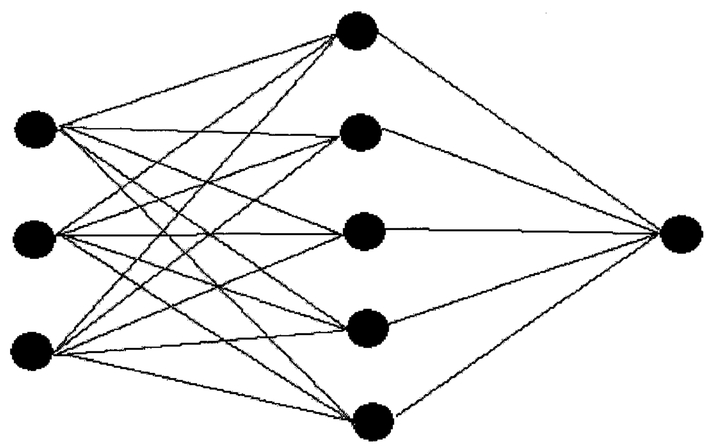


Figure 6. Basic structure for a multilayer perceptron ANN.

2.2.2 Learning Process

The number of input layer neurons equals the number of inputs (Ward Systems Group, Inc., 2003). When input data go into the input layer, they will be scaled (e.g. 0 to +1 or -1 to +1). They will then be multiplied by a weight factor and transferred to each of the hidden layer neurons. In the hidden layer neurons, the sum of the multiplied results will be processed by an activation function. If the result can activate a hidden neuron, it will be multiplied with another weight factor and transferred to the output layer.

In the output layer, the sum of multiplied results from the hidden layer will be processed by another activation function and an output function. Results will then be scaled to the real output data range. In an ANN with a back propagation algorithm, a result from the output layer will be compared with the actual output

(Ward Systems Group, Inc., 2003). If there is a large error, the model will change the connection weight between layers and repeat the above-mentioned steps until certain stop criteria (e.g. a small error or stop epochs) are satisfied.

2.2.3 Model Development

Because ANN is still a novel tool for modeling, there are uncertainties in the model development process. Therefore trial and error methods are used. The basic steps in model development and construction are listed below. More details involved in these steps are described in next chapter of this thesis:

- Data preprocessing -
    - Data sorting/formatting
    - Error screening
    - Statistical analysis
- Model construction -
    - Input and output selection
    - Model structure determination
- Analysis of model results
- Model application (sensitivity analysis)

2.2.4 Comparison with Other Models

Historically, statistical regression models have been used to make predictions about air quality. However, researchers are generally not satisfied with this method because of complexities of a modeled domain and the methods tend to produce poor results.

Other researchers have compared results from ANN models to regression models. Chaloulakou et al. (2003) constructed an ANN model and a multiple linear regression model. Using two-year meteorological data as inputs, PM_{10} concentrations were predicted for both models. Results from these models were compared and indicated that the ANN model had a smaller root mean square error (8.2% to 9.4% lower than a multiple linear regression model). A smaller root mean square error indicates that the model has a higher precision in predicting outcomes. These researchers concluded that with proper construction and training ANN models could be successfully used as an air quality prediction tool.

Baba et al. (1990) conducted a research on coagulant injection in a water treatment plant. They compared absolute mean error obtained from ANN models with results obtained from conventional multi-regression analysis. Results are listed in Table 3. Table 3 indicates that results from the ANN model had smaller errors indicating an ability to achieve a higher precision in predicting outcomes. This was more obvious under an abnormal condition (Baba et al., 1990). This conclusion was very useful because in real life abnormal conditions occur frequently and cannot be handled well by conventional models.

Table 3. **Comparison between ANN and multi-regression model for a water treatment application (adapted from Baba et al., 1990).**

Model	Absolute Mean Square Error for Normal Conditions	Absolute Mean Square Error for Abnormal conditions
ANN	1.14 mg/l (7.7%)	1.72 mg/l (8.8%)
Multi-regression	1.23 mg/l (8.3%)	2.60 mg/l (13.2%)

Valentin et al. (1999) also conducted research in the water coagulation field. Results from an ANN model and a linear regression model are compared in Figure 7 and Figure 8 from their research. Figure 7 (a) and (b) indicate that for the ANN model, the predicted coagulant dosage line could trace increases and decreases with a small error. Figure 8 (a) and 8 (b) indicate that for the linear regression model, prediction results were poorer.

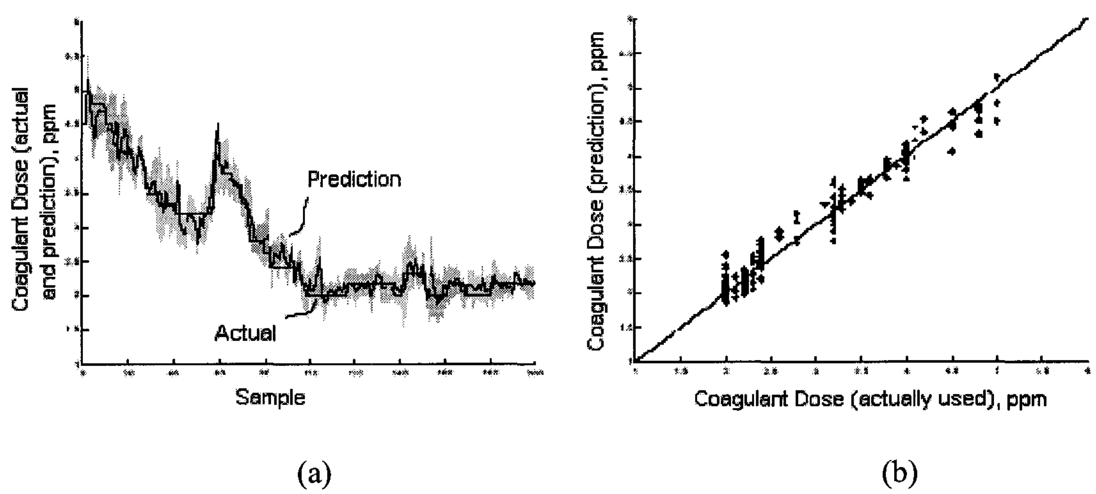
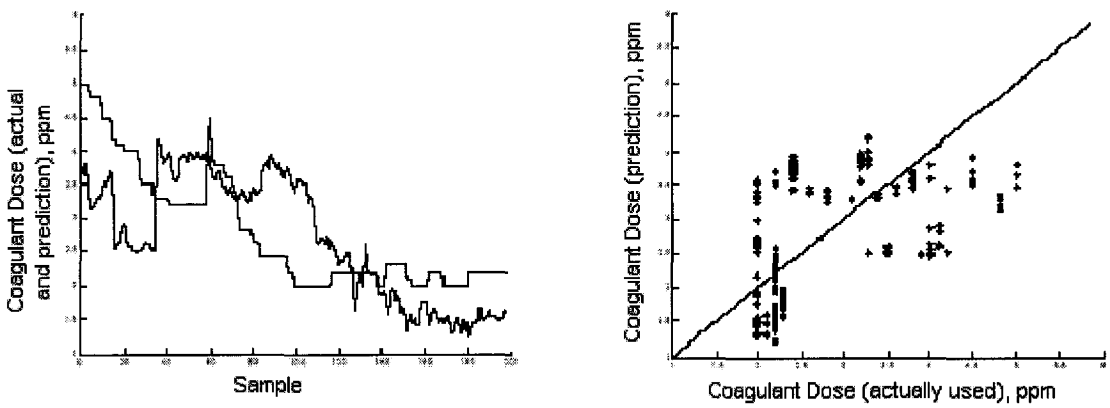


Figure 7. Result from the ANN model - Actual vs. Prediction for a water treatment application (adapted from Valentin et al., 1999).



(a) (b)

Figure 8. Result from the linear regression model- Actual vs. Prediction for a water treatment application (adapted from Valentin et al., 1999).

2.2.5 Applications with $PM_{2.5}$

ANNs have been used successfully for modeling O_3, SO_2, NO and NO_x (Boznar et al., 1991; Yi and Prybutok, 1996; Comrie, 1997; Spellman, 1999; Gardner and Dorling, 2000; Chelani et al., 2002; Hasham et al., 2004). However, there are much fewer publications related to modeling of inhalable and respirable particulate matter with ANN (Chaloulakou et al., 2003). Modeling of PM with ANN has been undertaken by Kohlemainen et al. (2000) in Kuopio, Finland, Chelani et al. (2002) in Jaipur, India, and Lu et al. (2002) in Hong Kong.

Most of these researchers agree that ANN has the potential to make predictions of airborne particulate matter behavior. Although McKendry (2000) stated that neural network models show little if any improvement over regression models when used to predict PM_{10} and $PM_{2.5}$ concentrations in the Lower Fraser

Valley area in British Columbia, Canada. However, McKendry (2002) noted that his study was specific to the geographical, transportation, and pollutant conditions in the Lower Fraser Valley area.

2.2.6 Advantages and Limitations of ANN

ANN is a tool simulating the thinking process of human brains. This makes it different with other conventional modeling methods. Advantages of ANN are reported to be (Baba et al., 1990):

1. It can be used to solve complex non-linear problems, which are difficult for current conventional deterministic and statistical methods.
2. It is a semi-black box method, which means it does not require a model constructor to provide detailed mathematical formula or algorithms to describe the problem. Therefore, it can be used when the target problem is complex and the relationship between input parameters and output parameter is not understood thoroughly.
3. The neural network can learn from and make predictions according to historical data.

There are limitations of ANNs. These limitations can result from insufficient databases for learning (Zvi Boger, 1992). For example if an important input is missed, a large test error can be produced. If there is a lack of learning samples in a certain region, ANN models may make unrealistic extrapolations.

On the other hand, results are sometimes hard to explain or understand because of the "semi-black box" property of ANN models.

3.0 ANN MODELING OF $PM_{2.5}$ IN FORT MCKAY

ANN model construction involved two components: source data pre-processing and model development. The first component involved screening input data to generate a suitable input data file for ANN model development. Model development included two steps. In the first step, sources and characteristics of available input parameters were analyzed to judge their fit as model inputs. Then a series of basic models with different inputs were constructed. Important model inputs were identified from this step. In the second step, models with different software settings were run and results (R^2 values) from these models were compared to identify optimal ANN software settings.

3.1 SOURCE DATA PRE-PROCESSING

Data collection and processing are important for ANN model building and these tasks can affect the learning accuracy (Joo et al., 2000). Baxter et al. (1999) indicated that blind application of original data could lead to a model with poor generalization ability. This means that a poor data set and/or poor understanding of a study domain can produce a model with poor prediction capabilities.

Generally a model constructor can only get raw data from a third party. This means errors and outliers may exist in the raw data set. This situation occurs often. Some of these errors and outliers are a result of instrument failure or malfunction; some are a result of data collection (e.g. typing error during data entry). If original data with these errors are used to train a network, the results can be very poor. Moreover, even when there are no errors in a data set, a poor model can be produced. This will happen when a model constructor does not understand the target processes well and uses the data blindly. For example, an ANN may be unsuitably used to solve a linear problem for which mathematic models will provide a better result. On the other hand, a poor model can be produced because irrelevant parameters are used as inputs.

Several methods are used to pre-process data. The first one is a conventional method, which is based on general logic. For example, raw data can be compared with an upper and lower limit (range) of normal values for a parameter. If a value is within this range, the data can be used as input and if it is beyond the range, it can be removed. Each parameter used in this study was analyzed using this method.

A second method is a statistical method. Joo et al. (2000) constructed data sets by using statistical methods and compared running results with results of original data sets. They proved that using a statistical method to eliminate data outliers data improved both the learning rate and prediction ability of an ANN model.

Another data preprocessing method uses the self-organization map (SOM) method introduced by Kohonen (Valentin et al., 2001). The SOM algorithm can do clustering and projection work, which can be used to validate data and do data reconstruction for multi-parameter data sets. This is a very important property for model constructors who want to use ANN as a process control tool, because instrument malfunction or failure occurs a lot in real life. By using a Kohonen network, instruments having problems can be found and missing data can be reconstructed. In this research project, the Kohonen network was used to separate data with different features, which could not be identified by general methods.

3.1.1 Statistical Analysis

The original data for this research were obtained from two air monitoring stations and an airport meteorological station in northeast Alberta. These data consisted of hourly average values for measured parameters during the period 1 October 2002 to 31 September 2003. One air monitoring station is located at Fort McKay, Alberta (AMS #1). The other air monitoring station is located adjacent to the Syncrude Canada Ltd. Mildred Lake site (AMS #13). The meteorological station is located at the Fort McMurray airport. The relative position of the two air monitoring stations is shown in Figure 9. The distance between these two stations is approximately 4.4 km (Appendix A).

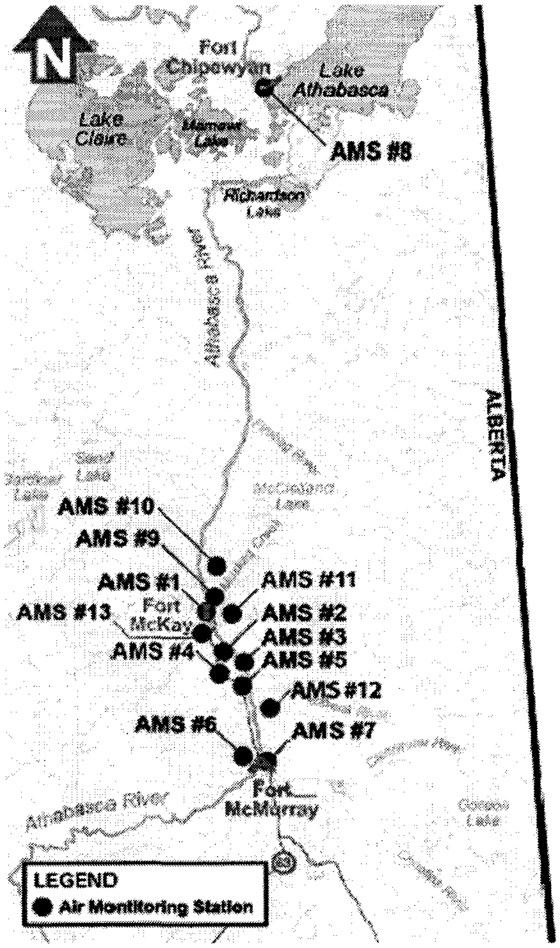


Figure 9. Position of WBEA air quality monitoring stations (adapted from WBEA, 2004).

Every parameter in the original input data were analyzed to estimate the maximum, minimum, mean, standard deviation, variance, 0.01, 0.05, 0.25, 0.75, 0.90, 0.95, 0.99 percentile values. These data are presented in Appendix B. Time series plots of each parameter were also generated in Excel software (Appendix C). The tabular results and plots were used to examine features of each parameter.

Parameters at Fort McKay Air Quality Monitoring Station (AMS #1)

SO_2 (Sulfur Dioxide): Negative concentration values in the original data were identified and deleted. The time series plot (Appendix C) indicated an increasing SO_2 concentrations were observed during some days. However, the intervals between any two increasing events were not same. Most of the peak values (i.e. above 10 ppb) occurred at noon.

TRS (Total Reduced Sulfur): Negative concentration values in original data were deleted.

THC (Total Hydrocarbon): Negative concentration values in the original data were deleted. THC concentrations fluctuated around 2 ppm.

O_3 (Ozone): Negative concentration values in the original data were deleted.

NO (Nitric Oxide): Negative concentration values in the original data were deleted. There were numerous high NO-concentration events during winter as shown in time series plots (Appendix C).

NO_2 (Nitrogen Dioxide): Negative concentration values in the original data were deleted. Same as for NO, there were numerous high NO_2-concentration events during winter as shown in time series plots (Appendix C).

NO_x (Oxides of Nitrogen): Negative concentration values in the original data were identified and deleted. Same as with NO and NO_2, there were much higher values observed during winter. This tendency is shown in the time series plot (Appendix C).

$PM_{2.5}$: Negative concentration values in the original data were deleted. According to the time series plot (Appendix C), more events with higher concentration were observed during summer. The mean $PM_{2.5}$ concentration in Fort McKay was around 4 $\mu g/m^3$, lower than that observed in Edmonton and Calgary (~10 $\mu g/m^3$) (CASA, 2004d), however consistent with that observed for rural areas in Alberta (3 to 6 $\mu g/m^3$) (Sandhu, 1998).

ETL (Temperature): Most of the temperature (ETL) values in the original set were within the normal range of -36 to +34°C. However, there were some outliers. For example, during summer several -50°C values were recorded in the original file. These data patterns were deleted.

PC (Precipitation): Precipitation (PC) data were listed in original data. Unfortunately there was a large body of invalid data (2,221 out of 8,760 data sets). As a result this parameter was not used as an input parameter.

RH (Relative Humidity): Most of the Relative Humidity (RH) values in the original data set were within the normal range (mean = 68%). RH fluctuated more during summer than in winter.

GR (Global Radiation): Negative values in the original data were identified and deleted. Global radiation increased during summer and decreased during winter. This tendency was shown clearly in the time series plot (Appendix C).

WS and WS_SD (Wind Speed and Wind Speed Standard Deviation): Wind speed (WS) and standard deviation of wind speed (WS_SD) were provided in the original data set. These parameters had a similar distribution in a time series plot and showed no obvious tendency (Appendix C).

WD and WD_SD (Wind Direction and Wind Direction Standard Deviation): Wind direction (WD) and standard deviation of wind (WD_SD) were provided in the original data set. Wind directions distributed more densely around 0° (i.e. from the north) and 180° (from the south). This is consistent with terrain features of the Athabasca Valley running in a north to south direction in the study area.

<u>Parameters at the Fort McMurray Airport</u>

TCO: Total cloud opacity was measured with 11 scales (from 0 to 10) – 0 = totally clear sky and 10 = totally cloud covered sky.

TCA: Total cloud amount was also measured with 11 scales (from 0 to 10) – 0 = totally clear sky and 10 = totally cloud covered sky.

Parameters at Syncrude Canada Ltd. Air Monitoring Station AMS #13

SO_2 (Sulfur Dioxide): Negative concentration values in the original data were deleted. The time series plot (Appendix C) indicated an increasing SO_2 concentration during certain days. However, the intervals between any two increasing events were not same. The distribution and magnitude of SO_2 concentrations at AMS #1 and #13 were quite similar (Appendix B).

TRS (Total Reduced Sulfur): Negative concentration values in the original data were deleted.

THC (Total Hydrocarbon): Negative concentration values in the original data were deleted.

O_3 (Ozone): Negative concentration values in original data were deleted.

NO, NO_2, and NO_x: Negative concentration values in the original data were deleted. Similar to the air monitoring result at AMS #1, there were much high NO concentration events during winter at AMS #13 (Appendix C).

$PM_{2.5}$: Negative concentration values in the original data were deleted. Time series plot (Appendix B) indicated numerous high-concentration events during summer. The mean $PM_{2.5}$ concentration value at AMS #13 was 3 μg/m^3, lower than that observed at AMS #1 (4 μg/m^3).

ET (Temperature): The temperature data distribution was similar to that measured at AMS #1. Outliers observed, e.g. -50°C values observed during summer were deleted.

WS and WS_SD (Wind Speed and Wind Speed Standard Deviation): Wind speed and standard deviation of wind speed were provided in the original data set. These parameters had a similar distribution in the time series plot and showed no obvious tendency. The magnitude of the wind speeds observed at AMS #13 were quite different from those recorded at AMS #1.

WD and WD_SD (Wind Direction and Wind Direction Standard Deviation): Wind direction and standard deviation of wind direction were provided in the original file. Similar to that observed at AMS #1, wind directions distributed more densely around 0° (i.e. from the north) and 180° (from the south) consistent with terrain features of the Athabasca Valley running in a north to south direction in the study area.

3.1.2 Data Screening and Re-formatting

Before the original data could be used as inputs to the model, they had to be processed to screen out invalid values in the original file (as discussed previously). Explanations for each flag number system accompanying the original data indicating valid/invalid values are listed in Table 4. According to Table 4, hourly data were only valid for a flag number "0". For TCA and TCO parameters obtained from Environment Canada for the Fort McMurray airport meteorological station, invalid data were labeled with "-99999".

Summaries of invalid data at these stations are derived and listed in Tables 5 through 7. The original data set included 8,760 sets. After deletion of invalid data, only 1,994 sets remained for ANN modeling.

Table 4. Explanation for flag numbers in original data files for AMS #1 and AMS #13 parameters.

Flag Number	Meaning
0	valid
1	zero/span
2	calibration
3	maintenance
4	data acquisition failure
5	analyzer failure
6	unstable operation
7	power failure
8	not in service

ANN modeling was performed using the commercial neural net software product NeuroShell®2 (Ward Systems Group Inc., Frederick, MD) operated on an IBM® compatible computer in a Microsoft® Windows environment. The software required arrangement of the data to follow a certain format.

After deletion of invalid data, the data file still included patterns with very high $PM_{2.5}$ values. This may have resulted from forest fires. However, the effect of a forest fire on $PM_{2.5}$ concentrations at a remote site may be complicated. This complication relates to fire area size, fire duration, distance between the air monitoring station and fire site, terrain features between air monitoring station and fire site, wind speed, wind direction, and precipitation.

For example, particulates from a forest fire can be transported directly to an air monitoring station by wind as illustrated in Figure 10(a). Depending upon micrometeorology and changes in atmospheric stability conditions during the course of a day, particulates may also arrive at an air monitoring station via a traverse path (Figure 10(b)), circular path (Figure 10(c)), or a reverse path (Figure 1 (d)). Thus the dispersion route of forest fire pollutants can be complicated depending upon micrometeorology and changes in atmospheric stability conditions.

Table 5. Number of invalid data with negative value and labeled by flag number at the Fort McKay air monitoring station (AMS #1) (refer to Table 4 for description of flag #).

Station 1	SO2	TRS	THC	O3	NO	NO2	NOX	PM2.5	ETL	RH	PC	GR	WS	WS_SD	WD	WD_SD
Negative value	950	266	59	24	2105	359	403	1586	0	0	0	2208	0	0	0	0
flag 1	365	365	365	365	365	365	365	0	0	0	0	0	0	0	0	0
flag 2	81	77	74	75	111	111	111	10	0	0	0	0	0	0	0	0
flag 3	7	8	12	4	8	8	8	28	0	0	0	0	4	4	3	3
flag 4	8	8	8	8	8	8	8	8	6	8	7	8	8	8	8	8
flag 5	0	0	0	0	0	0	0	0	0	0	0	0	798	798	65	65
flag 6	1	8	10	0	26	26	26	193	0	0	0	0	0	0	0	0
flag 7	40	41	49	39	42	42	42	45	34	34	7	30	35	35	35	35
flag 8	0	0	0	0	0	0	0	0	0	0	2207	0	0	0	0	0

Note: TRS = total reduced sulfur.
THC = total hydrocarbon.
ETL = temperature.
RH = relative humidity.
PC = precipitation.
WS = wind speed.
WD = wind direction.
NO_x = oxides of nitrogen.
$PM_{2.5}$ = particulate matter with a diameter smaller than 2.5 μm.
WS_SD = wind speed standard deviation.
WD_SD = wind direction standard deviation.
GR = global radiation.

Table 6. **Number of invalid data with negative value and labeled by flag number at AMS #13 (refer to Table 4 for description of flag #).**

Station 13	SO2	TRS	THC	O3	NO	NO2	NOX	PM2.5	ET	WS	WS_SD	WD	WD_SD
Negative value	2763	2473	23	281	2955	679	881	1712	0	13	0	0	0
flag 1	365	362	362	362	365	365	365	0	0	0	0	0	0
flag 2	74	96	57	76	86	86	86	8	0	0	0	0	0
flag 3	6	32	7	16	6	6	6	13	0	0	0	0	0
flag 4	0	0	0	0	0	0	0	0	0	0	0	0	0
flag 5	0	0	15	47	0	0	0	0	0	2	2	9	9
flag 6	3	47	65	1	0	0	0	237	0	0	0	0	0
flag 7	107	105	104	101	112	112	112	107	102	101	101	101	101
flag 8	0	0	0	0	0	0	0	0	0	0	0	0	0

Note: TRS = total reduced sulfur. WS = wind speed. WS_SD = wind speed standard deviation.
THC = total hydrocarbon. WD = wind direction. WD_SD = wind direction standard deviation.
ET = temperature. $PM_{2.5}$ = particulate matter with a diameter smaller than 2.5 μm.
NO_x = oxides of nitrogen.

Table 7. Number of invalid data (with value "-99999") at Fort McMurray airport meteorological monitoring station.

Fort McMurray Airport Station	TCO	TCA
-99999 (invalid)	19	19

Note: TCO = total cloud opacity.
TCA = total cloud amount.

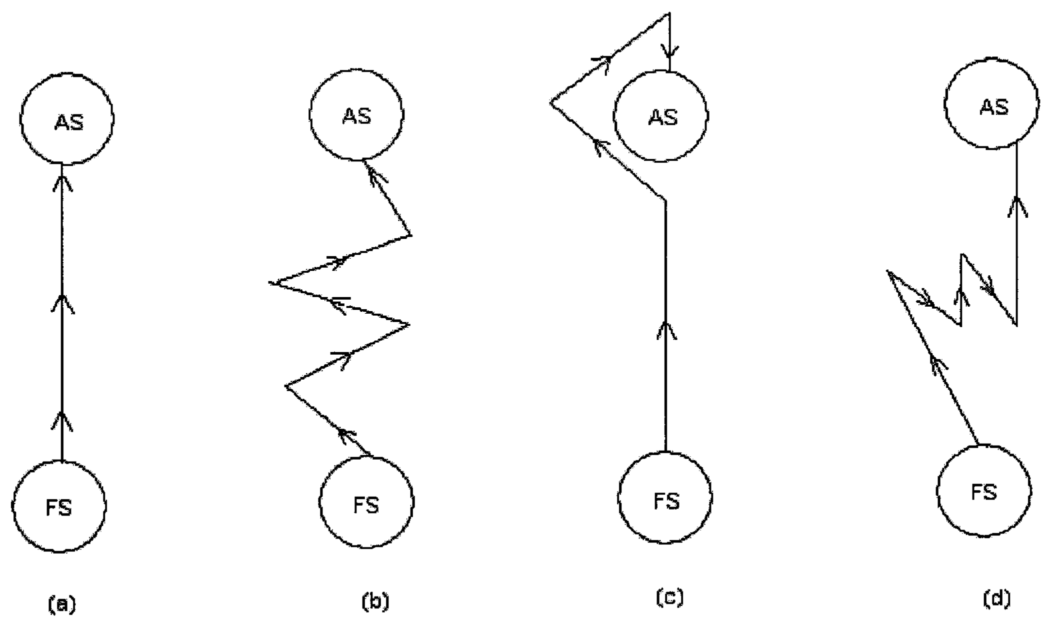


Figure 10. Forest fire generating Particulate matter and various transport pathways (AS = Air Monitoring Station. FS = Forest Fire Site).

According to the Alberta Government, 722 forest fires were recorded in northern Alberta (north of Edmonton) from 1 October 2002 to 31 September 2003. Some of these forest fires were confined in a small area (<1 hectare). Effects of these forest fires were ignored for the sake of minimizing the workload in incorporating these data into the study. There were 133 forest fire incidents that occurred in an area >1 hectare. For each of these fires, the recorded position was pinpointed on a map of northern Alberta. Then for each data pattern within the fire time (between fire start time and extinguish time), the $PM_{2.5}$ concentration at AMS #1 was examined. If $PM_{2.5}$ concentrations from AMS #1 were higher than the 90^{th} percentile value (17.3 ug/m^3); wind speed and direction and relative position of the forest fire were examined to identify whether particulate matter from the forest fire may have been influencing $PM_{2.5}$ concentrations at AMS #1. If this was the case, the hourly data sets between fire start time and extinguish time were deleted from the data set.

3.2 MODEL DEVELOPMENT

3.2.1 Input and Output Parameter Selection

An ANN model will learn from historical data to seek out relationships between input parameters and output parameters. To get a model with good prediction capabilities, inputs and outputs selected must be related to each other. This means that a model constructor must do a thorough examination of a target problem (Baxter et al., 1999). In addition, the number of parameters selected will affect model results. Ratnaweera and Blorm (1995) showed that models with more input parameters have a better ability to make predictions and have a better ability to reduce instrument failure effects than models with less input parameters.

Generally only one parameter is used as an output because a single-output ANN model will be more accurate compared to a multi-output ANN model (Baxter et al., 1999). This approach may not always be the case. Maier et al. (2003) examined the water coagulation process and constructed a model with three outputs: turbidity, color and UVA-254. They also constructed models that used turbidity, color and UVA-254, respectively as single outputs. Results indicated that there were no significant differences in prediction capabilities between these models. Therefore they used the three-output model to make the model simple. This case may have been caused by a strong correlation of the three outputs. Thus, when there are target outputs with strong correlations, a multi-output ANN model might be more beneficial. $PM_{2.5}$ concentration at AMS #1 was the single target output parameter for this ANN study.

Parameters used at the WBEA Air Quality Monitoring Stations

SO_2: SO_2 is a colorless gas with a pungent odor. According to AEP (1996), the most significant sources in Alberta were natural gas processing plants. It was reported that these plants contributed 42% to the total amount of SO_2 emissions in Alberta. Oil sands and power plants contributed 26% and 18%, respectively. Other sources were indicated as gas plant flares, oil refineries, pulp and paper mills and fertilizer plants. In current study domain, oil sands plants were the predominant source. As described previously, SO_2 is also a precursor for formation of secondary $PM_{2.5}$.

TRS: TRS includes hydrogen sulfide, mercaptans, dimethyl sulfide, dimethyl disulfide and other sulfur compounds. According to CASA (2004e), TRS can be produced from oil sands plants and other industrial and natural sources. Because oil sand mining is the main industrial activity in northern Alberta, TRS may be a parameter to represent these activities for ANN modeling.

THC: THC includes a series of compounds containing hydrogen and carbon atoms. Components of THCs can react with oxides of nitrogen to produce ozone in the presence of sunlight (CASA, 2004f). THC can also originate from natural, anthropogenic and industrial sources, like vegetation, vehicle emissions and industrial facilities.

O_3: Instead of being produced directly from human activities, O_3 is produced from complex reactions between oxides of nitrogen and volatile organic compounds (VOCs) in the presence of sunlight. O_3 tends to be higher in spring and summer because of greater hours of sunlight and more vertical mixing from the upper atmosphere (CASA, 2004g). Moreover, rural areas tend to have higher O_3 concentrations than urban areas because in urban areas there is more nitric oxide that can scavenge it. O_3 can represent a precursor to secondary and condensable primary $PM_{2.5}$.

NO, NO_2, NO_x: NO_x includes nitrogen dioxide (NO_2) and nitric oxide (NO). Most of the NO_x in Alberta is produced from transportation sources (CASA, 2004h).

Other sources are oil and gas industry, power plants, gas and fuel combustion in homes, and forest fires. These parameters represent high temperature burning processes that are a principal source of fine particles.

ETL: As described previously, temperature changes with altitude will determine atmospheric stability and the degree of mixing of pollutants in atmosphere. Temperature also represents weather conditions that can affect formation of fine particles. For example, cold temperatures tend to slow down many reactions that occur more-readily at ambient temperatures.

RH: Relative humidity reflects the amount of water vapor in air. This vapor can affect the formation and deposition of fine aerosols.

GR: Global radiation is the total of direct solar radiation and diffuse sky radiation received by a unit horizontal surface. Solar energy received by the earth's surface will control the temperature profile near the earth surface.

WS, WS_SD, WD, WD_SD: Wind speed and wind direction determine the direction and speed of that air pollutants travel. The rate of change of wind speed and wind direction (wind speed standard deviation and wind direction standard deviation) reflect the stability of the air and the corresponding amount of mixing that will occur.

Parameters at The Fort McMurray Airport

TCO: Total cloud opacity represents the incoming solar radiation condition, which is important for the formation of O_3.

TCA: Total cloud amount also represents the incoming solar radiation condition, which is also important for the formation of O_3.

NO (nitric oxide) and PC (precipitation amounts) were not used as inputs. There were too many missing values for these parameters in the original file (as indicated in Tables 5 and 6). All of the other air quality and meteorological parameters may affect formation and transportation of $PM_{2.5}$ in some manner. This situation reflects the complexity of modeling $PM_{2.5}$ and might explain why there are not many published research papers related to this issue.

3.2.2 Model Evaluation Criteria

After training a network, model performance must be evaluated. Generally, there are two criteria to do this. The first criterion is the coefficient of determination (R^2) defined by following equation:

$$R^2 = 1 - \frac{\sum_{i=1}^{i=n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{i=n} (y_i - \bar{y}_i)^2} \tag{2}$$

Where: n is the number of predictions

y_i is the actual output,

$\bar{y}$ is the mean of the predictions, and

$\hat{y}_i$ is the network prediction output.

The R^2 indicates the proportion of variance in the model (dependent) variable – or output variable – that is explained by the input (independent) variables. The R^2 represents a comparison between a real model and a virtual model (Gamal El-Din and Smith, 2002). The virtual model uses the mean of all the samples as its prediction. Results from ANN models with different structures can be directly assessed without confusion using this criterion. When the R^2 value is near one, it means the model will have good prediction capabilities. R^2 value near zero indicates a model will have poor prediction capabilities. Although hard rules do not exist, models with R^2 values <0.5 are not acceptable for making predictions, models with R^2 values from 0.5 to 0.7 can be considered adequate or suitable for making predictions, while models with R^2 values >0.7 tend to be better for making predictions.

A second criterion is the absolute mean error. This criterion is a comparison between an actual prediction and a network prediction (Baxter et al., 1999). Consequently, it can show the magnitude of the inconsistency in a prediction. The absolute mean error can be used as a tool to judge whether a model is suitable for a process control application (Baxter et al., 1999). For example, if an absolute mean error value is within the interval of process adjustment, it will be suitable for process control. Otherwise, it will cause an accuracy problem for process control.

For current research, R^2 was used as the only criterion to evaluate model performance. Pre-screened data sets were separated into a training data set, testing data set, and production data set. The data sets were separated using the following steps:

1. The pre-screened data set was sorted in a descending sequence for the output parameter (hourly $PM_{2.5}$ concentration at AMS #1) in a Microsoft ® Excel file.
2. A new column was added to the file. Cells in this column were filled with a character "T", "P", or "V" in a repeating sequence of T, P, T, V, T, etc. "T" represented training set data, "P" represented testing set data, and "V" represented production set data. Accordingly, 60% of the data sets were assigned for training, 20% of the data sets were assigned for testing, and 20% of the data sets were assigned for production.
3. These data sets were then sorted back to their original sequence.
4. In NeuroShell 2.0, the training set, testing set, and production set data were extracted according to the character "T", "P", or "V" assigned to it.

The pre-screened data sets were separated into training, testing, and production sets in a ratio of 3:1:1 in this manner. At the same time the training, testing, and production sets retained high, average, and low values of the output parameter (hourly $PM_{2.5}$ concentration at AMS #1). Using this approach the NeuroShell software can learn most of the features of the problem domain, train the system accordingly, and obtain an optimum model.

Four sets of R^2 values were computed during the model training and application process ("whole," "training," "testing," and "production"):

- The R^2 values for the "whole" data set represent the extent of ANN model fit in terms of its ability to predict variability in the model output in relation to variability in model inputs for the overall data.
- The R^2 values for the "training" and "testing" data sets represent the extent of interim ANN model fit after initial exposure to (i.e. training and testing of) input data.
- The R^2 values for the "production" data set represent the extent of interim ANN model fit to input data that the model had never seen before. This criterion (R^2 value for the "production" data set) was used as the principal measurement for evaluating model results.

3.2.3 Model Development Step One

3.2.3.1 Methodology

A series of ANN models were initially trained to assess the suitability of using individual input parameters for the current research. After these initial training runs, each model was carefully examined to identify the important contribution factors (i.e. those input parameters that had the largest weighting to the model output). The R^2 values were also examined to evaluate a model's prediction capability. The weighting factors can be used to roughly evaluate the importance of an input parameter in predicting an output, relative to other input parameters in the same network (Ward Systems Group, Inc., 2003). ANN models that were constructed are presented in Table 8 and discussed below:

Model 1: Model 1 used all of the parameters described in Section 3.2.1 as inputs.

Model 2: This model had all of the parameters used in model 1 plus a time variable "hour of day" as inputs. The objective of this variable was to relate emission activities and possibly meteorological conditions to hour of day. The 24 hours of a day were transformed into a "0 to 23" numerical system. For example, if there were valid data obtained at 4:00 pm, an input parameter "16" was added. If there were valid data obtained at 11:00 pm, "23" was added as an input.

Model 3: This model had nine optimum parameters as inputs (optimum 1). These optimum parameters were based on analysis of the top nine contribution factors in model #1 and model #2.

Model 4: This model had 15 optimum parameters as inputs (optimum 2).

Model 5: Only air quality parameters from AMS # 1 station (Fort McKay) and the Fort McMurray weather station were used as inputs in this model.

Table 8. Computed R^2 values and relative factor weightings for ANN models constructed.

Model #	Model Inputs Description		R^2 Value				Importance of Contribution Factor			
			Whole data set	Training set	Testing set	Production set	Biggest	Second	Third	Four
1	All input parameters		0.53	0.6	0.43	0.43	$PM_{2.5}$[2]	SO_2[2]	WS[1]	O_3[1]
2	All input parameters with hour of day		0.58	0.66	0.46	0.46	$PM_{2.5}$[2]	O_3[1]	Hour	TCO[3]
3	Optimum 1 input parameters		0.49	0.55	0.38	0.44	$PM_{2.5}$[2]	O_3[1]	RH[1]	Hour
4	Optimum 2 input parameters		0.58	0.67	0.5	0.39	RH[1]	Hour	$PM_{2.5}$[2]	O_3[1]
5	Input parameters from AMS #1 only		0.30	0.32	0.30	0.26	ET[1]	TCO[3]	NO_2[1]	O_3[2]
6	Input parameters from AMS #13 only		0.45	0.5	0.37	0.37	$PM_{2.5}$[2]	O_3[2]	WS_SD[2]	WS[2]
7	All input parameters with hourly forest fire data deleted		0.59	0.70	0.55	0.30	$PM_{2.5}$[2]	RH[1]	O_3[2]	Hour
8a	Classified by season (two categories)	Oct to Apr	0.46	0.45	0.47	0.48	ET[1]	ET[2]	TCO[3]	$PM_{2.5}$[2]
8b		May to Sep	0.72	0.88	0.55	0.41	O_3[1]	Hour	RH[1]	O_3[2]
9a	Classified by meteorological parameters using Kohonen network (two categories)	Category #1 (1233)	**0.63**	**0.67**	**0.61**	**0.47**	WD[1]	$PM_{2.5}$[2]	O_3[2]	O_3[1]
9b		Category #2 (761)	**0.58**	**0.60**	**0.54**	**0.55**	Hour	WS[1]	RH[1]	$PM_{2.5}$[2]
10a	Classified by meteorological parameters using Kohonen network (three categories)	Category #1 (1015)	0.59	0.59	0.52	0.69	Hour	$PM_{2.5}$[2]	RH[1]	WD[2]
10b		Category #2 (463)	0.72	0.78	0.73	0.52	Hour	RH[1]	$PM_{2.5}$[2]	O_3[1]
10c		Category #3 (516)	0.43	0.53	0.23	0.26	$PM_{2.5}$[2]	Hour	RH[1]	WD[1]
11a	Classified by air quality parameters using Kohonen network (two categories)	Category #1 (948)	0.57	0.70	0.32	0.44	$PM_{2.5}$[2]	RH[1]	WS_SD[1]	Hour
11b		Category #2 (1046)	0.62	0.70	0.46	0.56	RH[1]	$PM_{2.5}$[2]	O_3[1]	WS[1]
12a	Classified by air quality parameters using Kohonen network (three categories)	Category #1 (384)	0.56	0.72	0.29	0.35	RH[1]	$PM_{2.5}$[2]	Hour	WS_SD[2]
12b		Category #2 (859)	0.41	0.49	0.27	0.35	$PM_{2.5}$[2]	Hour	O_3[2]	GR[1]
12c		Category #3 (751)	0.72	0.83	0.53	0.61	Hour	O_3[1]	$PM_{2.5}$[2]	RH[1]

[1] Factor at AMS #1 (Fort McKay station).
[2] Factor at AMS #13.
[3] Factor at Fort McMurray Airport weather station.

Model 6: Only air quality parameters from AMS # 13 were used as inputs in this model.

Model 7: The data for input parameters used in model # 2 were screened to delete sets related to forest fires (as described in Section 3.1.2).

Models 8a and 8b: The data for input parameters used in model # 2 were classified into two categories according to temperature/seasonal conditions. The categorized data were fed into two sub-models (8a and 8b) individually. The first category included data from October 2002 to April 2003. A majority of the hourly temperature values were below zero during this period. The second category included data from May 2003 to September 2003. A majority of the hourly temperature values were above zero during this period.

Models 9a and 9b: The data for input parameters used in model # 2 were classified into two categories by a Kohonen network according to meteorological parameters. These categorized data were fed into two sub-models (9a and 9b). Meteorological parameters used in the Kohonen network were: temperature, relative humidity, global radiation, wind speed, wind speed standard deviation, wind direction, and wind direction standard deviation.

Models 10a, 10b and 10c: The data for input parameters used in model # 2 were classified into three categories by a Kohonen network according to same meteorological parameters as those used in models 9a and 9b.

These categorized data were fed into three sub-models (10a, 10b and 10c).

Models 11a and 11b: The data for input parameters used in model # 2 were classified into two categories by a Kohonen network according to air quality parameters. These categorized data were fed into two sub-models (11a and 11b). Air quality parameters used in the Kohonen network were: SO_2, O_3, NO_x, and $PM_{2.5}$.

Models 12a, 12b and 12c: The data for input parameters used in model # 2 were classified into three categories by the Kohonen method according to the same air quality parameters as those used in models 11a and 11b. These categorized data were fed into three sub-models 12a, 12b and 12c.

All of these models were used with the following default software settings in the training (and learning) process:

- Pattern selection method: Rotation
- Weight update method: Turboprop
- Activation function for hidden layer: Logistic
- Activation function for output layer: Logistic
- Number of hidden layer neurons: May differ for each model
- Number of learning epochs: May differ for each model

3.2.3.2 Results and Discussion

Models #1 thru #6

Compared with model #1, model #2 had a higher R^2 value for production, testing, training and the whole data set (Table 8). By adding a time variable (hour of day), model prediction capabilities and overall fit were improved. As a result, all subsequent models had the time variable as an input parameter.

Models #3 and #4 – only used the top 9 and 15 optimum parameters, respectively identified in model #2 as inputs – had lower R^2 values than model #2. Only these optimum variables was worse for model performance. Models #5 and #6 only used air quality parameters from separate monitoring stations. R^2 values for the production data set from these two models were 0.26 and 0.37, which were much more lower than R^2 values obtained for model #2.

Results from models #3 thru #6 illustrated that prediction of ambient $PM_{2.5}$ concentrations is very complex. For example, using only optimum input parameters or a subset of all the input parameters was not successful in improving model fit. As a result, all of the parameters were used as inputs for subsequent model development steps.

Generally, the weighting of individual input parameters from each model cannot be compared among models. However, if an input parameter has high relative weighting in most of the models, it can be considered an important input variable for modeling the output parameter. The weighting factor for each parameter was examined after the training process for each model. The four most important input parameters in terms of weighting are indicated in Table 8 for each model. $PM_{2.5}$ concentration at AMS #13 was among the four most important input factors in terms of relative weighting for all models constructed, except for model #5 (which only used parameters from AMS #1). This demonstrates that $PM_{2.5}$ concentrations from AMS #13 were a very important parameter for modeling $PM_{2.5}$ concentrations at AMS #1. The next most important input parameters contributing to model fit in all of the models constructed were "hour of day" and "relative humidity at AMS #1."

Model #7

This model was used to examine the possible influence of forest fires by deleting data sets that corresponded to forest fire occurrences. Table 8 indicates a poorer model suggesting that removal of these data decreased ANN model prediction capabilities. Another aspect to deal with forest fire data that was not tried here would be to use all the data sets and to add another input parameter to categorize presence/absence of a forest fire (i.e. 1 for forest fire present and 0 for forest fire absent).

Models #8 Thru #12

The original data were classified into two or three categories in the models before they were used as inputs. The objective was to separate data with different properties and to improve model prediction capabilities and overall fit. The Kohonen network automatically classified the data sets for models #9 thru #12:

- When original data were classified into two categories, R^2 values increased for all sub-models and all data sets compared to model #2 (e.g. models #9a/b, #11a/b).
- When original data were classified into three categories, R^2 values increased for one or two of the sub-models compared to model #2. The R^2 value for other sub-model(s) was much lower than for model #2. This was observed for models #10a/b/c and #12a/b/c.
- When original data were classified by meteorological parameters using the Kohonen network, R^2 values increased (models #9a/b).
- When original data were classified by season, R^2 values were lower than those classified by meteorological parameters and in two categories (e.g. model #7 compared to models #9a/b).
- When original data were separated into two categories, R^2 values were similar for those classified by meteorological parameters and air quality parameters (e.g. models #9a/b compared to models #11 a/b).

3.2.3.3 Conclusion

Based on the above results and interpretations, it was found that using meteorological or air quality parameters as categorization criteria and separating data sets into two groups by a Kohonen network resulted in sub-models with the best overall fit for the conditions tested (i.e. models #9a/b and #11a/b). This suggests that the Kohonen network could learn the features of original data more precisely than using intuitive knowledge. Meteorological factors are acknowledged as being important in influencing the movement of air masses and the dispersion behavior of pollutants in the atmosphere (Wark et al., 1998). Consequently, models #9a/b were used for further construction and development in the remainder of this study.

3.2.4 Model Development Step two

3.2.4.1 Methodology

In order to further optimize a chosen ANN model for making predictions, other important aspects that are important for model constructors to resolve are answers to the following questions:

1. What is the optimal pattern selection method and weight update method?
2. What is the optimal activation function in the hidden layer and output layer?
3. How many hidden layer neurons should be used?

4. When the original data set is classified into two categories, what is the minimum number of epochs required for ANN model learning?

These questions relate to determining the optimum software settings for running the ANN models. Answering these questions is very important. For example, when training a model, if the training process is stopped with too few epochs passed, the system cannot learn all of the important properties. An incomplete learning problem will be produced in this situation. On the other hand, if the learning process is stopped after too many epochs pass, an over-learning problem may exist (Gamal El-Din and Smith, 2002). This means that the model output will fit exactly to the training data set. When the model is used to make predictions about a data set which it has not seen before, the results will be poor.

Baxter et al. (1999) suggested using a factorial design method to obtain answers to the above questions. This is a statistical design approach with results indicating an optimal direction for changing the magnitude of an ANN software setting, but it cannot tell a model constructor what is the best value for that factor.

Another method indicated by Gamal El-Din and Smith (2002) is a systematic method. When these authors develop an ANN model, they use a multi-layer feed forward back propagation network with a single hidden layer, which is the most popularly structure used today. Then the hidden layer size is increased from three nodes to 10 nodes in one-node increment each time.

The hidden layer size with the best result (highest R^2 value) is selected as the optimal one. By doing this, the lowest hidden layer size in a model with good converging and generalizing ability is achieved. Optimal training epochs are decided in the same way (increased from 10 to 1000 epochs). In the end, a model with a simple structure and optimal results can be constructed (Gamal El-Din and Smith, 2002).

A model constructor can program a series of runs with the NeuroShell 2 software using the systematic approach described above and set the software to run in batch mode. This systematic approach was used for further development of models #9a/b.

3.2.4.2 Results and Discussion

The data for input parameters used in models #9a/b were separated into two categories by the Kohonen network using meteorological parameters as classification criteria. The category #1 data set included 1,233 data patterns and the category #2 data set included 761 data patterns. Optimized ANN sub-models for the category #1 and #2 data sets were developed individually according to the following steps:

1. Data pattern selection method and weight update method.

The available software settings for data pattern selection methods were Rotation and Random. The available software settings for weight update methods were Vanilla, Momentum and TurboProp. The possible bonds between these data pattern selection and weight update methods are listed in Table 9. Resulting R^2 values for sub-models with different bonds are displayed in Figure 11 (category #1 data set) and Figure 12 (category #2 data set).

For the category #1 data set, use of a bond of Momentum weight update and Rotation pattern selection method was able to obtain the best R^2 value for the production data set (Figure 11). For the category #2 data set, use of a bond of TurboProp weight update and Rotation pattern selection method was able to obtain the best R^2 value for the production data set (Figure 12). These software settings were fixed for the next model development step.

Table 9. Different bonds of weight update methods and data pattern selection methods for ANN modeling.

Setting No.	Weight Update Method	Data Pattern Selection Method
1	Momentum	Rotation
2	Momentum	Random
3	Vanilla	Rotation
4	Vanilla	Random
5	TurboProp	Rotation

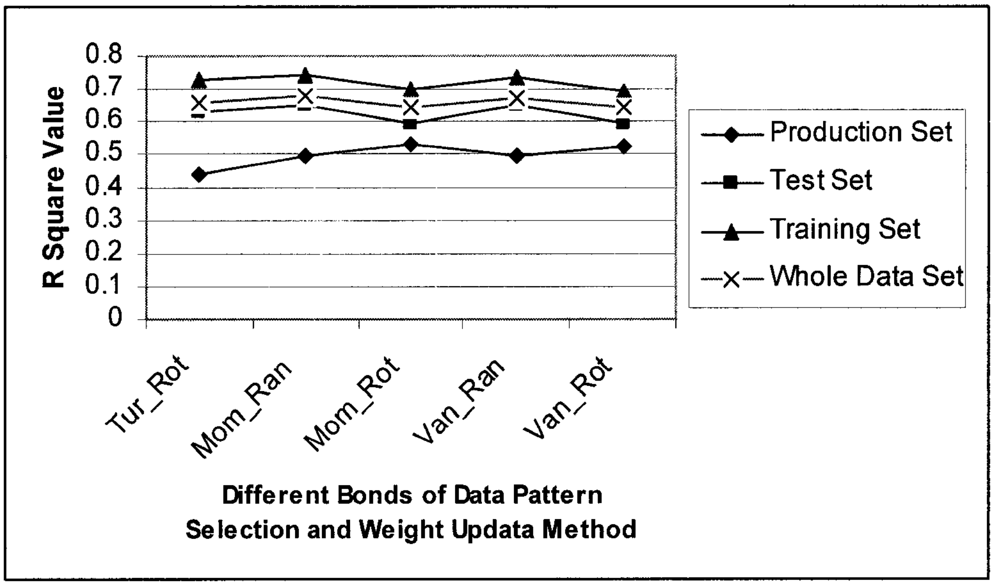


Figure 11. **R^2 values versus data pattern selection and weight update method settings for ANN sub-model using category #1 data set.**

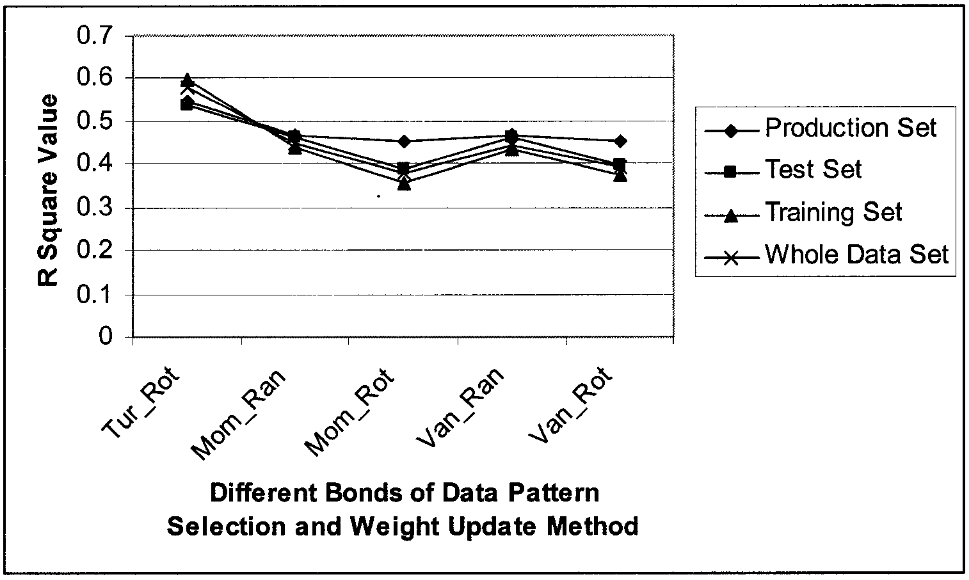


Figure 12. **R^2 values versus data pattern selection and weight update method settings for ANN sub-model using category #2 data set.**

2. **Hidden layer and output layer activation function settings.**

Different settings related to the activation function were evaluated. The available activation functions were (Ward Systems Group, Inc., 2003):

- Logistic
- Symmetric Logistic
- Tanh
- Linear
- Tanh15
- Gaussian
- Gaussian Complement
- Sine

Because the hidden layer and output layer can have different activation functions, there were a lot of possible settings. Resulting R^2 values for the different activation function settings are shown in Table 10 (category #1 data set) and Table 11 (category #2 data set). For the category #1 data set, the best R^2 value for the production set was obtained when the setting was "t15_syl," which means that the hidden layer activation function was "Tanh15" and the output layer activation function was "Systemetric Logistic" (Table 10). For the category #2 data set, the best R^2 value for production set was obtained when the setting was "log_t15," which means that the hidden layer activation function was "Logistic" and output layer activation function was "Tanh15" (Table 11). These software settings were fixed for the next model development step.

3. Amount of hidden layer neurons.

The number of hidden layer neurons evaluated for the category #1 data set were: 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 74, 78 and 82. The numbers of hidden layer neurons evaluated for the category #2 data set were: 30, 32, 34, 36, 38, 40, 44, 48, 52, 56, 60, 64, 68, 72.

R^2 values for models with these settings are shown in Figure 13 (category #1 data set) and Figure 14 (category #2 data set). Figure 13 shows that for the sub-model using category #1 data, R^2 values started increasing from 30 hidden layer neurons and peaked at 46 hidden layer neurons. For the sub-model using category #2 data, the highest R^2 value was obtained by using 36 hidden layer neurons (Figure 14). The sub-model may get the same R^2 value when different numbers of hidden layer neurons were used. At this condition fewer hidden layer neurons was preferred in order to save model running time. These software settings (46 hidden layer neurons for the model using the category #1 data set and 36 hidden layer neurons for the model using the category #2 data set) were fixed for the next model development step.

4. Number of learning epochs.

For the sub-model using the category #1 data set, learning epochs ranged from 50 to 1600 in intervals of 100. For the sub-model using the category #2 data set, learning epochs ranged from 100 to 1500 in intervals of 100. Resulting R^2 values are shown in Figure 15 (category #1 data set) and Figure 16 (category #1 data set).

Table 10. R^2 values for different activation function settings in ANN sub-model using category #1 data set.

Activation Function	Production Set	Test Set	Training Set	Whole Data Set
tan_log	0.529	0.591	0.701	0.648
tan_lin	-6.156	-5.696	-5.493	-5.654
tan_tan	0.429	0.497	0.462	0.463
tan_gau	0.539	0.504	0.580	0.557
tan_t15	0.220	0.219	0.212	0.215
tan_syl	0.515	0.564	0.598	0.576
tan_gco	0.499	0.521	0.534	0.525
log_log	0.489	0.522	0.577	0.550
log_lin	-0.100	-0.095	-0.093	-0.095
log_tan	0.380	0.441	0.491	0.461
log_gau	0.459	0.483	0.540	0.514
log_t15	0.453	0.282	0.405	0.389
log_syl	0.480	0.547	0.603	0.569
log_gco	0.412	0.407	0.447	0.432
gau_log	0.510	0.567	0.793	0.697
gau_lin	0.411	0.330	0.445	0.416
gau_tan	0.286	0.411	0.496	0.441
gau_gau	0.312	0.455	0.575	0.503
gau_t15	-0.966	-0.588	-0.745	-0.754
gau_syl	0.521	0.521	0.753	0.664
gau_gco	0.337	0.318	0.369	0.353
t15_log	0.539	0.596	0.735	0.672
t15_lin	-0.153	-0.134	-0.130	-0.135
t15_tan	0.448	0.411	0.503	0.475
t15_gau	0.431	0.527	0.784	0.669
t15_t15	-0.869	-0.482	-0.716	-0.697
t15_syl	**0.544**	**0.538**	**0.674**	**0.623**
t15_gco	0.035	0.307	0.019	0.079
syl_log	0.478	0.581	0.630	0.593
syl_lin	0.465	0.476	0.474	0.473
syl_tan	0.429	0.549	0.538	0.521
syl_gau	0.518	0.553	0.601	0.576
syl_t15	0.457	0.452	0.486	0.474
syl_syl	0.489	0.601	0.641	0.605
syl_gco	0.513	0.539	0.547	0.539
gco_log	0.377	0.586	0.818	0.692
gco_lin	0.193	0.257	0.174	0.194
gco_tan	0.504	0.565	0.707	0.641
gco_gau	0.420	0.586	0.708	0.631
gco_t15	0.350	0.277	0.291	0.299
gco_syl	0.426	0.579	0.730	0.644
gco_gco	0.425	0.500	0.592	0.544

Table 11. R^2 values for different activation function settings in ANN sub-model using category #2 data set.

Activation Function	Production Set	Test Set	Training Set	Whole Data Set
tan_log	0.468	0.522	0.592	0.555
tan_lin	0.444	0.454	0.582	0.531
tan_tan	0.486	0.519	0.678	0.611
tan_gau	-0.524	-0.349	-0.628	-0.552
tan_t15	0.403	0.489	0.494	0.476
tan_syl	0.482	0.584	0.794	0.695
tan_gco	0.344	0.478	0.670	0.572
log_log	0.546	0.537	0.598	0.577
log_lin	0.413	0.600	0.568	0.546
log_tan	0.458	0.545	0.587	0.555
log_gau	0.480	0.499	0.567	0.538
log_t15	**0.611**	**0.539**	**0.655**	**0.623**
log_syl	0.509	0.529	0.624	0.584
log_gco	0.520	0.546	0.581	0.563
gau_log	0.287	0.445	0.596	0.510
gau_lin	0.004	0.193	0.676	0.456
gau_tan	0.384	0.425	0.772	0.632
gau_gau	-1.134	-0.518	0.103	-0.247
gau_t15	0.112	0.283	0.716	0.519
gau_syl	0.329	0.341	0.675	0.545
gau_gco	-0.518	-0.247	0.701	0.289
t15_log	0.535	0.535	0.581	0.563
t15_lin	0.378	0.411	0.496	0.457
t15_tan	0.401	0.531	0.593	0.545
t15_gau	-3.252	-1.364	-1.789	-1.968
t15_t15	0.377	0.439	0.735	0.611
t15_syl	0.471	0.569	0.653	0.603
t15_gco	-0.172	0.019	0.602	0.344
syl_log	0.535	0.535	0.581	0.563
syl_lin	0.508	0.554	0.548	0.542
syl_tan	0.470	0.475	0.551	0.521
syl_gau	-1.799	-1.409	-1.383	-1.464
syl_t15	0.488	0.547	0.756	0.665
syl_syl	0.503	0.529	0.636	0.590
syl_gco	0.479	0.541	0.584	0.556
gco_log	0.339	0.496	0.635	0.553
gco_lin	0.091	0.263	0.887	0.617
gco_tan	0.349	0.440	0.752	0.616
gco_gau	0.022	0.274	0.658	0.465
gco_t15	0.225	0.306	0.712	0.542
gco_syl	0.469	0.470	0.640	0.575
gco_gco	-0.080	0.159	0.721	0.462

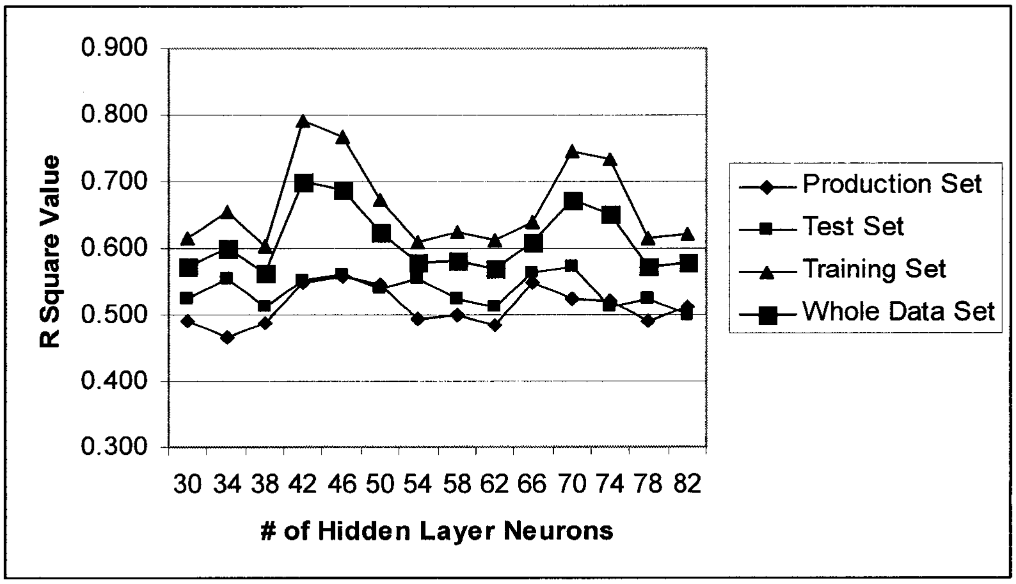


Figure 13. **R^2 values versus number of hidden layer neurons for ANN sub-model using category #1 data set.**

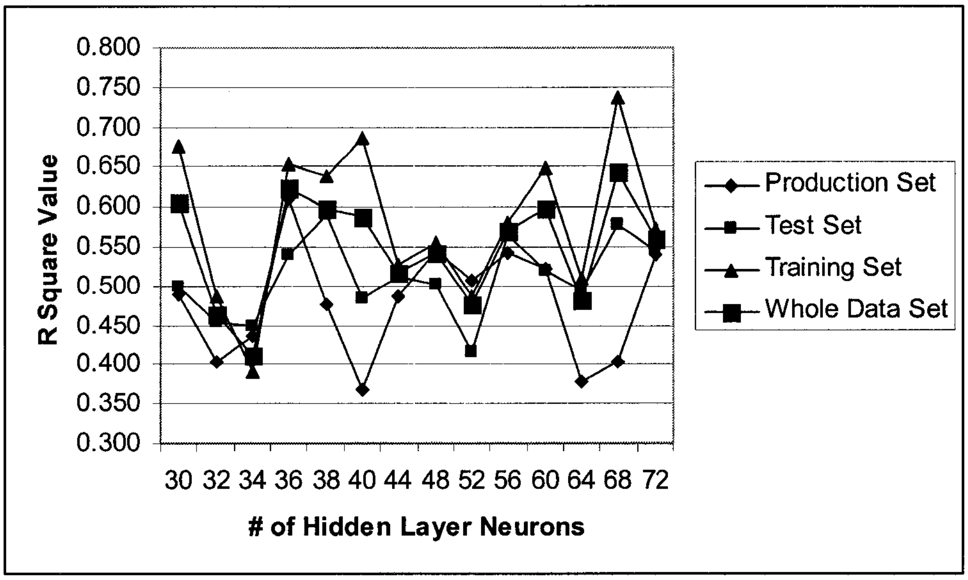


Figure 14. **R^2 values versus number of hidden layer neurons for ANN sub-model using category #2 data set.**

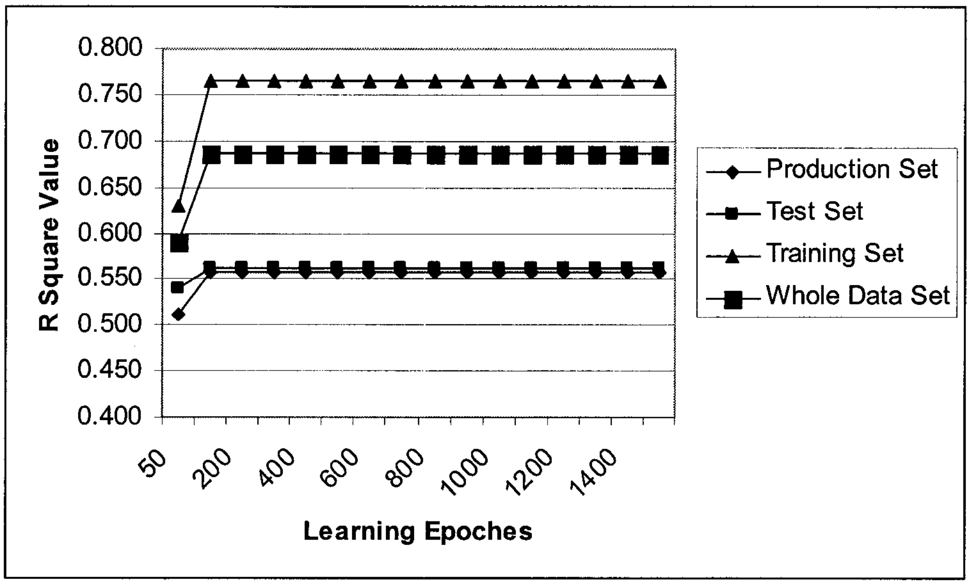


Figure 15. **R^2 values versus learning epochs for ANN sub-model using category #1 data set.**

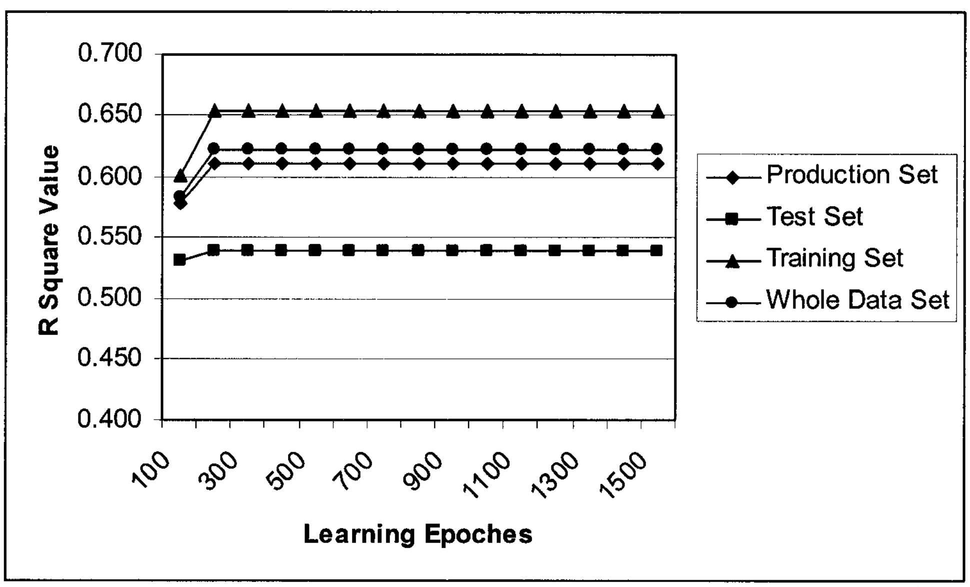


Figure 16. **R^2 values versus learning epochs for ANN sub-model using category #2 data set.**

Figure 15 shows that for the sub-model using the category #1 data set, R^2 values peaked at a software setting of one hundred learning epochs and did not improve with more learning epochs. For the sub-model using the category #2 data set, this peak occurred at two hundred learning epochs (Figure 16).

3.2.4.3 Conclusion

The resulting optimum software setting for the ANN sub-models using category #1 and #2 data are listed below. For the ANN sub-model using the category #1 data set:

- Pattern selection method: Rotation
- Weight update method: Momentum
- Activation function for hidden layer: Tanh15
- Activation function for output layer: Symmetric Logistic
- Number of hidden layer neurons: 46
- Number of learning epochs: 100

For the ANN sub-model using the category #2 data set:

- Pattern selection method: Rotation
- Weight update method: TurboProp
- Activation function for hidden layer: Logistic
- Activation function for output layer: Tanh15
- Number of hidden layer neurons: 36
- Number of learning epochs: 200

4.0 MODEL CHECK AND APPLICATION

4.1 MODEL STABILITY AND OVERALL FIT

4.1.1 Model Stability

After optimizing the two ANN sub-models, the test and production data sets were exchanged and the models were re-trained and applied. Results from this exercise were compared with results from models that were trained with the original data sets in order to check the stability of the models. If similar R^2 values are observed, it indicates that the models are stable and one would expect to get similar results when the models process different data. Model instability – as evidenced by different R^2 values after exchanging test data sets and production data sets – indicates that further refinements in the models are necessary.

Results of exchanging test and production data sets for the optimized ANN sub-models are shown in Table 12. Table 12 results indicate comparable R^2 values (rounded to two significant digits) after data exchange and re-training and indicate stable models were constructed. The optimum ANN sub-model using category #2 data exhibited slightly less stability than the optimum ANN sub-model using category #1 data as R^2 values decreased slightly (Table 12).

Table 12. Comparison of R^2 for optimum ANN sub-models after training with original and exchanged production data and testing data sets.

Model applied data		R^2 for production data set	R^2 for testing data set	R^2 for training data set	R^2 for whole data set
Category #1	Original	0.56	0.56	0.77	0.69
	Exchanged	0.56	0.56	0.77	0.69
Category #2	Original	0.61	0.54	0.65	0.62
	Exchanged	0.53	0.61	0.65	0.62

4.1.2 Overall Model Fit

NeuroShell 2 software creates a file that can be exported and used by other software programs (e.g. Microsoft Excel) to make predictions. In this case it was used to make predictions about $PM_{2.5}$ concentrations based on available input data. Resulting predicted versus actual $PM_{2.5}$ concentrations are shown Figures 17 and 18 below.

Figure 17 shows predicted versus actual $PM_{2.5}$ concentrations for the ANN sub-model using category #1 data corresponding to $PM_{2.5}$ levels ranging from 1 to 23 μg/m^3. These levels correspond to the "typical range of fluctuation" observed in $PM_{2.5}$ levels at AMS #1. Figure 18 shows predicted versus actual $PM_{2.5}$ concentrations for the ANN sub-model using category #1 data corresponding to $PM_{2.5}$ levels ranging from 1 to 80 μg/m^3. These levels correspond to a much greater range and it encompassed peak concentration events.

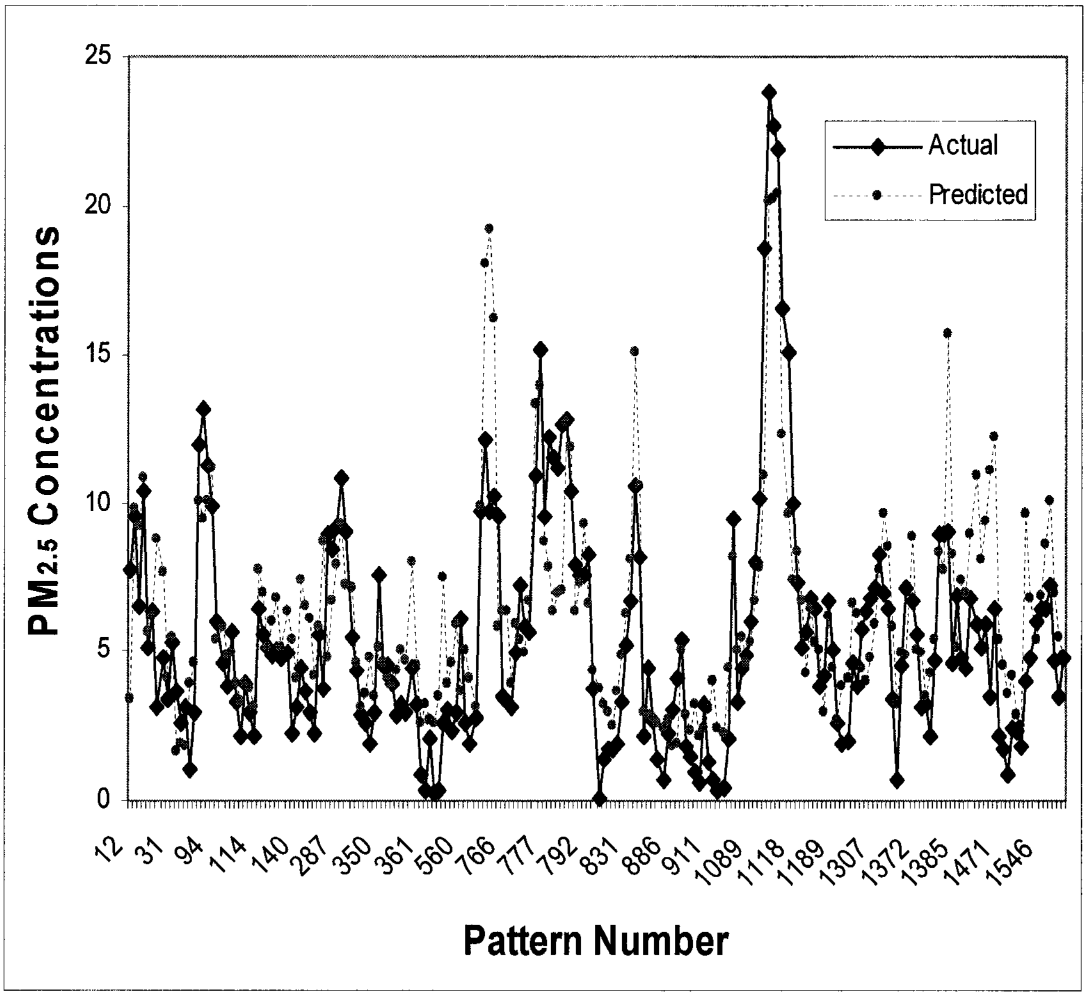


Figure 17. Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [$\mu g/m^3$] for ANN sub-model using category #1 data (205 data patterns) corresponding to typical range of fluctuation observed in $PM_{2.5}$ levels.

As shown in Figures 17 and 18, the ANN sub-model using category #1 data was able to predict the magnitude of $PM_{2.5}$ concentrations (e.g. pattern #'s 1100 to 1200 in Figure 17; pattern #'s 5650 to 5750 in Figure 18). The ANN sub-model using category #1 data was also able to respond to increasing and decreasing $PM_{2.5}$ concentrations (e.g. pattern #'s 800 to 900 in Figure 17; pattern #'s 5775 to 5975 in Figure 18). There are numerous instances where the ANN sub-model both underestimates (i.e. predicted level < actual level) and overestimates (i.e. predicted level > actual level) in these figures.

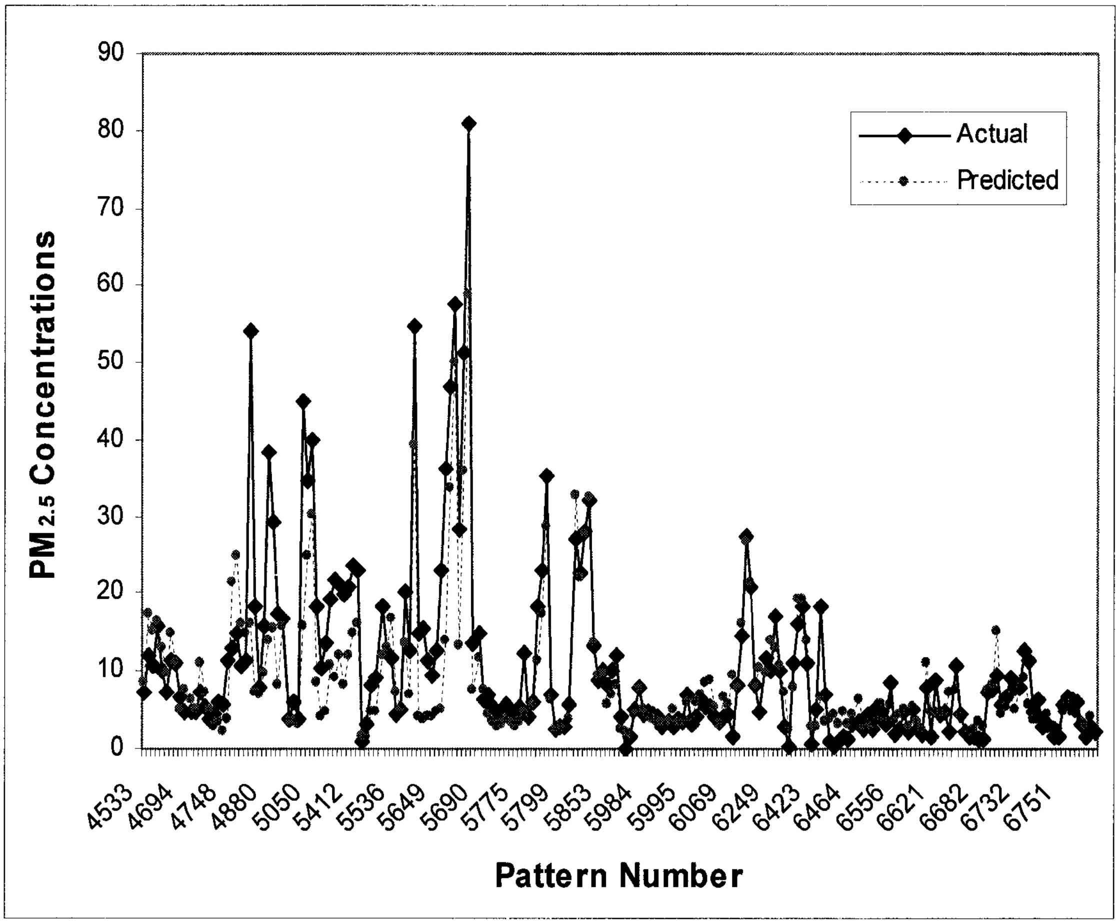


Figure 18 Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [μg/m^3] for ANN sub-model using category #1 data (205 data patterns) corresponding to a range encompassing peak concentration events.

Figure 19 shows predicted versus actual $PM_{2.5}$ concentrations for the ANN sub-model using category #2 data corresponding to $PM_{2.5}$ levels ranging from 1 to 26 μg/m^3. These levels correspond to the "typical range of fluctuation" observed in $PM_{2.5}$ levels at AMS #1. Figure 20 shows predicted versus actual

PM2.5 concentrations for the ANN sub-model using category #2 data corresponding to PM2.5 levels ranging from 1 to 75 µg/m^3. Similar to Figure 18, this range encompassed peak concentration events.

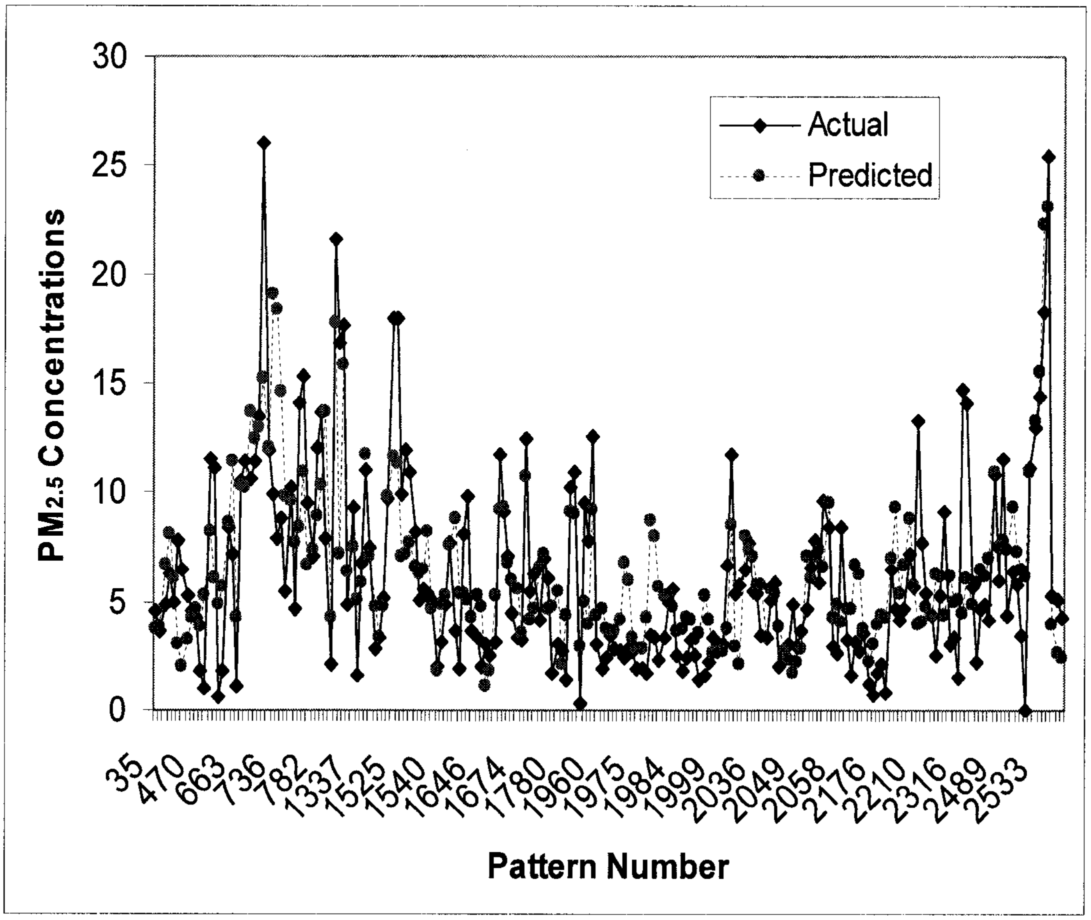


Figure 19 Actual versus predicted hourly PM2.5 concentrations at Fort McKay [µg/m^3] for ANN model using category #2 data (205 data patterns) corresponding to typical range of fluctuation observed in PM2.5 levels.

Similar to the ANN model using category #1 data, Figures 19 and 20 show that the ANN sub-model using category #2 data was able to predict the magnitude of PM2.5 concentrations – however not as well (e.g. pattern #'s 680 to

780 in Figure 19; pattern #'s 5650 to 5750 in Figure 20). The ANN sub-model using category #2 data was also able to respond to increasing and decreasing $PM_{2.5}$ concentrations (e.g. pattern #'s 2300 to 2600 in Figure 19; pattern #'s 4800 to 5200 in Figure 20). There are numerous instances where the ANN sub-model both underestimates (i.e. predicted level < actual level) in Figure 19.

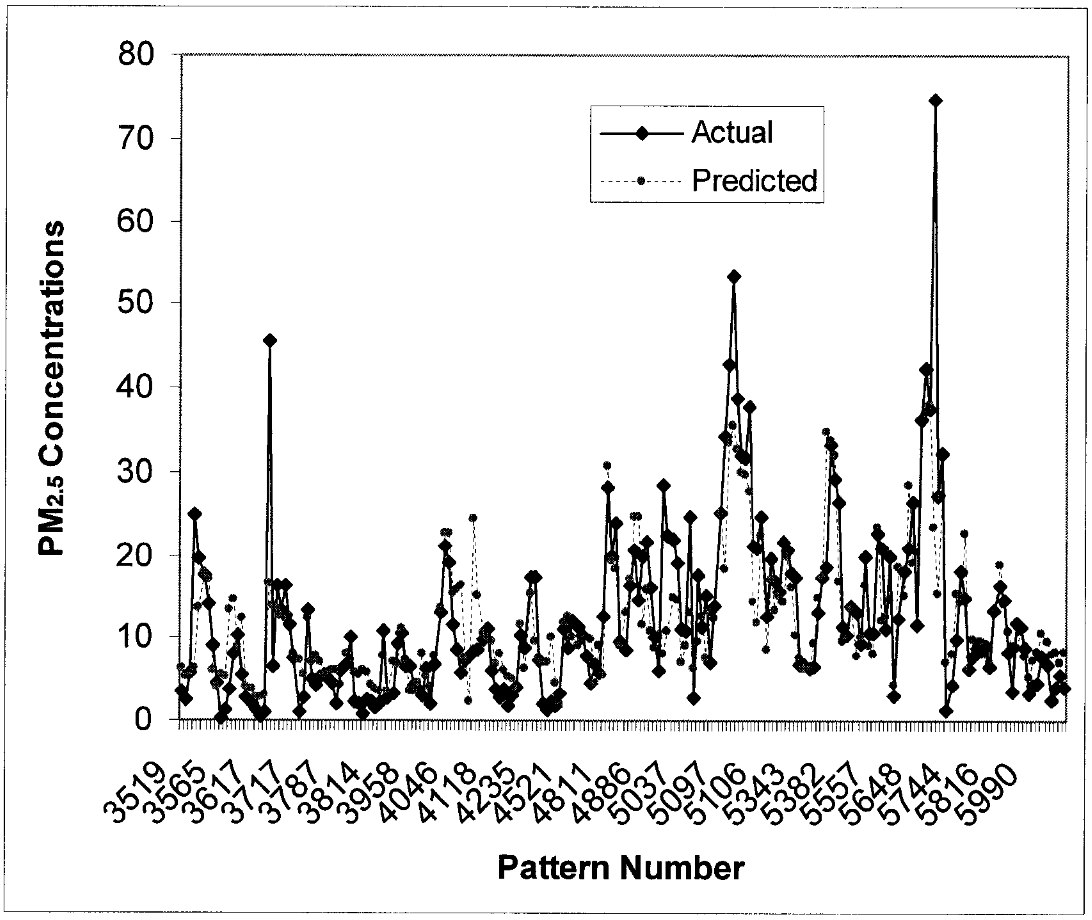


Figure 20 **Actual versus predicted hourly $PM_{2.5}$ concentrations at Fort McKay [μg/m³] for ANN model using category #2 data (205 data patterns) corresponding to a range encompassing peak concentration events.**

An unfavourable attribute of both sub-models observed in Figures 17 through 20 was that they were unable to agree exactly with actual $PM_{2.5}$ concentrations for high (>50 μg/m^3) values. This suggests that there are other unexplained factors contributing to hourly $PM_{2.5}$ concentrations at the Fort McKay station that were not taken into account in the modeling process. The absence of influences from local and/or near-field sources including road dust emissions, vehicle exhaust, residential wood burning associated with home-heating in the community and near the station may be an explanation.

4.2 SENSITIVITY ANALYSIS

Two sensitivity tests were undertaken to examine whether the ANN models exported to Microsoft Excel would respond to simulated increases in several of the important input factors. These input factors consisted of air quality parameters from AMS #13 that represented the "near-field" influence of PM-generation activities at the Syncrude Canada Ltd. site. If increases in "near-field" air quality (i.e. at AMS #13) occur as a result of future increases in PM-generation activities at the Syncrude Canada Ltd. Site, the issue of whether similar responses would occur at the Fort McKay air monitoring station could be addressed.

The first sensitivity test involved increasing $PM_{2.5}$ and O_3 concentrations at AMS #13. The concentrations of these two air pollutants were separately increased by a factor of 25% and 50% and the resulting $PM_{2.5}$ concentration at the Fort McKay station (AMS #1) was examined. $PM_{2.5}$ concentration at AMS #13 was shown to be an important input parameter for the ANN sub-models adopted (Table 8).

The second sensitivity test similarly involved increasing concentrations of $PM_{2.5}$ at AMS #13 by a factor of 25% and 50% and observing the resulting (predicted) $PM_{2.5}$ concentration at the Fort McKay station (AMS #1). However, this test was simulated for winds blowing from a direction with minor $PM_{2.5}$ generation from oil sand and other undefined activities and compared to the condition when winds blow from directions with oil sand and other undefined activities.

Sensitivity Test Based on Increased $PM_{2.5}$ and O_3 Concentrations at AMS #13

Results of the first sensitivity test are presented in Figures 21 through 28 and discussed below. Figure 21 shows the predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using the ANN sub-model based on category #1 data for a 25% simulated increase in $PM_{2.5}$ and O_3 concentrations at AMS #13. Figure 22 shows the predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using the ANN sub-model based on category #1 data for a 50% simulated increase in $PM_{2.5}$ and O_3 concentrations at AMS #13.

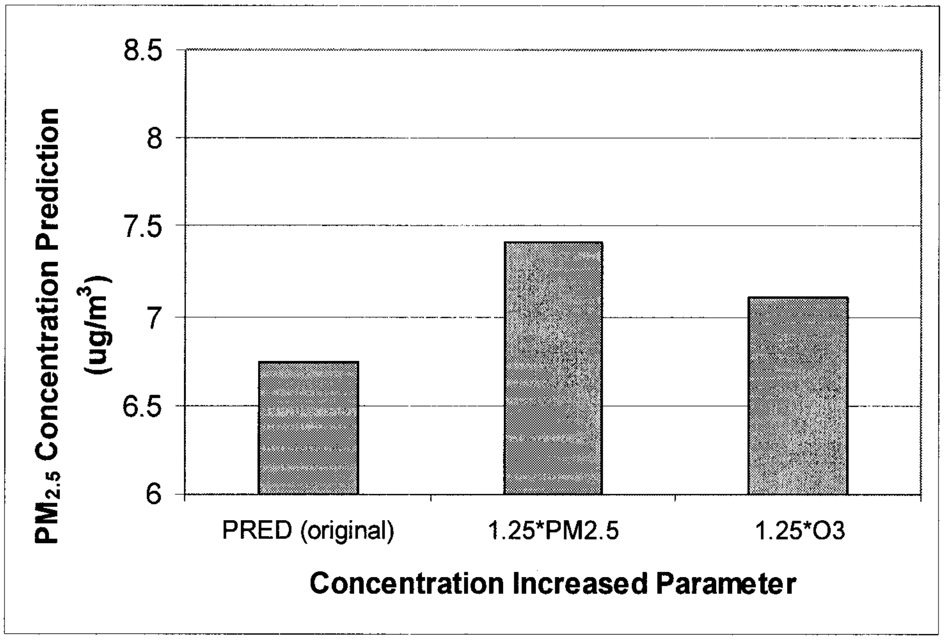


Figure 21 **Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 25% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.**

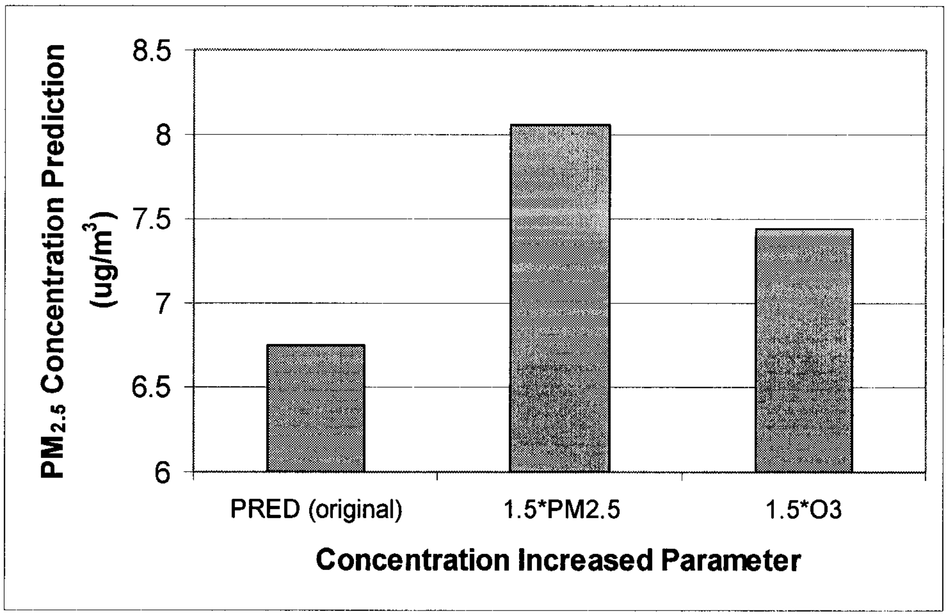


Figure 22 **Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.**

Figure 23 shows $PM_{2.5}$ concentrations at Fort McKay station for the ANN sub-model using category #1 data (205 data patterns) and a 50% increase in $PM_{2.5}$ concentrations at AMS #13 for a typical concentration range at the Fort McKay station (1 to 21 $\mu g/m^3$). Predicted increases of up to 33% were observed.

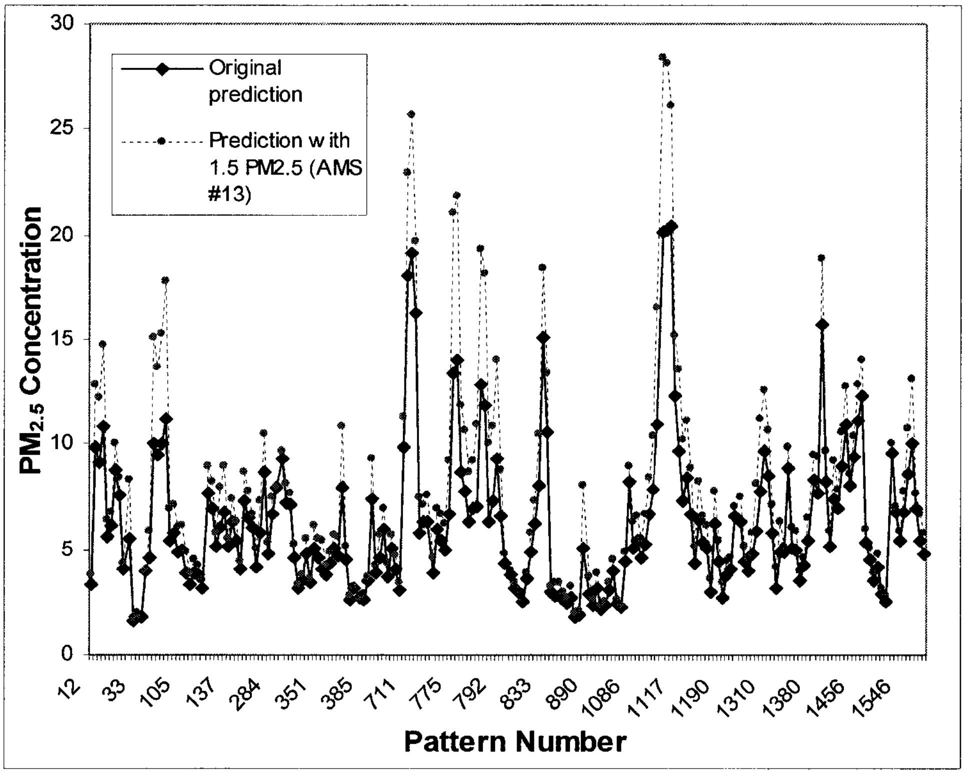


Figure 23 **$PM_{2.5}$ concentrations at Fort McKay station [$\mu g/m^3$] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.**

Figure 24 shows $PM_{2.5}$ concentrations at Fort McKay station for the ANN sub-model using category #1 data (205 data patterns) and a 50% increase in $PM_{2.5}$ concentrations at AMS #13 for a typical concentration range at the Fort McKay station (1 to 60 $\mu g/m^3$). Predicted increases of up to 20% were observed.

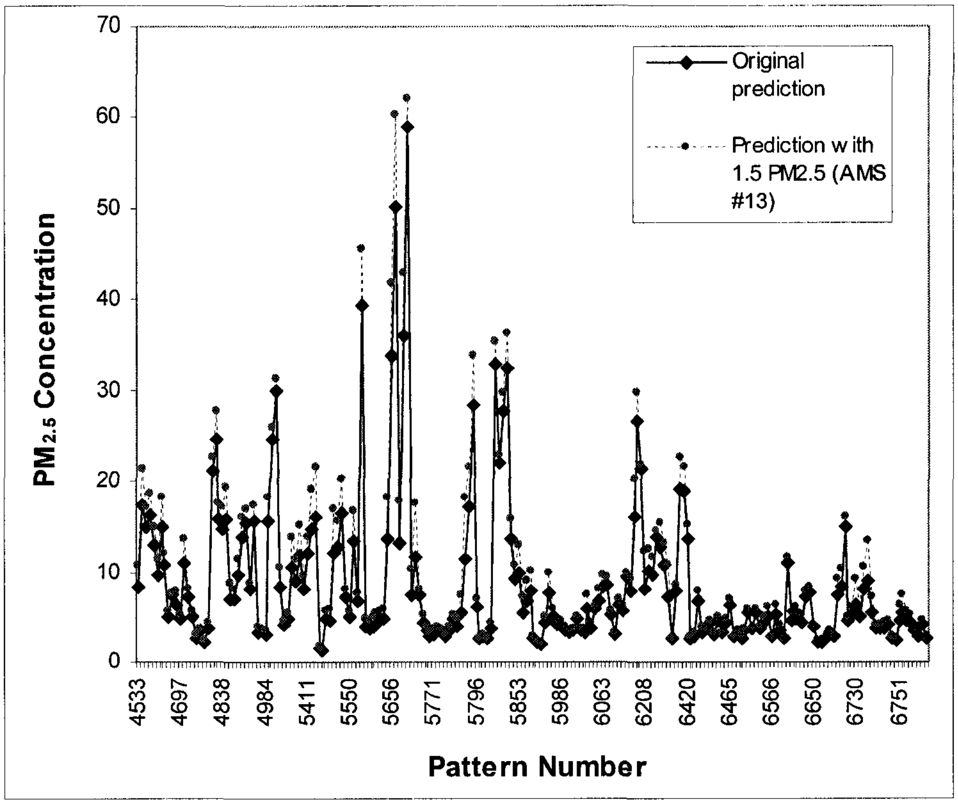


Figure 24 **$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.**

Figure 25 shows the predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using the ANN sub-model based on category #2 data for a 25% simulated increase in $PM_{2.5}$ and O_3 concentrations at AMS #13. Figure 26 shows the predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using the ANN sub-model based on category #2 data for a 50% simulated increase in $PM_{2.5}$ and O_3 concentrations at AMS #13.

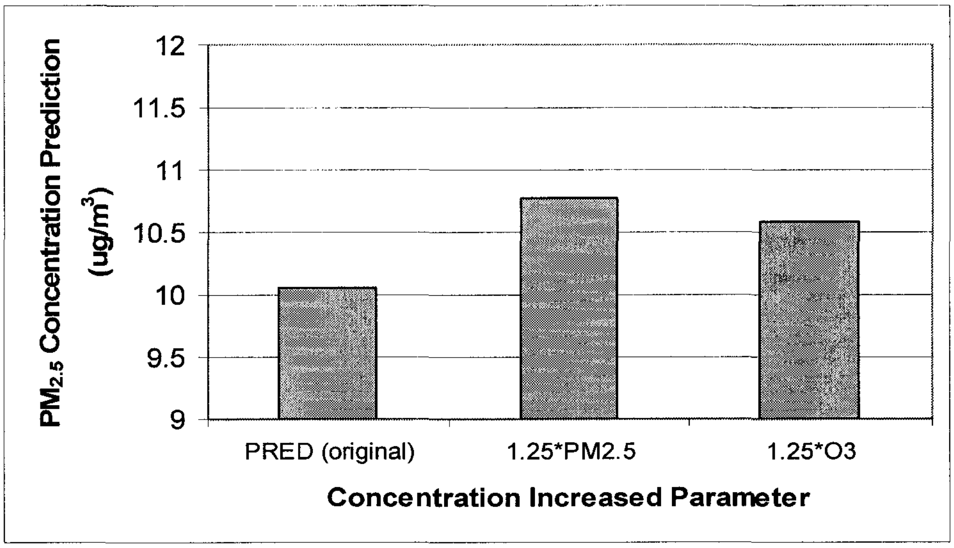


Figure 25 **Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 25% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.**

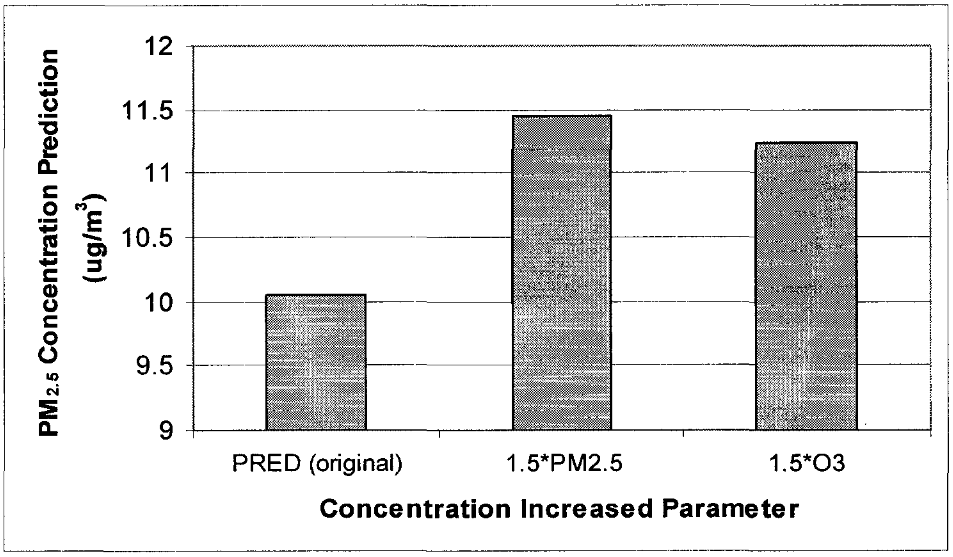


Figure 26 **Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 50% increase in $PM_{2.5}$ and O_3 concentration at AMS #13.**

Figure 27 shows $PM_{2.5}$ concentrations at Fort McKay station for the ANN sub-model using category #2 data (205 data patterns) and a 50% increase in

$PM_{2.5}$ concentrations at AMS #13 for a typical concentration range at the Fort McKay station (1 to 23 $\mu g/m^3$). Predicted increases of up to 35% were observed.

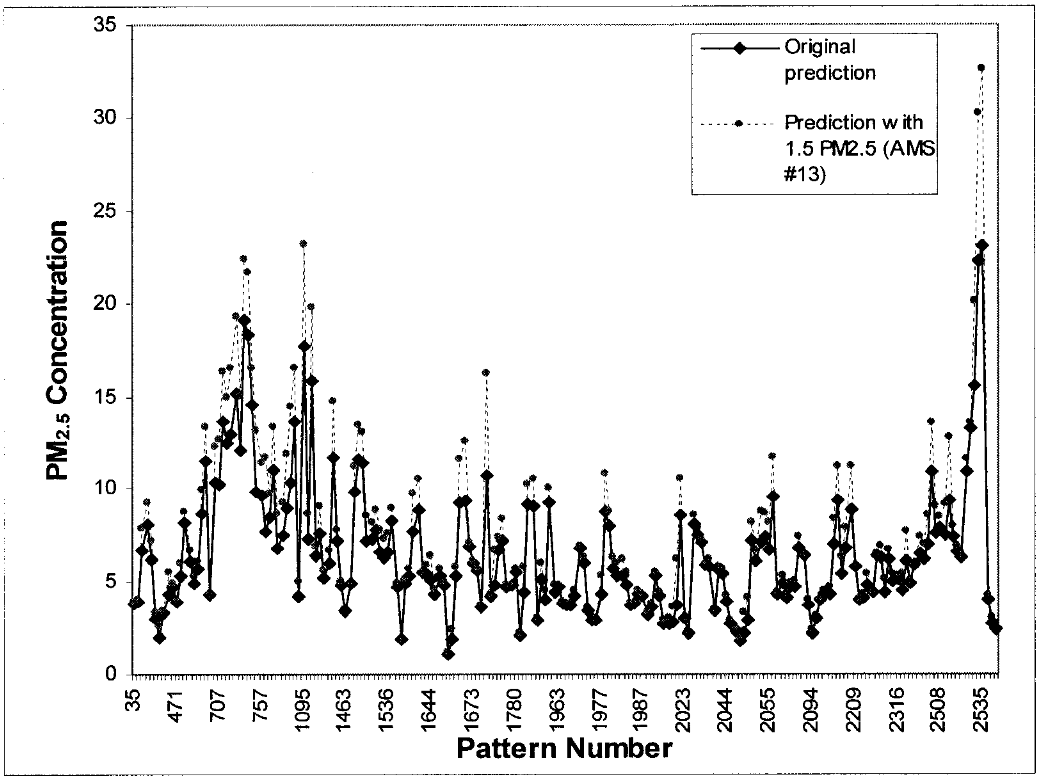


Figure 27 **$PM_{2.5}$ concentrations at Fort McKay station [$\mu g/m^3$] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.**

Figure 28 shows $PM_{2.5}$ concentrations at Fort McKay station for the ANN sub-model using category #1 data (205 data patterns) and a 50% increase in $PM_{2.5}$ concentrations at AMS #13 for a typical concentration range at the Fort McKay station (1 to 38 $\mu g/m^3$). Predicted increases of up to 20% were observed. Predicted $PM_{2.5}$ concentrations at Fort McKay station for both ANN sub-models and a 25% increase in $PM_{2.5}$ concentrations at AMS #13 are presented in Appendix D.

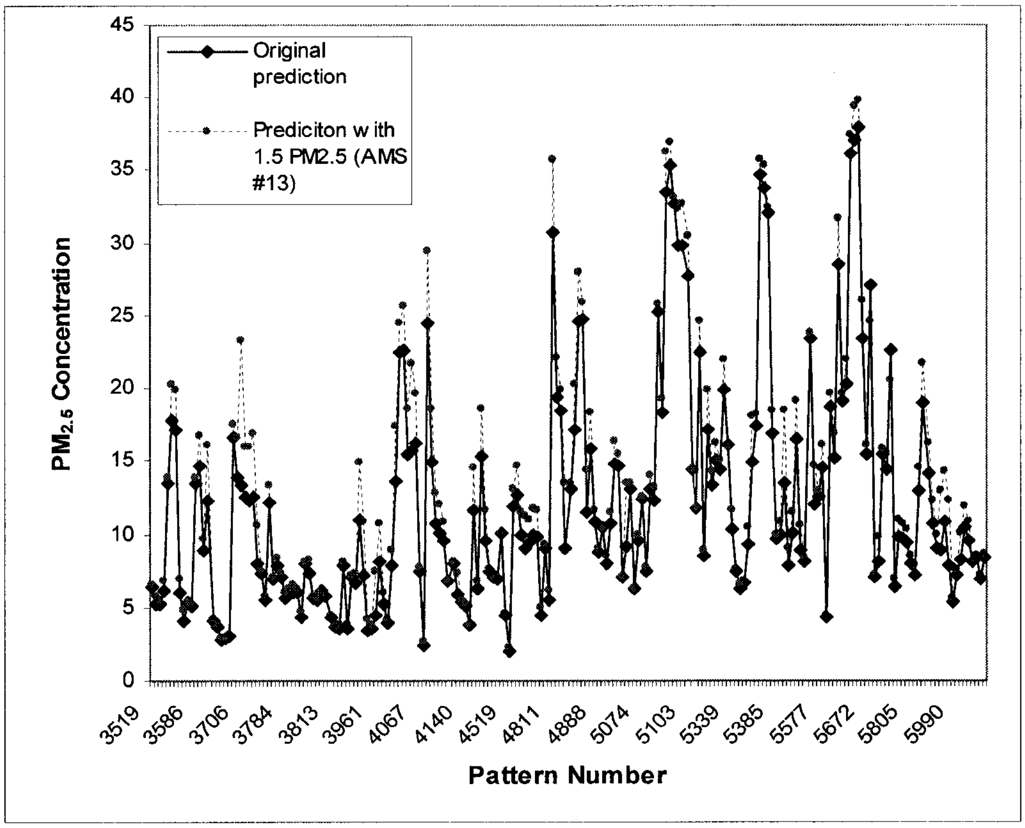


Figure 28 **$PM_{2.5}$ concentrations at Fort McKay station [µg/m^3] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 50% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.**

Figure 22 (ANN sub-model based on category #1 data) and Figure 26 (ANN sub-model based on category #2 data) predict that hourly mean $PM_{2.5}$ concentrations at Fort McKay will increase by up to 1.5 µg/m^3 when $PM_{2.5}$ concentrations at AMS #13 are increased by 50%. Changes to hourly $PM_{2.5}$ concentrations associated with peak concentration events >30 µg/m^3 (Figures 24 and 28) are ~5 µg/m^3. A general implication of these observations is that hourly $PM_{2.5}$ concentrations at Fort McKay would only be expected to increase by small amounts (i.e. up to 5 µg/m^3). This is not considered a large increase.

<u>Sensitivity Test Based on Different Wind Directions</u>

The original data were separated into six groups – each corresponding to winds blowing from within 60°-angles – as illustrated in Figure 29. Using the two ANN sub-models, $PM_{2.5}$ concentrations were predicted for three 60°-angles ranges shown in Figure 29: winds blowing within a 120 to 180° angle; 180 to 240° angle; and 240 to 300° angle.

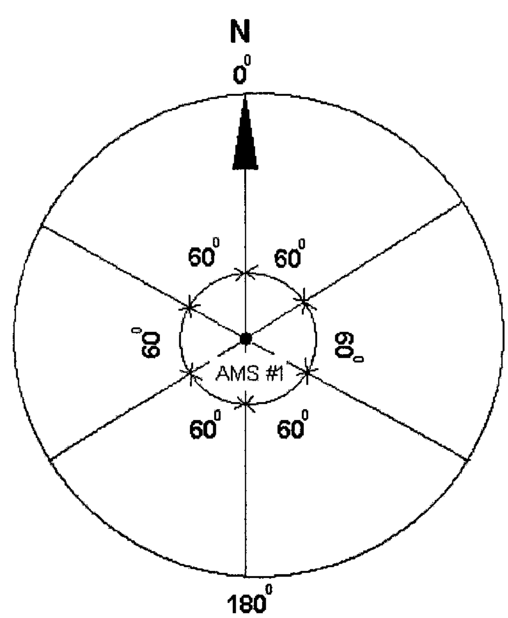


Figure 29 WBEA Fort McKay Station showing 60° angles used for examination of wind direction and corresponding $PM_{2.5}$ concentrations.

The reason for comparing model results among the three 60°-angles ranges is based upon the following rationale. Winds blowing within a 240 to 300° angle were assumed to represent the nearest to "background" conditions in terms of upwind $PM_{2.5}$ sources affecting the Fort McKay station (AMS #1) compared to winds blowing from other directions. A simplistic assumption was made that winds blowing from directions within the 240 to 300°-angle range shown in Figure 29 pass over land with minor $PM_{2.5}$ generation from oil sand and other undefined activities compared to winds are blowing from the other directions.

Several oil sand operations exist upwind of the Fort McKay station (AMS #1) within the 240 to 300° angle range. This is shown in Figure 30 after Harper (2004). However these operations are sub-surface Steam Assisted Gravity Drainage (SAGD) projects that do not have $PM_{2.5}$ emissions of the same magnitude as surface oil sand mining, and extraction and refining operations. These latter types of operations exist upwind of the 120 to 180°-angle range (e.g. Suncor upgrader) and upwind of the 180 to 240°-angle range (e.g. Syncrude Mildred Lake upgrader and open pit mine).

Actual $PM_{2.5}$ concentrations at the Fort McKay station were examined to confirm whether concentrations for winds blowing from directions within the 240 to 300°-angle range (shown in Figure 29) were less than concentrations for winds blowing from all directions. This comparison is shown in Figure 31. Several observations can be made of Figure 31. $PM_{2.5}$ concentrations at the Fort McKay station (AMS #1) were clearly lower when winds blow from within the 240 to 300°-angle range. For example, the 50th percentile concentration was 2.5 μg/m^3 when winds blow from within the 240 to 300°-angle range compared to 4 μg/m^3 when winds blow from all directions. The 90th percentile concentration was 9 μg/m^3 when winds blow from within the 240 to 300°-angle range compared to 12.5 μg/m^3 when winds blow from all directions.

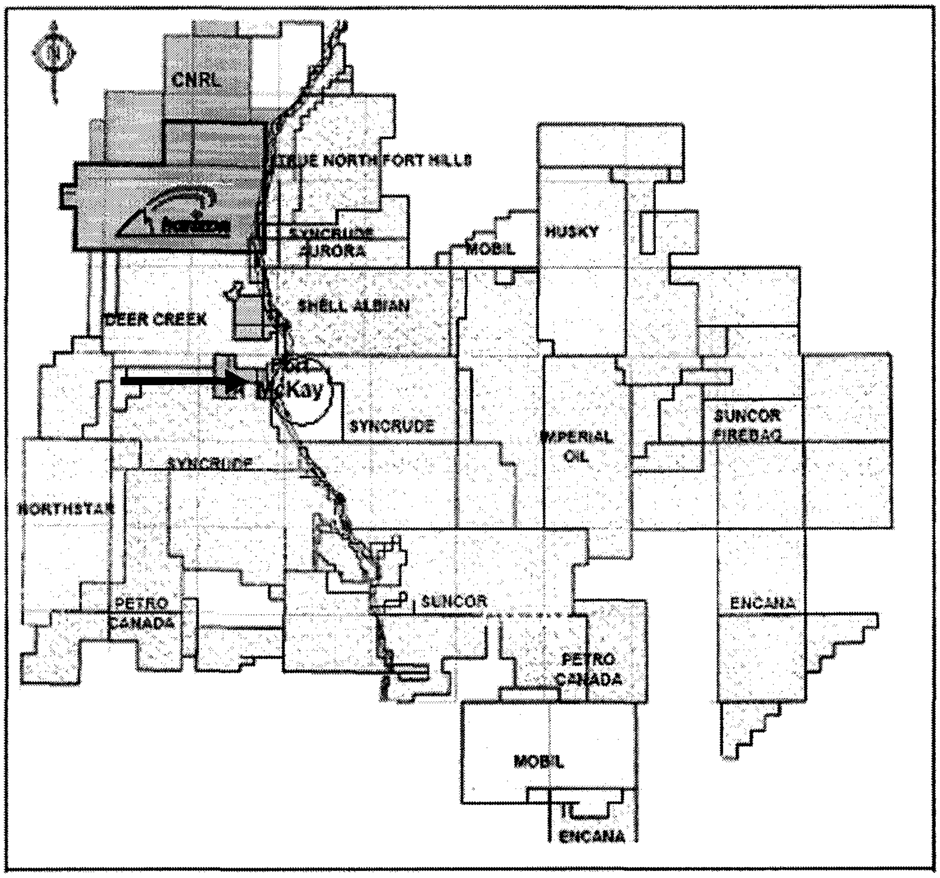


Figure 30 Location of oil sand mining leasehold areas around Fort McKay (adapted from Canadian Natural Resource Limited, 2002).

The other observation to be made of Figure 31 is that the slope of both curves beyond the 95% percentile concentration is virtually the same. This suggests that peak hourly $PM_{2.5}$ concentration events at the Fort McKay station (e.g. concentrations >15 to 20 µg/m^3) are unrelated to any specific wind direction.

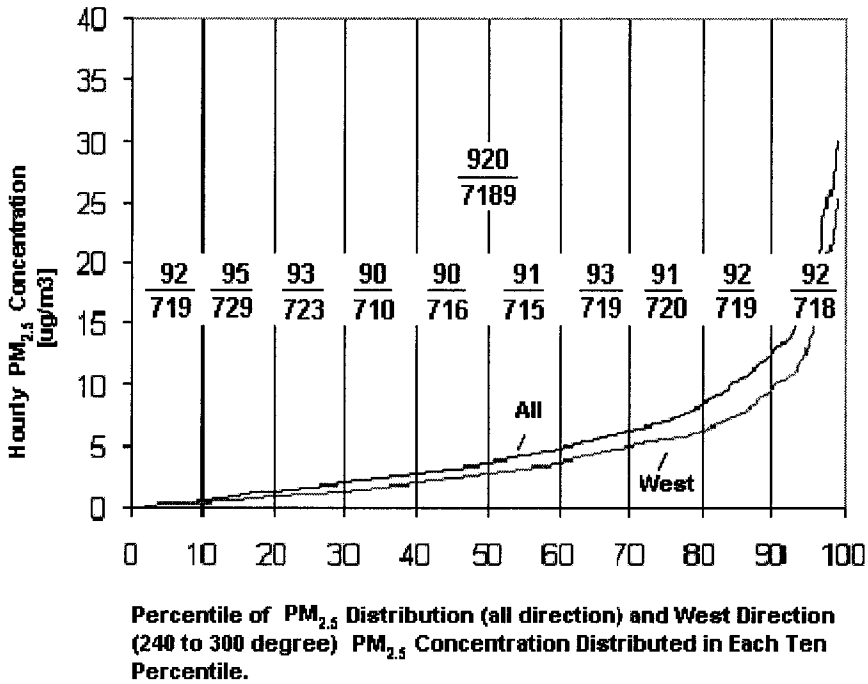


Figure 31 **Cumulative distribution of hourly average $PM_{2.5}$ concentrations at Fort McKay station.**
(All = winds blowing from all directions; West = winds blowing from within the 240 to 300° angle range).

Results of the sensitivity test are presented in Figures 32 and 33 for the ANN sub-models using category #1 data and category #2 data, respectively. Both figures indicate that the magnitude of mean hourly $PM_{2.5}$ concentrations for the original, 25% and 50% increase conditions are the smallest for winds blowing from within the 240 to 300°-angle range. This finding was expected.

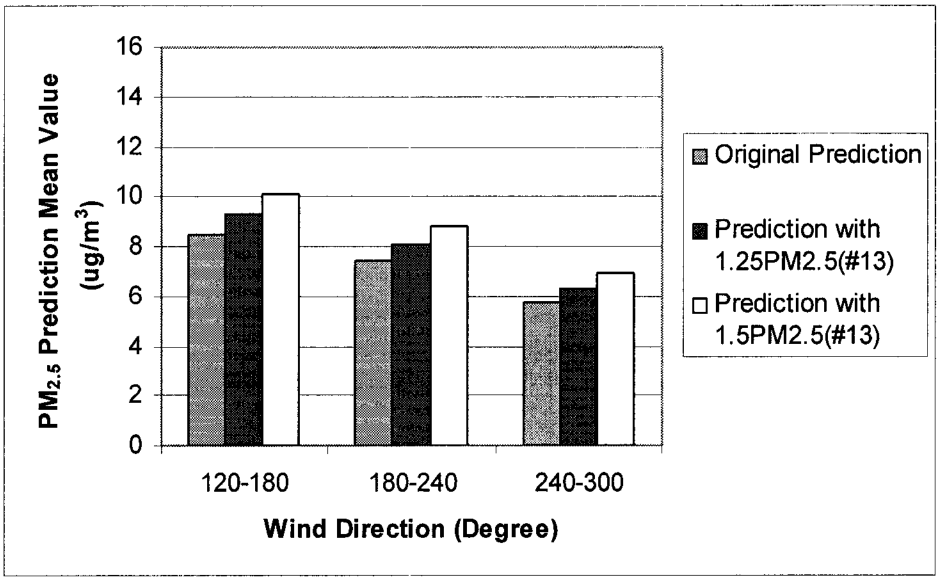


Figure 32 Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #1 data for 25% and 50% increase in $PM_{2.5}$ concentration at AMS #13 and for winds originating from three 60°-angles ranges.

Another finding that was expected was that the mean hourly $PM_{2.5}$ concentration increased regardless of wind direction. AMS #13 is located approximately 4.4 km south of AMS #1 (Appendix A). When winds blow directly from the south towards the Fort McKay station (e.g. from within a narrow 175 to 185°-angle range), a true source-to-receptor trajectory relationship should exist between AMS #13 and AMS #1. However when winds blow from within a much broader range (i.e. from within a 120 to 300° angle range), it is possible that both AMS #1 and AMS #13 are responding to $PM_{2.5}$ concentrations in the same air mass carried by wind. As a result, $PM_{2.5}$ concentrations observed at both stations could be related.

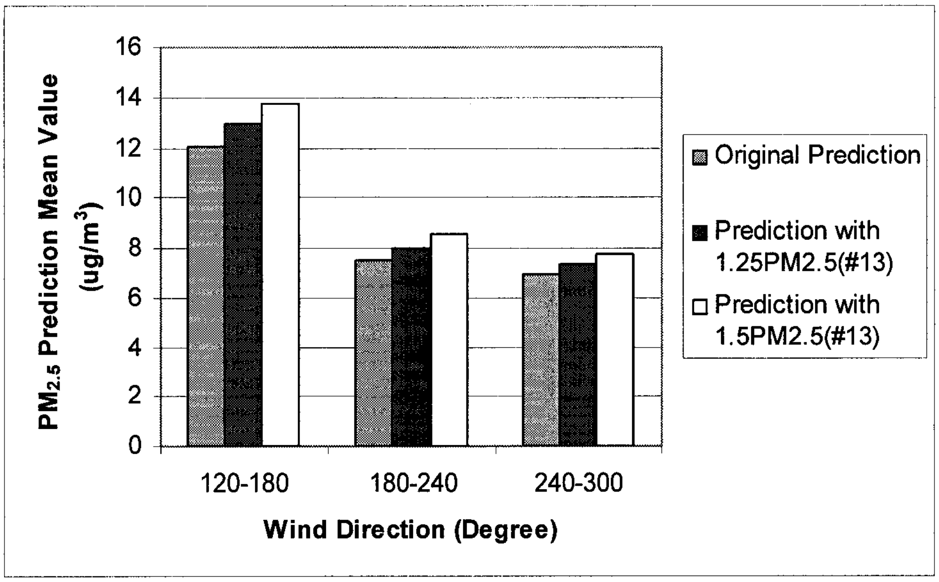


Figure 33 Predicted mean hourly $PM_{2.5}$ concentration at the Fort McKay station using ANN sub-model based on category #2 data for 25% and 50% increase in $PM_{2.5}$ concentration at AMS #13 and for winds originating from three 60°-angles ranges.

Figure 32 (ANN sub-model based on category #1 data) and Figure 33 (ANN sub-model based on category #2 data) predict that hourly mean $PM_{2.5}$ concentrations at Fort McKay will increase by <2 μg/m^3 when $PM_{2.5}$ concentrations at AMS #13 are increased by up to 50% for winds originating from three 60°-angles ranges. This magnitude of this increase is not considered large.

5.0 FINDINGS AND RECOMMENDATIONS

Surface mining and other related PM-generation activities at the Syncrude Canada Ltd. Oil sands site are perceived as having a direct influence on resulting ambient $PM_{2.5}$ concentrations at the Fort McKay air monitoring station. An ANN model study was undertaken to examine whether and to what extent this influence occurs:

- The first objective of this study was to evaluate the feasibility of using ANN to predict ambient $PM_{2.5}$ concentrations at the Fort McKay air monitoring station.
- Upon development of an optimum ANN model structure, a second objective of this study was to conduct a sensitivity analysis. The sensitivity analysis examined whether the optimized ANN model would respond to simulated increases in several of the important input factors.

The following findings are noted:

1. Among the most important input parameters contributing to all of the ANN models evaluated was the "$PM_{2.5}$ concentration measured at AMS #13 (at the Syncrude Canada Ltd. oil sands site)". The implication of this is that $PM_{2.5}$ concentrations measured at the Fort McKay air monitoring station are related to concentrations measured at the Syncrude Canada Ltd. oil sands site. A parameter that took into account "hour of day" was also

useful for improving the prediction capabilities of the ANN models evaluated.

2. A Kohonen network that classified input data into two categories according to meteorological parameters was shown to provide the best ANN model structure in terms of performance. These meteorological parameters included: temperature, relative humidity, global radiation, wind speed, standard deviation of wind speed, wind direction, and standard deviation of wind direction. Two ANN sub-models constructed were able to relate 55% and 61% of variance, respectively in hourly $PM_{2.5}$ concentrations at the Fort McKay station to the input parameters used to construct the models.

3. An unfavourable attribute of the ANN sub-models constructed was that they were unable to predict actual hourly $PM_{2.5}$ concentrations for high values (>30 $\mu g/m^3$). This suggests that there are other unexplained factors contributing to hourly $PM_{2.5}$ concentrations at the Fort McKay station not taken into account in the modeling process. Influences from local and/or near-field sources – including road dust emissions, vehicle exhaust, and residential wood burning associated with home-heating in the community and/or near the station – were not taken into account in the modeling process and may offer a partial explanation.

4. Hourly $PM_{2.5}$ concentrations at the Fort McKay station when winds blew from the west were compared to corresponding concentrations when winds blew from all directions. Winds blowing from the west were assumed to represent the nearest to "background" conditions in terms of upwind $PM_{2.5}$ sources affecting the Fort McKay station compared to winds blowing from other directions. A reasonable assumption was made that winds blowing from the west pass over land with minor $PM_{2.5}$ generation from oil sand and other undefined activities compared to winds blowing from all directions:
   - It was found that $PM_{2.5}$ concentrations at the Fort McKay station were lower when winds blew from the west compared to all directions. The 50th percentile concentration was 2.5 $\mu g/m^3$ when winds blew from the west compared to 4 $\mu g/m^3$ when winds blew from all directions. The 90th percentile concentration was 9 $\mu g/m^3$ when winds blew from the west compared to 12.5 $\mu g/m^3$ when winds blew from all directions.
   - It was also found that the cumulative distribution of hourly $PM_{2.5}$ concentrations beyond the 95th percentile concentration was virtually the same for winds blowing from the west and winds blowing from all directions. This suggests that peak hourly $PM_{2.5}$ concentration events at the Fort McKay station (e.g. concentrations >15 to 20 $\mu g/m^3$) are unrelated to any specific wind direction.

5. A sensitivity analysis examined whether the ANN sub-models constructed would respond to simulated increases in one of the important input parameters. This parameter consisted of the hourly $PM_{2.5}$ concentration at AMS #13 that represented the "near-field" influence of PM-generation activities at the Syncrude Canada Ltd. site. It was found that by increasing hourly $PM_{2.5}$ concentrations by 50% at AMS #13, hourly $PM_{2.5}$ concentrations at the Fort McKay station would increase by <5 $\mu g/m^3$. Overall, the magnitude of this increase is considered small.

Artificial Neural Network is considered a powerful tool for modeling. The following are recommendations for future research that may improve upon what was found in this study:

1. This study originally considered one full year of hourly data for analysis corresponding to 8,760 hourly data sets. After examining these data sets to remove missing and anomalous data, only 1,994 valid data sets remained for ANN modeling purposes. This corresponded to 6,766 invalid data sets. Given the high proportion of invalid data sets, future research should consider initially working with a larger body of data, e.g. two to three years worth of hourly data, in attempts to improve the forecasting ability of the models.

2. In developing an ANN model to forecast $PM_{2.5}$ concentrations at the Fort McKay station, input data from only one additional air monitoring station were considered (monitoring station #13). Future research could consider input data from additional air monitoring stations in the region in attempts to improve the forecasting ability of the models.

3. A three layer back propagation network was used in ANN models evaluated in this study. Other model structures – four layer, five layer, multi-hidden slabs – could offer improved forecasting ability and should be considered in future research.

REFERENCES

AEP. 1996. Air Quality Monitoring in Alberta, 1994 Summary Report. Alberta Environment, Edmonton, AB.

AEP. 2001. Air Quality Monitoring in The Bow Corridor. Final Report, December 1999 to August 2001. Alberta Environment, Edmonton, AB.

Baba, K., Matuzaki, H. and Nogita, S. 1990. Intelligent Support System for Water and Sewage Treatment Plants Which Includes A Past History Learning Function – Coagulant Injection Guidance System Using Neural Net Algorithm. Proceedings of the 5th International Association on Water Pollution Research and Control Workshop. Yokohama and Kyoto, Japan. Pergammon Press.

Baxter, C.W., Stanley, S.J. and Zhang, Q. 1999. Development of A Full-scale Artificial Neural Network Model for The Removal of Natural Organic Matter by Enhanced Coagulation. J Water SRT – Aqua, 48: 129-136.

Beychok, M.R. 1996. Fundamentals of Stack Gas Dispersion, Third Edition. Beychok, M.R., Irvine, California.

Boznar, M., Lesjak, M. and Mlakar, P. 1991. A Neural Network-Based Method for Short-term Predictions of Ambient SO_2 Concentrations in Highly Polluted Industrial Areas of Complex Terrain. Atmospheric Environment, 27: 221-230.

Burnett, R.T., Dales, R.E., Krewski, D., Vincent, R., Dann, T.F. and Brook, J.F. 1995. Associations Between Ambient Particulate Sulfate and Admissions to Ontario Hospitals for Cardiac and Respiratory Diseases. American Journal of Epidemiology, 142: 15-22.

Canadian Natural Resources Limited. 2002. Application for Approval: Horizon Oil Sand Project. Submitted to: Alberta Energy and Utility Board and Alberta Environment, Edmonton, AB.

CASA. 2004a. Data Warehouse, What is air quality, wind. Clean Air Strategic Alliance, Edmonton, AB. Available at: http://www.casadata.org/whatis/wind.asp (accessed 27 March 2004).

CASA. 2004b. Data Warehouse, What is air quality, turbulence. Clean Air Strategic Alliance, Edmonton, AB. Available at: http://www.casadata.org/whatis/turbulence.asp (accessed 27 March 2004).

CASA. 2004c. Data Warehouse, Monitoring Methods. Clean Air Strategic Alliance, Edmonton, AB. Available at: http://www.casadata.org/airquality/ index.asp (accessed 27 March 2004).

CASA. 2004d. Data Warehouse, Long-term Trends. Clean Air Strategic Alliance, Edmonton, AB. Available at: http://www.casadata.org/longterm/ (accessed 6 April 2004).

CASA. 2004e. Data Warehouse, Monitoring Methods. Clean Air Strategic Alliance, Edmonton, AB. Available at: http:// www.Casadata.org/ pollutants/hydrogen_sulphide.asp (accessed 8 April 2004).

CASA. 2004f. Data Warehouse, Monitoring Methods. Clean Air Strategic Alliance, Edmonton, AB. Available at: http:// www.Casadata.org/pollutants/ hydrocarbons.asp (accessed 8 April 2004).

CASA. 2004g. Data Warehouse, Monitoring Methods. Clean Air Strategic Alliance, Edmonton, AB. Available at: http:// www.Casadata.org/ pollutants/ozone.asp (accessed 8 April 2004).

CASA. 2004h. Data Warehouse, Monitoring Methods. Clean Air Strategic Alliance, Edmonton, AB. Available at: http:// www.Casadata.org/ pollutants/oxides_nitrogen.asp (accessed 8 April 2004).

CCME. 2000. Canada-Wide Standards for Particulate Matter (PM) and Ozone. Canadian Council of Ministers of the Environment, Winnipeg, MB.

CEPA/FPAC. 1999. National Ambient Air Quality Objectives for Particulate Matter. A Report by CEPA/FPAC Working Group on Air Quality Objectives and Guidelines, Ottawa, ON.

Chaloulakou, A., Grivas, G. and Spyrellis, N. 2003. Neural Network and Multiple Regression Models for PM10 Prediction in Athens: A Comparative Assessment. Journal of The Air and Waste Management Association, 53: 1183-1190.

Chelani, A.B., Chalapati Rao, C.V., Phadke, K.M. and Hasan, M.Z. 2002. Prediction of Sulphur Dioxide Concentration Using Artificial Neural Networks. Environmental Modeling and Software, 17: 161-168.

Chelani, A.B., Gajghate, D.G. and Hasan, M.Z. 2002. Prediction of Ambient PM_{10} and Toxic Metals Using Artificial Neural Network Model. ASCE Journal of Environmental Engineering, 128: 1146-1157.

Cheng, L., Sandhu, H.S., Angle, R.P. and Myrick, R.H. 1998. Characteristics of Inhalable Particulate Matter in Alberta Cities. Atmospheric Environment, 32: 3835-3844.

Chow, J.C. 1995. Measurement Methods to Determine Compliance with Ambient Air Quality Standards for Suspended Particles. Journal of The Air and Waste Management Association, 45: 320-382.

Comrie, A.C. 1997. Comparing Neural Networks and Regression Models for Ozone Forecasting. Journal of The Air and Waste Management Association, 47: 653-663.

Deslauriers, M. 1996. Canadian Emissions Inventory of Criteria Air Contaminants (1990) (Report EPS 5/AP/7E). Pollution Data Branch, Environmental Protection Service, Environment Canada, Ottawa, ON.

Dobbins, R.A. 1976. Atmospheric Motion and Air Pollution. Published by John Wiley & Sons, Inc., Toronto, ON.

Erishman, J. W. and Schaap, M. 2004. The Need for Ammonia Abatement with Respect to Secondary PM Reductions in Europe. Environmental Pollution, 129: 159-163.

Gamal El-Din, A. and Smith, D.W. 2002. Modeling a Full-scale Primary Sedimentation Tank Using Artificial Neural Networks. Environmental Technology, 23: 479-496.

Gardner, M.W. and Dorling, S.R. 1999. Neural Network Modeling and Prediction of Hourly NO_x and NO_2 Concentration in Urban Air in London. Atmospheric Environment, 31: 709-719.

Harper, A. 2004. Water and Wastewater Management Practices at SAGD Facilities. M.Eng. Report Prepared for Department of Civil and Environmental Engineering at University of Alberta, Edmonton, AB.

Harvey, S. and Harvey, R. 1998. An Introduction to Artificial Intelligence. Appita, J, 51 (1): 20-24.

Hasham, F. A., Kindzierski, W. B. and Stanley, S. J. 2004. Modeling of Hourly NO_X concentrations using artificial neural networks, Journal of Environmental Engineering Science, 3(Suppl. 1): S111-S119.

Hornberg, C., Seemayer, N.H. and Kainka, E. 1998. Strong Genotoxicity on Human Tracheobronchial Epithelial Cells (BEAS-2B) In Vitro by The Coarse (PM-10) and Fine Fraction (PM-2.5) of Airborne Particulates as An

Indicator of Potential Adverse Health Effects. Journal of Aerosol Science, 29 Supp. 1: 317-S318.

Idriss, A. 2003. Air Quality Model Guidelines. Alberta Environment, Edmonton, AB.

Jain, A.K., Mao, J. and Mohiuddin, K.M. 1996. Artificial Neural Networks: A Tutorial. IEEE Computer, 29: 31-44.

Jones, C. 1996. New $PM_{2.5}$ Standard Ensnares More than Just Particulates. Power, November/December: 57-60.

Joo, D., Choi, D. and Park, H.Y. 2000. The Effects of Data Preprocessing in The Determination of Coagulant Dosing Rate. Water Research, 34: 3295-3302.

Katrinak, K.A., Rez, P., Perkes, P.R. and Buseck, P.R. 1993. Fractal Geometry of Carbonaceous Aggregates From an Urban Aerosol. Environmental Science & Technology, 27: 539-547.

Kohlemainen, M., Martikainen, H., Hiltunen, T. and Ruuskanen, J. 2000. Forecasting Air Quality Parameters Using Hybrid Neural Network Modeling. Environmental Monitoring and Assessment, 65: 277-286.

Lu, W.Z., Fan, H.Y., Leung, A.Y.T. and Wong, J.C.K. 2002. Analysis of Pollutant Levels in Central Hong Kong Applying Neural Network Method with Particle Swarm Optimization. Environmental Monitoring and Assessment, 79: 217-230.

Lu, W.Z., Wang, W.J., Fan, H.Y., Leung, A.Y.T., Xu, Z.B., Lo, S.M. and Wong, J.C.K. 2002. Prediction of Pollutant Levels in Causeway Bay Area of Hong Kong Using an Improved Neural Network Model. ASCE Journal of Environmental Engineering, 128: 1146-1157.

Maier, H.R., Morgan, N. and Chow, C.W.K. 2003. Use of Artificial Neural Networks for Predicting Optimal Alum Dose and Treated Water Quality Parameters. Environmental Modeling and Software, 19: 485-494.

McKendry, I. 2000. Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM_{10} and $PM_{2.5}$) Forecasting. Journal of The Air and Waste Management Association, 52: 1096-1101.

Pacific Environmental Services, Inc. 1999. Getting Started: Emissions Inventory Methods For PM-2.5, Prepared for: PM-2.5 Committee Emission Inventory Improvement Program. Washington, DC..

Ratnaweera, H. and Blorm, H. 1995. Optimization of Coagulant Dosing Control Using Real-time Models Selective to Instrumental Errors. Water Supply, 13: 285-289.

RERI. 1994. The Air Quality Valuation Model. Draft report prepared by Regional Economic Research, Inc., for California Energy Commission, Sacramento, CA.

Sandhu, H.S. 1998. Ambient Particulate Matter in Alberta, Report Prepared for Science and Technology Branch, Alberta Environment, Edmonton, AB.

Santachiara, G., Prodi, F., and Vivarelli, F. 1989. Absorption of Sulphur Dioxide on Monodisperse Water Droplets and Catalytic Activity of Carbon Particles. Atmospheric Environment, 23: 1775-1782.

Schwartz, J., Dockery, D.W., Neas, L.M., Wypij, D.,Ware, J.H., Spengler, J.D., Koutrakis, P., Speizer, F.E. and Ferris Jr., B.G. 1994. Acute Effects of Summer Air Pollution on Respiratory Symptom Reporting in Children. American Journal of Respiratory Critical Care Medicine, 150: 1234-1242.

Seinfeld, J.H. 1986. Atmospheric Chemistry and Physics of Air Pollution. John Wiley & Sons, New York, NY.

Sloss, L.L. and Smith, I.M. 2000. PM10 and PM2.5: An International Perspective. Fuel Processing Technology, 65-66: 127-141.

Smith, W.H. 1990. Air Pollution and Forests, Interactions Between Air Contaminants and Forest Ecosystems. 2nd Edition. Springer-Verlag, New York, NY.

Spellman, G. 1999. An Application of Artificial Neural Networks to The Prediction of Surface Ozone Concentrations in the United Kingdom. Applied Geography, 19: 123-136.

Tucker, W.G. 2000. An Overview of $PM_{2.5}$ Sources and Control Strategies. Fuel Processing Technology, 65-66: 379-392.

US Environmental Protection Agency. 1996a. EPA's Proposal on The Particulate Matter Standard, Fact Sheet. US Environmental Protection Agency, Washington, DC..

US Environmental Protection Agency. 1996b. Air Quality Criteria for Particulate Matter, Report NO. EPA/600/p-95/001abcF. National Center for Environmental Assessment, US Environmental Protection Agency, Research Triangle Park, NC.

US Environmental Protection Agency. 1997. PM-2.5 Composition and Sources, Prepared for FCFA National and Regional Strategies Workgroup, Prepared by Emissions, Monitoring and Analysis Division Office of Air Quality Planning and Standards. US Environmental Protection Agency, Washington, DC..

US Environmental Protection Agency. 2000. PM – How Particulate Matter Affect the Way We Live & Breathe, Published by The US EPA Office of Air Quality Planning and Standards, November 2000. US Environmental Protection Agency, Washington, DC.. Available at: http://www.epa.gov/air/urbanair/pm/hlth1.html / (accessed 28 March 2004).

US Environmental protection Agency. 2003a. Terminology Reference System, Office of Pollution Prevention and Toxics. Last Updated on September 16th, 2003a. US Environmental Protection Agency, Washington, DC.. Available at: http://oaspub.epa.gov/trs/trs_proc_qry.navigate_term?p_term_id =1072901&p_term_cd=TERMDIS (accessed 28 March 2004).

US Environmental Protection Agency. 2003b. National Ambient Air Quality Standards (NAAQS), Last Updated on December 10th, 2003b. US Environmental Protection Agency, Washington, DC.. Available at: http://www.epa.gov/air/criteria.html (accessed 29 March 2004).

Valentin, N., Denoeux, T., Fotoohi, F. 1999. Modeling of Coagulant Dosage In a Water Treatment Plant, Engineering Applications of Neural Networks Proceedings of the 5th International Conference on Engineering Applications of Neural Networks, Wydawnictwo Adam Marszalek, Torun, Poland.

Vedal, S. 1995. Health Effects of Inhalable particles: Implications for British Columbia, Prepared for The Air Resources Branch, British Columbia Ministry of Environment, Lands and Parks, Victoria, BC.

Vedal, S. 1996. Evaluation of health impacts due to fine inhalable particles (PM2.5). Contract report prepared from Health Canada, final report.

Ward Systems Group, Inc.. 2003. NeuroShell 2, Release 4.0.

Wark, K., Warner, C. and Davis, W.T. 1998. Air Pollution Control: Its Origin and Control, 3rd Edition. Addison Wesley, Boston, MA.

WBEA. 2004. Maps of air monitoring stations. Available at: http://www.wbea.org/html/air/ map.html (accessed 6 April 2004).

Wilson W.E. and Suh, H.H. 1999. Fine and Coarse Particles: Concentration Relationships Relevant to Epidemiological Studies. Journal of The Air and Waste Management Association, 47: 1238-1249.

Wilson, W.E., Chow, J.C., Claiborn, C.S., Fusheng, W., Engelbrecht, J., and Watson, J.G., 2002. Monitoring of Particulate Matter Outdoors. Chemosphere, 49: 1009-1043.

Yi, J., and Prybutok, V.R. 1996. A Neural Network Model Forecasting for Prediction of Daily Maximum Ozone Concentration in an Industrialized Urban Area. Environmental Pollution, 92: 349-357.

Zhang, Q. and Stanley, S.J. 1999. Real Time Water Treatment Process Control with Artificial Neural Networks. Journal of Environmental Engineering, 125:153-160.

APPENDIX A – DISTANCE BETWEEN AMS #1 AND AMS #13

A degree (°) of latitude or longitude can be subdivided into 60 parts called minutes ('). Each minute can be further subdivided into 60 seconds ("). One degree of latitude equals approximately 69 miles (111 km). One minute of latitude is just over a mile, and one second is around 100 feet. The length of a degree of longitude varies from 69 miles at the equator to 0 at the poles" (GeoSystems Global Corporation, 1998).

Table 1. Geographical position of AMS #1 and AMS #13.

Air Monitoring Station #	Latitude	Longitude
1	57° 11' 20.9"	111° 38' 25.9"
13	57° 08.951'	111° 38.547'

Vertical distance between AMS #1 and AMS #13:

1'=111/60 = 1.85 km and 1"= 1.85/60 = 0.031 km

Vertical distance = [(11 x 1.85) + (20.9 x 0.031)] – [8.951 x 1.85] = 4.44 km

Horizontal distance between AMS #1 and AMS #13:

At 57° North

1° = ((90-57)/90) * 111 = 40.7 km

1' = 40.7/60 = 0.68 km

1" = 0.68/60 = 0.011 km

Horizontal distance = 0.547 x 0.68 – 25.9 x 0.011 = 0.09 km

The distance between AMS #1 and AMS #13 = $\sqrt{(4.44)^2 + (0.09)^2}$ = 4.44 km.

Therefore the distance between AMS #1 and AMS #13 almost equals the vertical distance. The horizontal distance between AMS #1 and AMS #13 AMS #1 and AMS #13 longitude positions is 90 m. Therefore, AMS #13 is located at almost exact south side of AMS #1.

APPENDIX B – SUMMARY STATISTICS FOR PARAMETERS USED IN ANN MODELING

Table A1. Parameters used from AMS #1.

Parameter	SO2	TRS	THC	O3	NO	NO_2	NO_X	$PM_{2.5}$	ETL	RH	GR	WS	WS_SD	WD	WD_SD
Units	ppb	ppb	ppm	ppb	ppb	ppb	ppb	µg/m³	°C	%					
Mean	1.20	0.73	1.83	21.98	2.16	5.80	7.94	4.13	0.61	67.89	108.63	7.33	3.21	201.52	32.05
Median	0.23	0.54	1.78	21.79	0.15	2.60	2.90	2.74	0.60	72.24	3.36	6.34	2.77	191.20	29.10
Std. Dev.	3.63	0.68	0.38	13.79	6.50	7.17	12.04	7.58	15.49	19.17	378.91	5.61	2.15	196.05	17.39
Var.	13.14	0.46	0.14	190.17	42.29	51.42	144.85	57.42	240.02	367.39	143576.00	31.51	4.63	38434.78	302.36
Minimum	-4.31	-1.79	-0.31	-0.97	-4.65	-3.33	-6.87	-50.00	-50.00	0.00	-4.87	0.00	0.00	0.00	0.00
P(0.01)	-0.18	-0.42	0.09	0.36	-0.38	-0.53	-0.69	-5.51	-36.09	17.00	-3.53	0.21	0.00	1.00	8.10
P(0.05)	-0.06	0.11	1.49	1.14	-0.14	0.05	0.03	-1.67	-25.70	29.99	-2.52	0.48	0.00	7.60	12.30
P(0.25)	0.08	0.40	1.67	10.90	0.00	0.62	0.71	0.57	-10.72	57.19	-0.17	3.08	1.81	145.80	18.90
P(0.75)	0.77	0.80	1.95	31.64	0.91	8.79	10.42	6.15	13.08	82.16	134.29	10.33	4.36	292.20	40.30
P(0.90)	2.54	1.61	2.21	39.81	5.56	16.90	22.65	11.39	19.95	89.02	395.78	15.02	6.23	340.16	54.40
P(0.95)	5.59	2.12	2.42	44.51	12.07	21.61	32.13	15.85	22.95	91.41	536.84	17.92	7.35	351.80	67.40
P(0.99)	17.01	3.25	2.86	55.40	34.87	29.06	57.94	28.77	28.39	94.79	707.53	25.21	9.49	358.45	87.25
Maximum	84.74	7.98	11.07	186.64	81.08	50.96	141.52	80.92	232.88	356.69	31420.57	36.50	19.64	14376.33	293.93

Table A2. Parameters used from AMS #13.

Parameter	SO2	TRS	THC	O3	NO	NO2	NOX	PM25	ET	WS	WS_SD	WD	WD_SD
Units	ppb	ppb	ppm	ppb	ppb	ppb	ppb	µg/m^3	°C				
Mean	1.20	0.23	2.18	21.34	1.93	5.77	7.66	3.05	0.04	4.23	1.86	176.80	36.50
Median	0.11	0.13	2.07	21.51	0.15	2.37	2.57	1.92	0.17	3.46	1.52	192.40	31.90
Std. Dev.	4.48	0.57	0.48	14.66	7.66	8.17	12.81	8.58	15.42	3.13	1.36	105.22	16.83
Var.	20.11	0.32	0.23	215.03	58.66	66.77	164.22	73.61	237.80	9.82	1.84	11071.32	283.11
Minimum	-1.11	-4.20	-0.74	-25.30	-13.98	-6.11	-12.96	-50.00	-50.00	-0.02	0.00	0.00	0.00
P(0.01)	-0.41	-0.99	1.12	-0.43	-0.85	-0.68	-1.11	-5.41	-37.24	0.27	0.10	1.90	11.20
P(0.05)	-0.24	-0.36	1.87	0.27	-0.36	-0.12	-0.29	-1.62	-27.06	0.61	0.17	8.10	16.40
P(0.25)	-0.04	-0.02	1.97	9.01	-0.08	0.50	0.48	0.37	-11.31	1.74	0.80	89.28	26.00
P(0.75)	0.54	0.36	2.25	31.62	0.90	7.96	9.58	4.32	12.23	6.17	2.62	253.10	42.40
P(0.90)	2.33	0.79	2.64	40.87	4.95	16.07	23.08	8.45	19.97	8.71	3.78	313.61	61.50
P(0.95)	5.88	1.16	2.98	45.98	10.97	23.04	33.77	12.02	23.05	10.35	4.54	343.01	73.80
P(0.99)	22.77	2.00	3.69	55.08	31.12	38.93	53.45	22.05	28.79	13.31	5.89	358.14	89.54
Maximum	91.82	13.24	15.41	81.29	436.02	57.54	437.29	444.97	34.65	18.22	8.69	359.90	102.70

Table A3. Parameters used from Fort McMurray meteorological monitoring station.

Parameter	TCO	TCA
Mean	-74	-73
Median	7	8
Std. Dev.	2829	2829
Var.	8003525	8003679
Minimum	-99999	-99999
P(0.01)	0	0
P(0.05)	0	0
P(0.25)	2	4
P(0.75)	10	10
P(0.90)	10	10
P(0.95)	10	10
P(0.99)	10	10
Maximum	10	10

APPENDIX C – TIME SERIES PLOT FOR PARAMETERS (RAW DATA) USED IN ANN MODELING

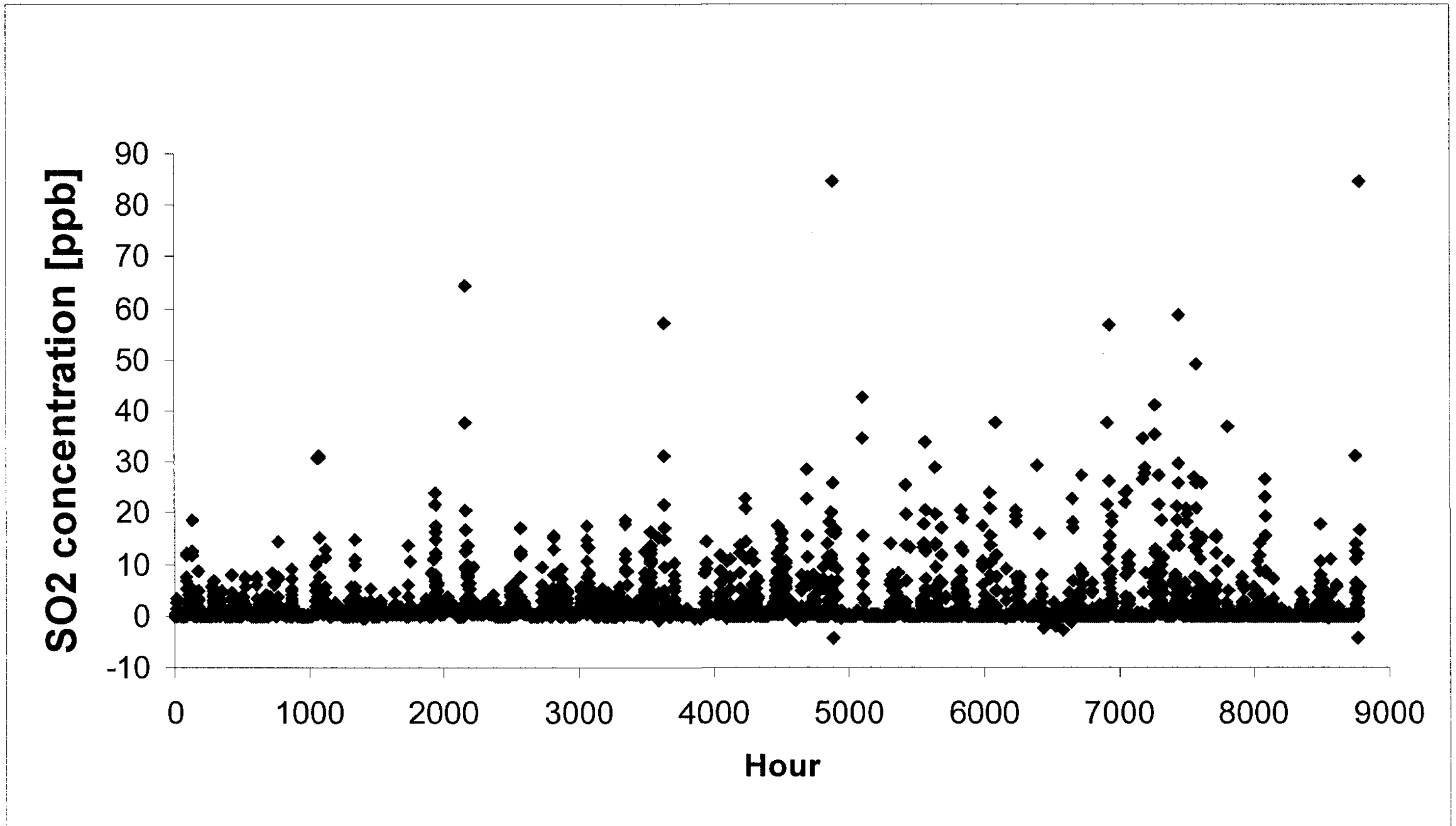


Figure A1. SO_2 [ppb] versus hour plot for AMS #1.

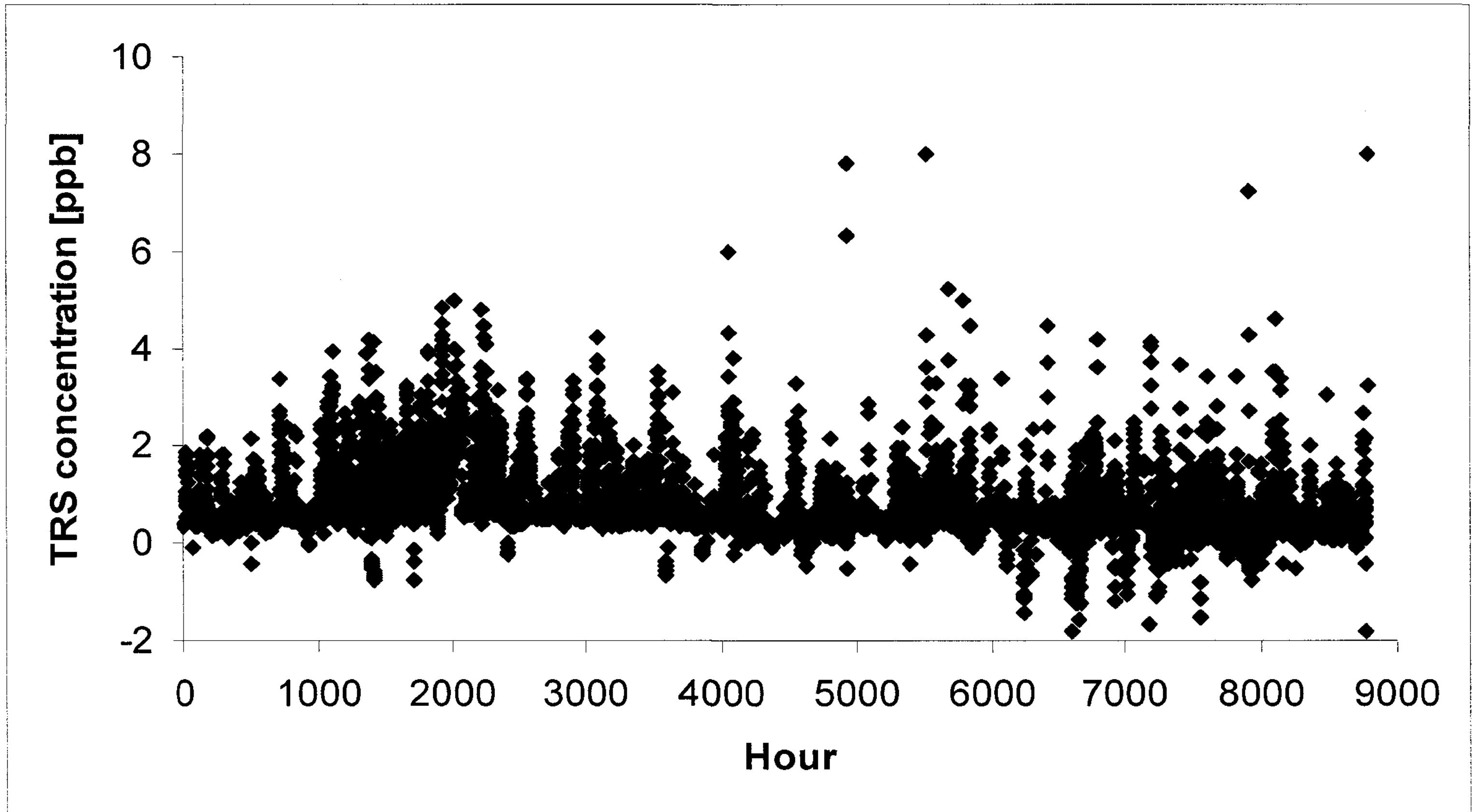


Figure A2. TRS [ppb] versus hour plot for AMS #1.

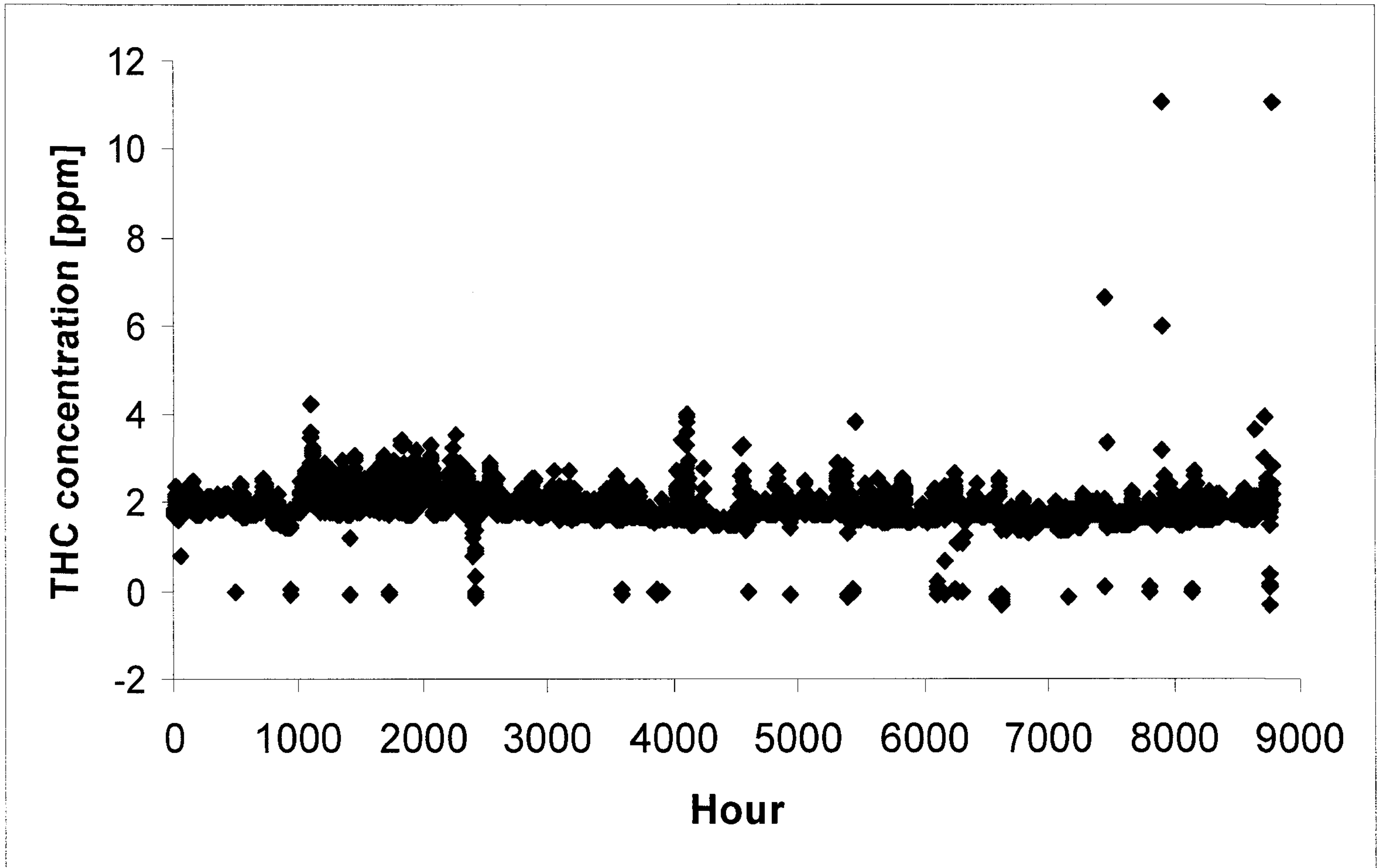


Figure A3. THC [ppm] versus hour plot for AMS #1.

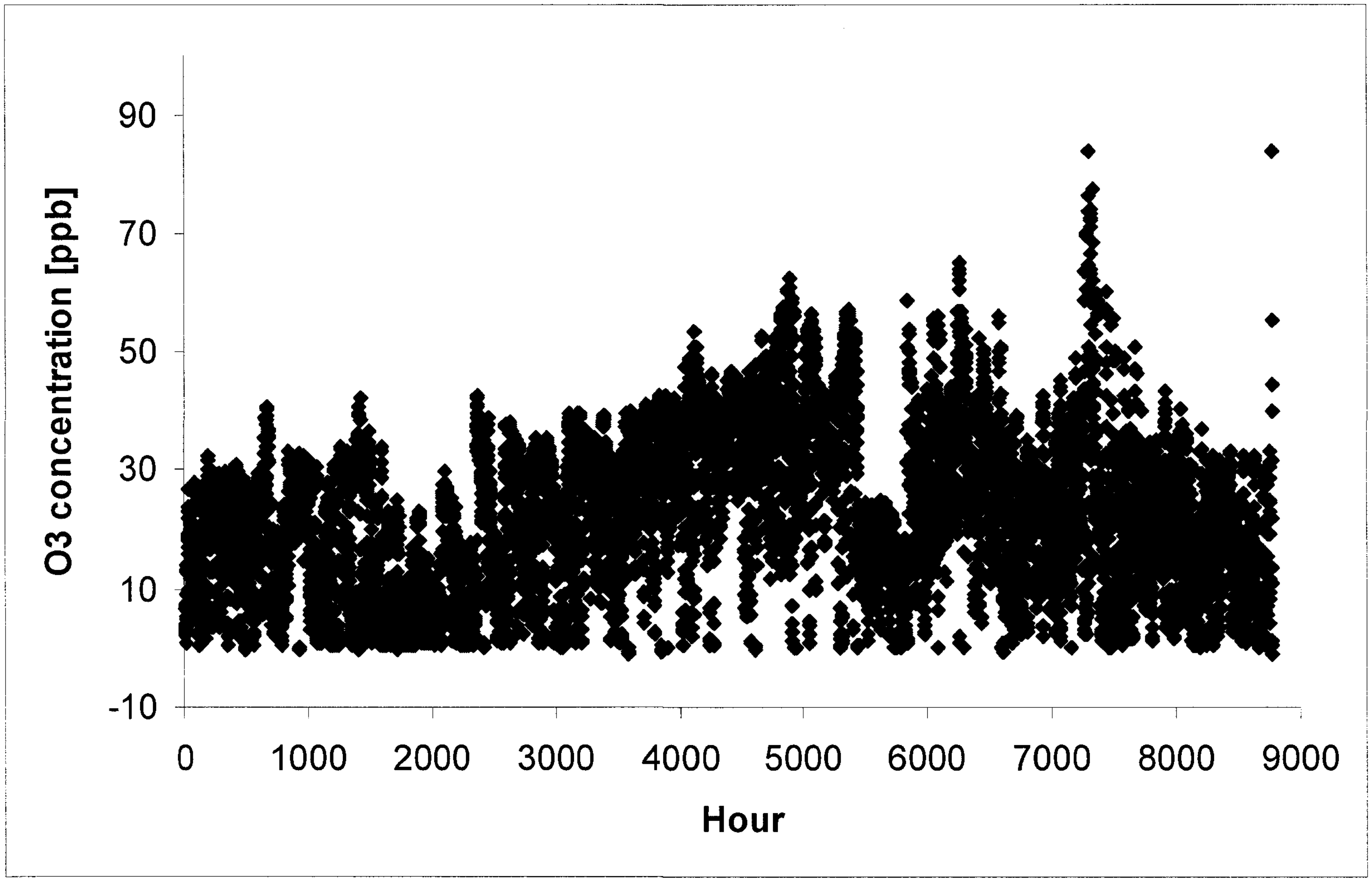


Figure A4. O_3 [ppb] versus hour plot for AMS #1.

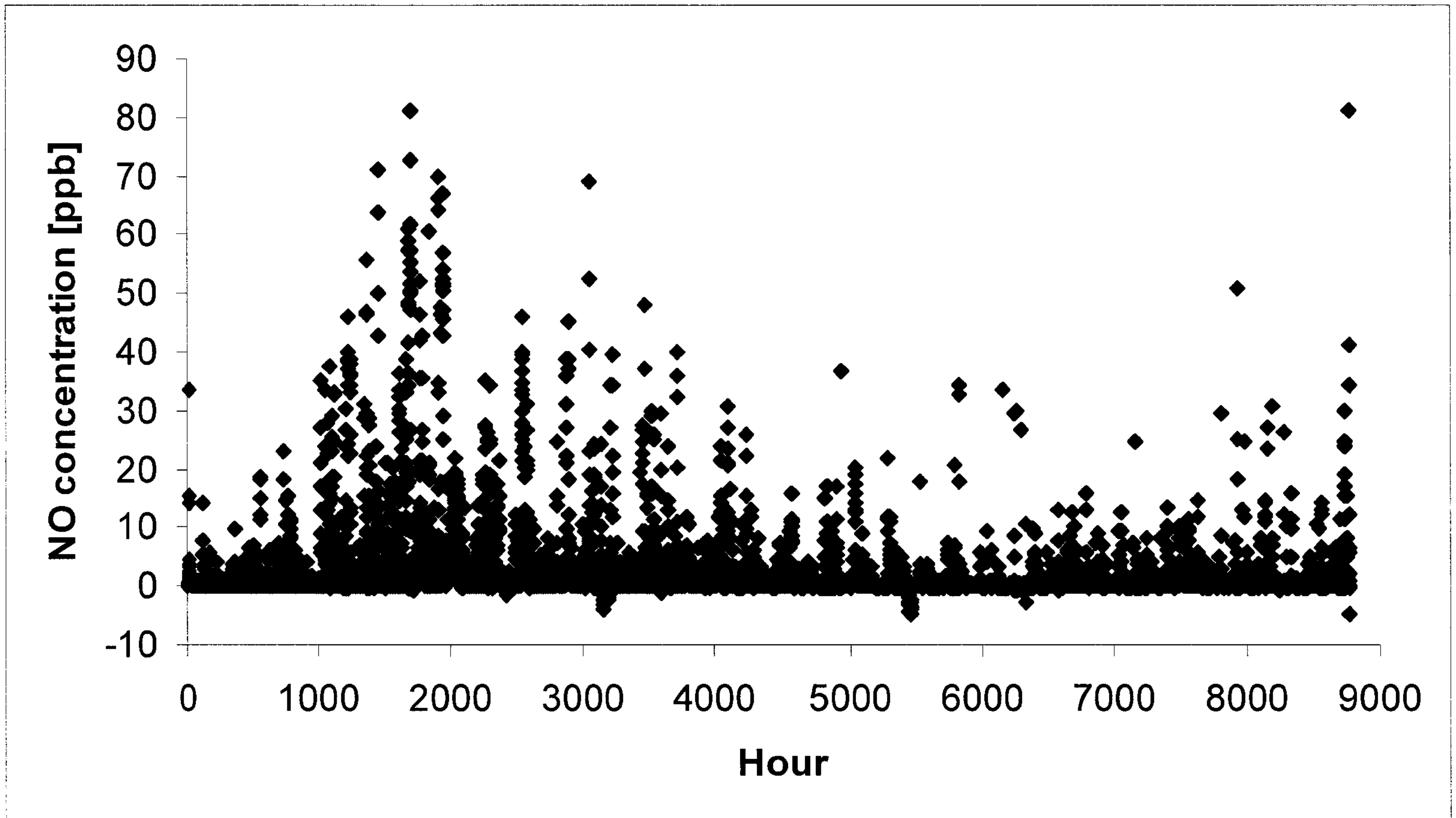


Figure A5. NO [ppb] versus hour plot for AMS #1.

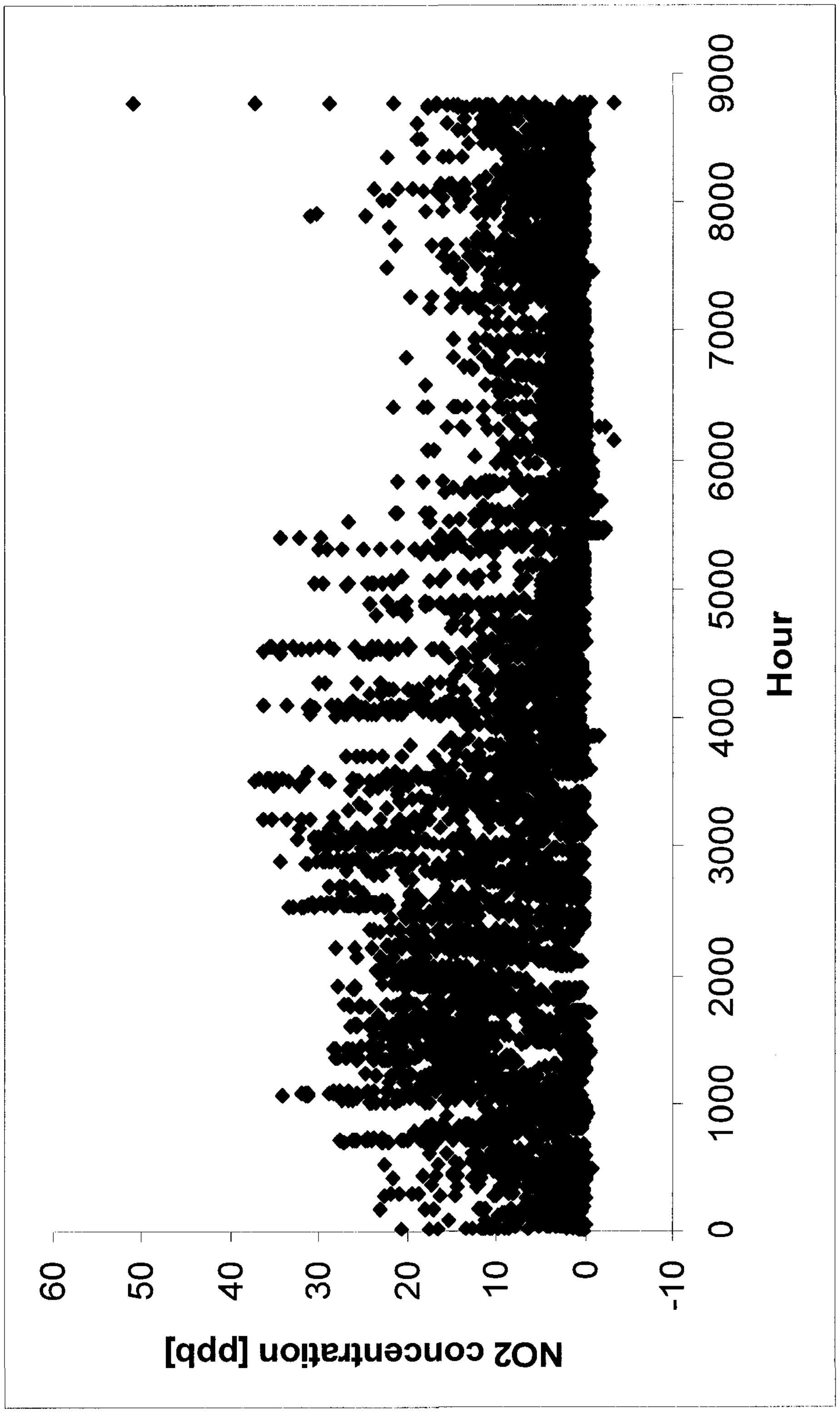


Figure A6. NO_2 [ppb] hour plot for AMS #1.

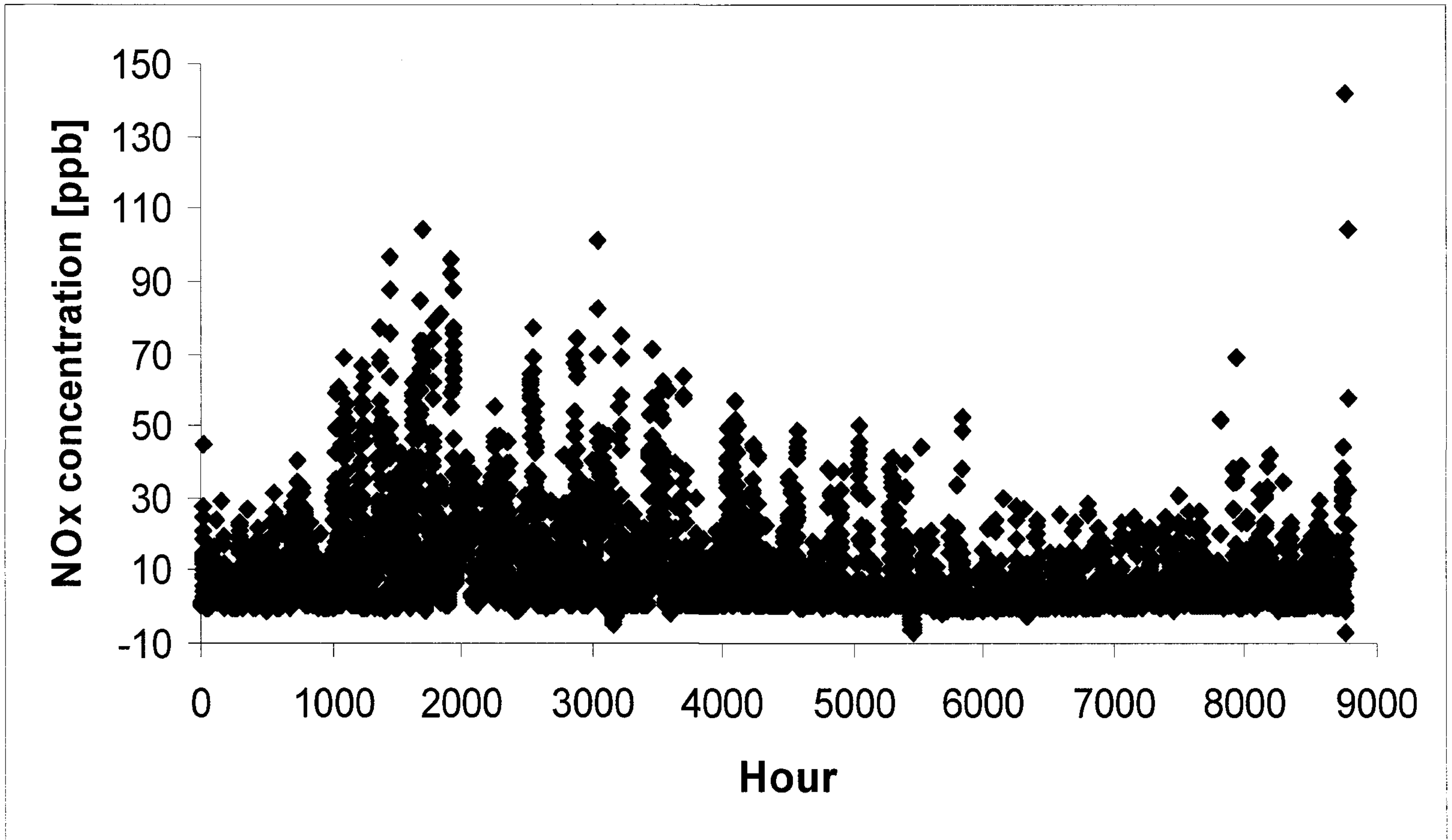


Figure A7. NO_x [ppb] versus hour plot for AMS #1.

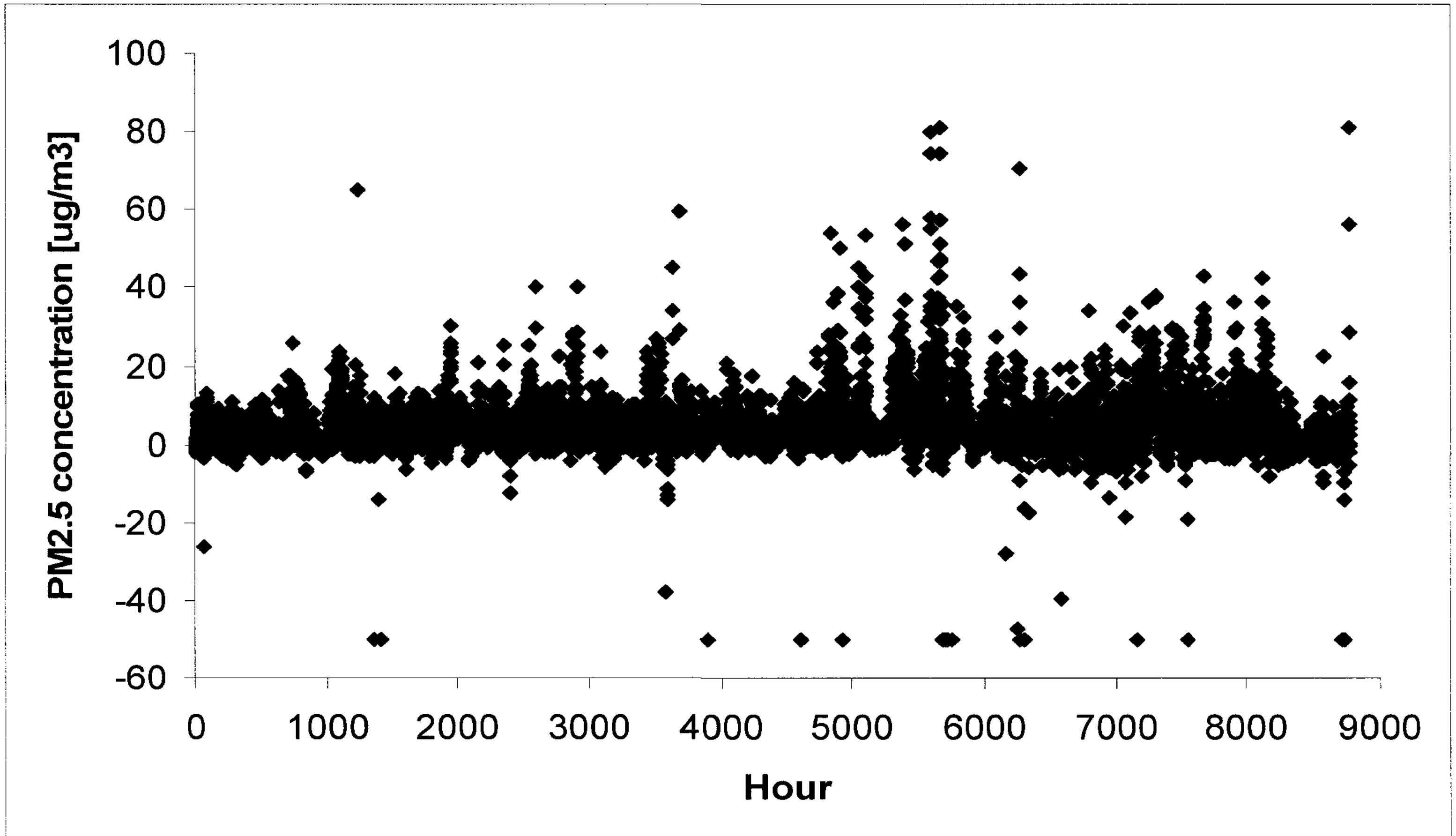


Figure A8. $PM_{2.5}$ [µg/m3] versus hour plot for AMS #1.

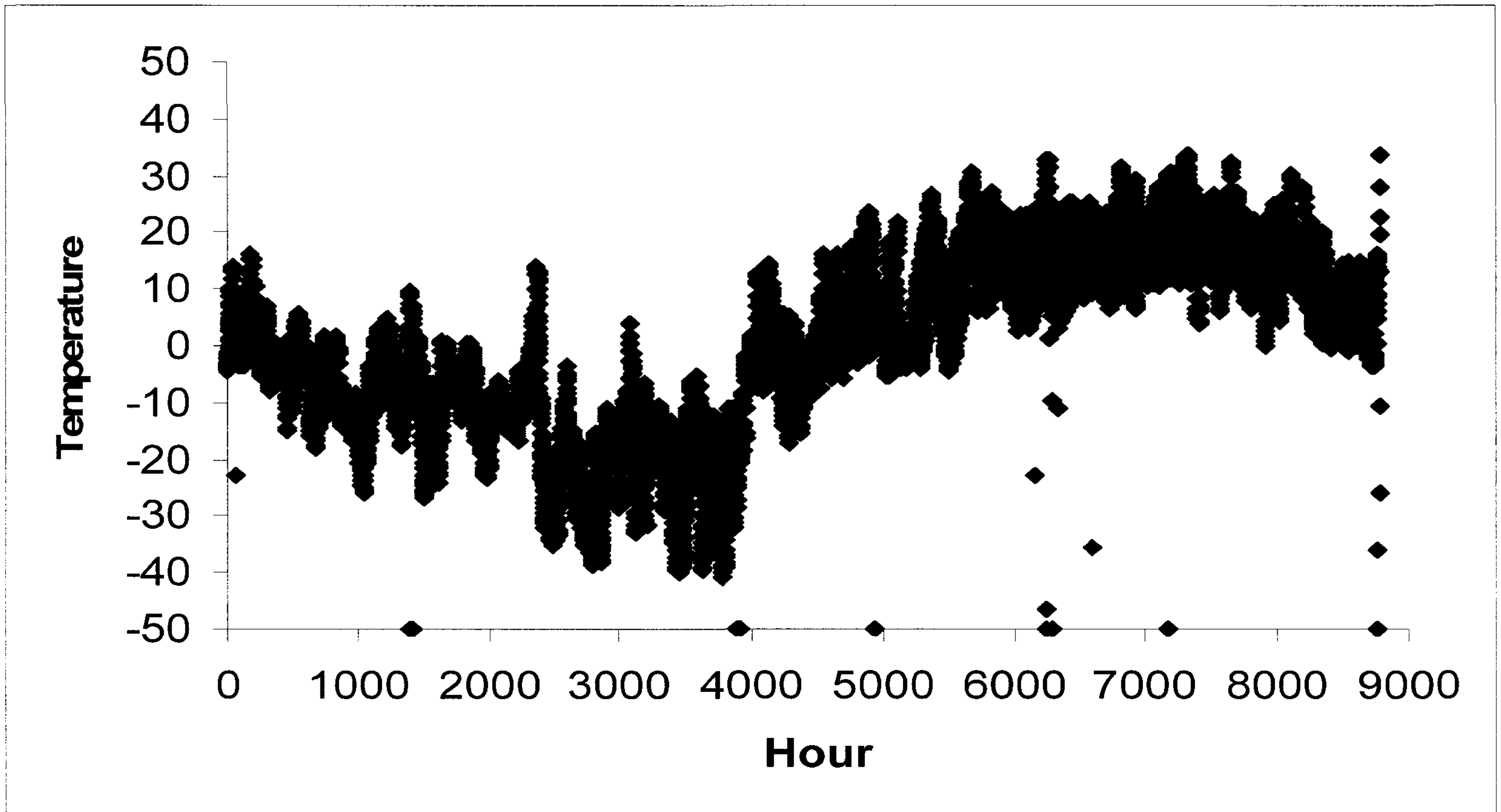


Figure A9. Temperature [°C] versus hour plot for AMS #1.

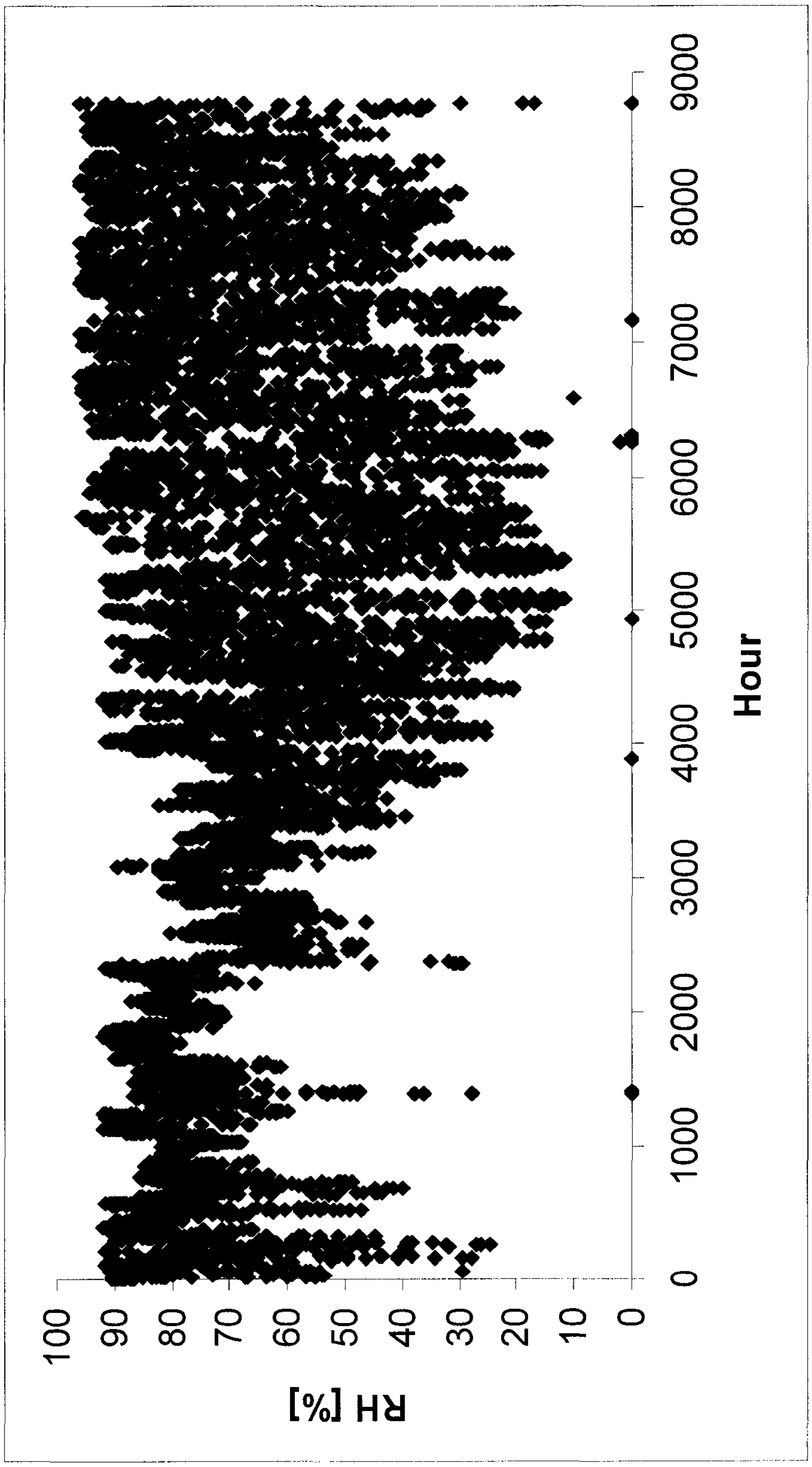


Figure A10. RH [%] versus hour plot for AMS #1.

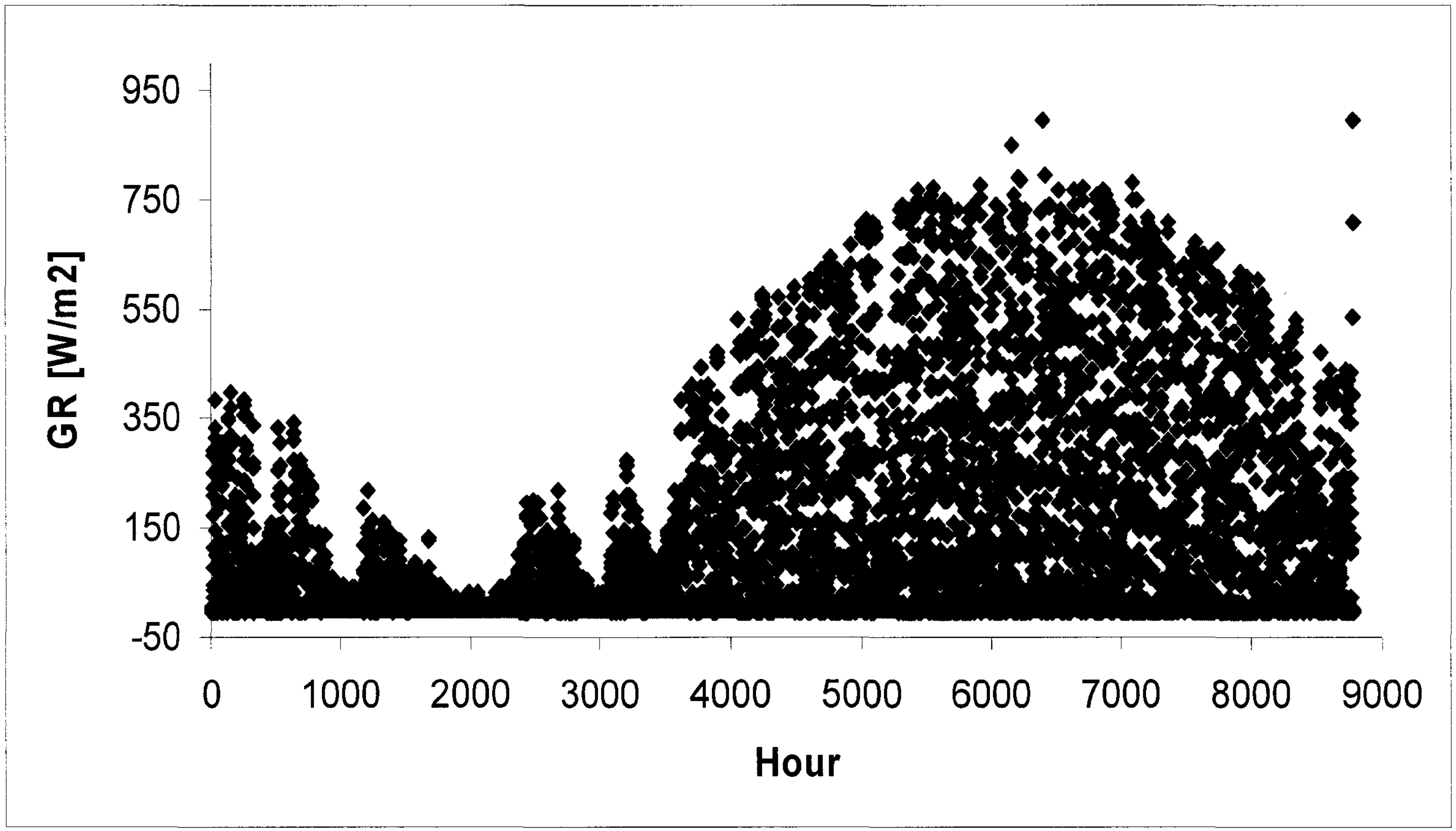


Figure A11. GR [W/m^2] versus hour plot for AMS #1.

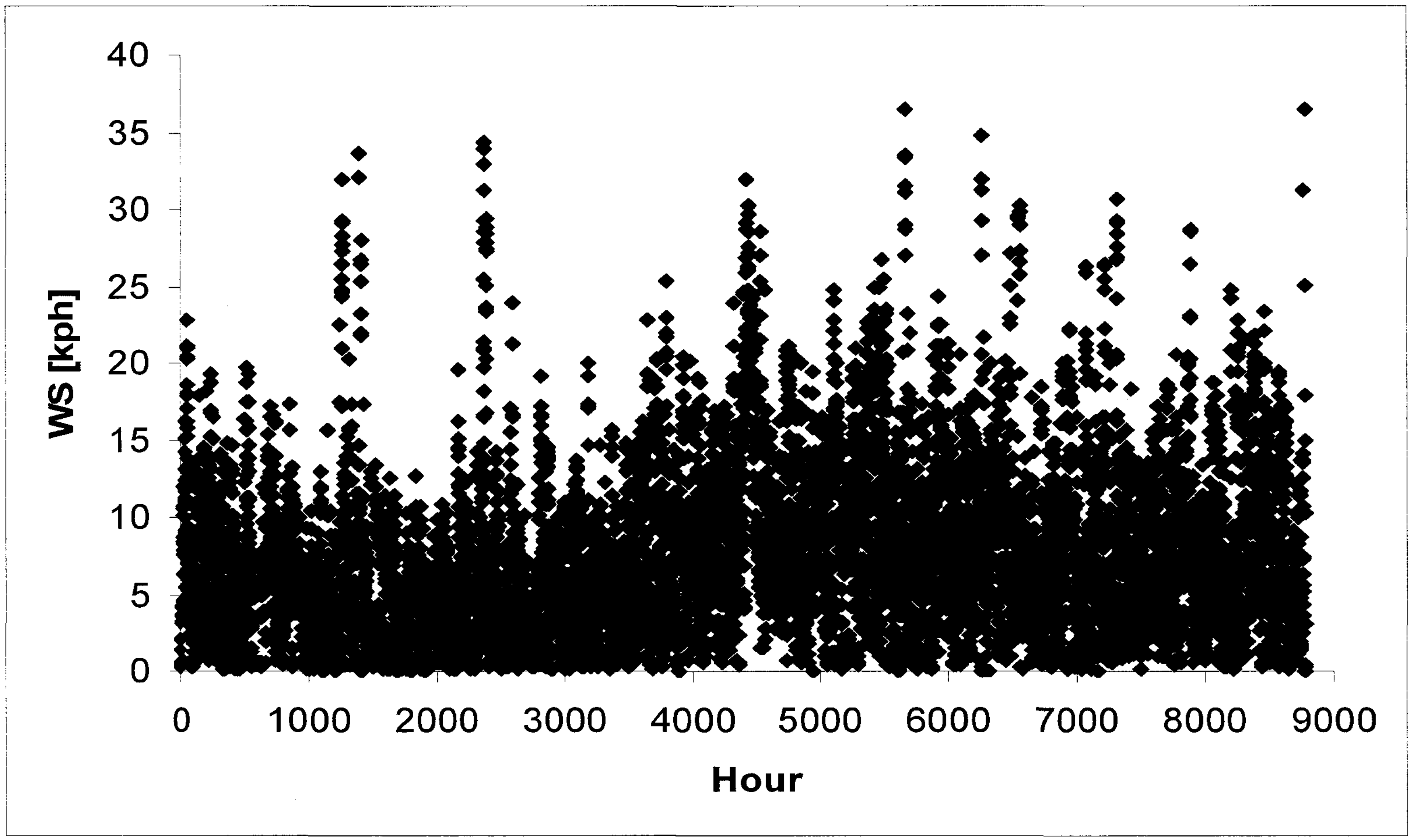


Figure A12. WS [kph] versus hour plot for AMS #1.

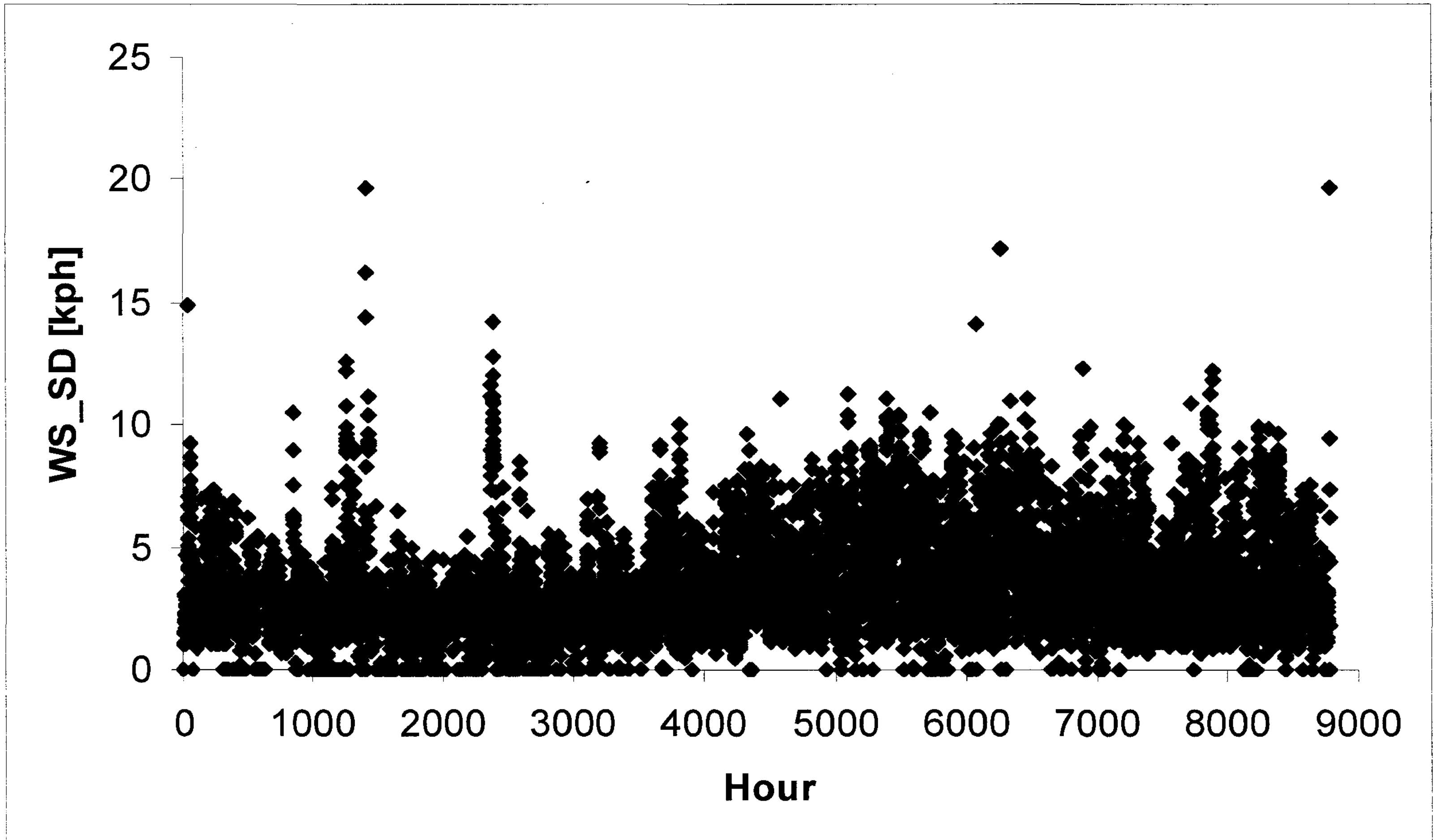


Figure A13. WS_SD [kph] versus hour plot for AMS #1.

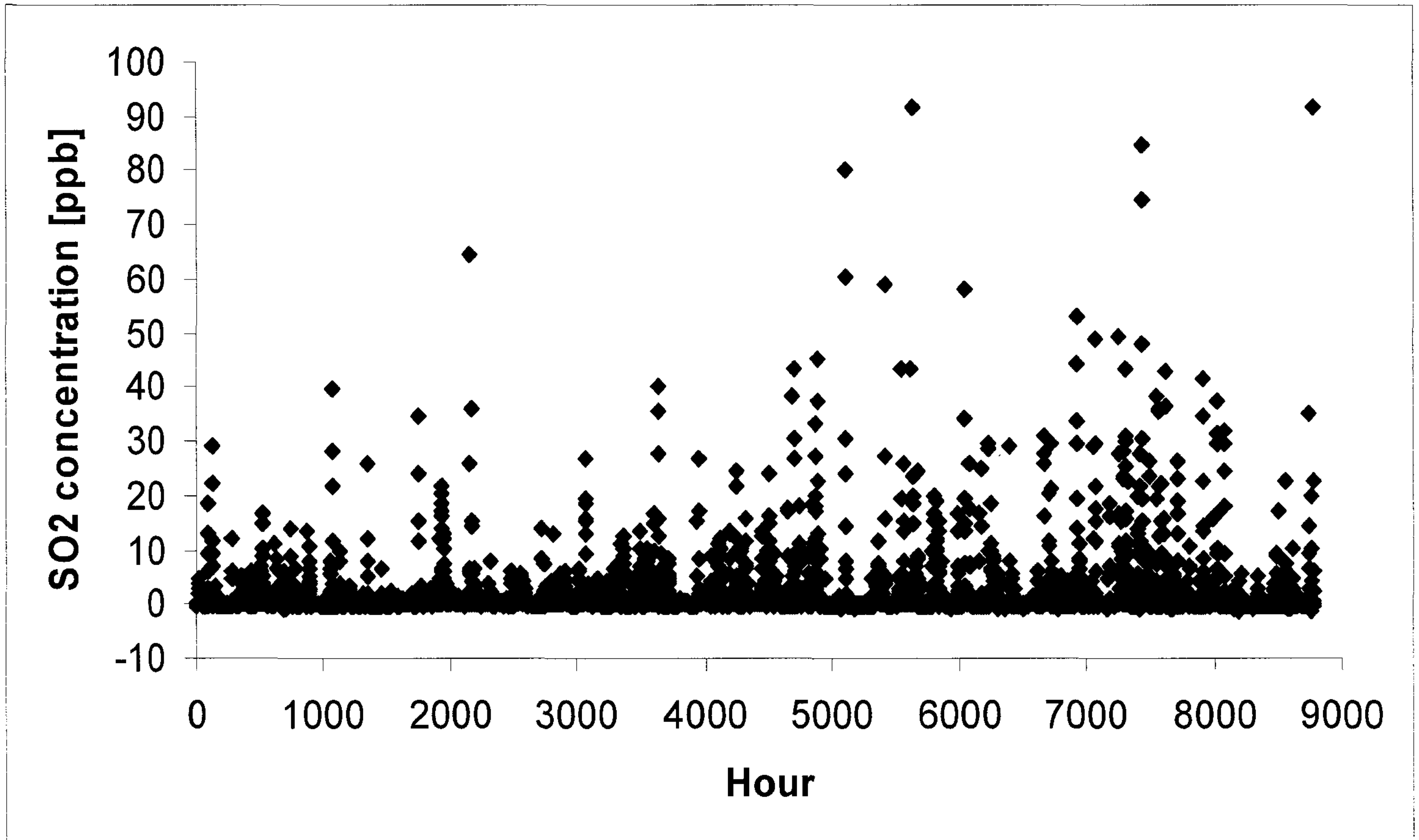


Figure A14. SO_2 [ppb] versus hour plot for AMS #13

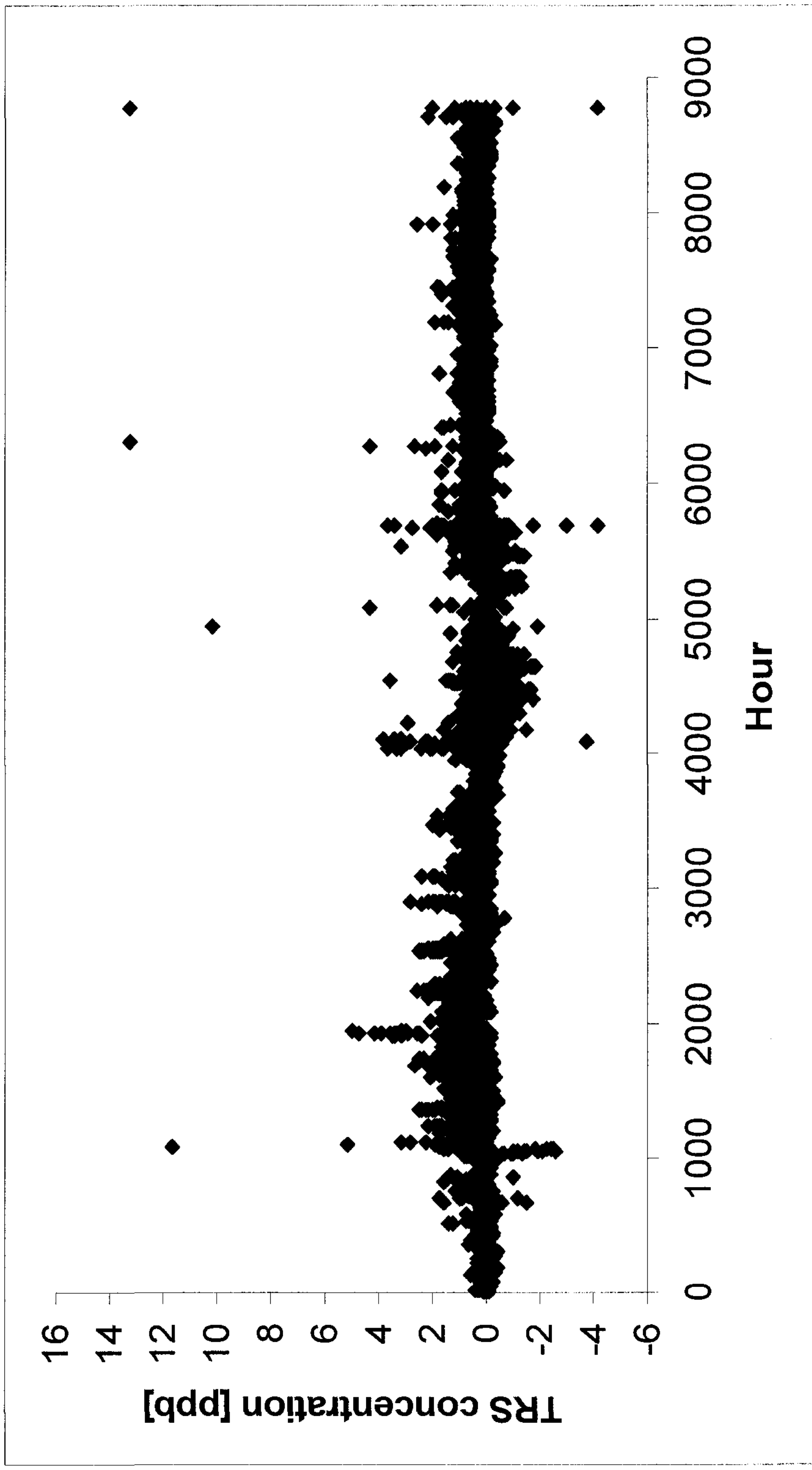


Figure A15. TRS [ppb] versus hour plot for AMS #13.

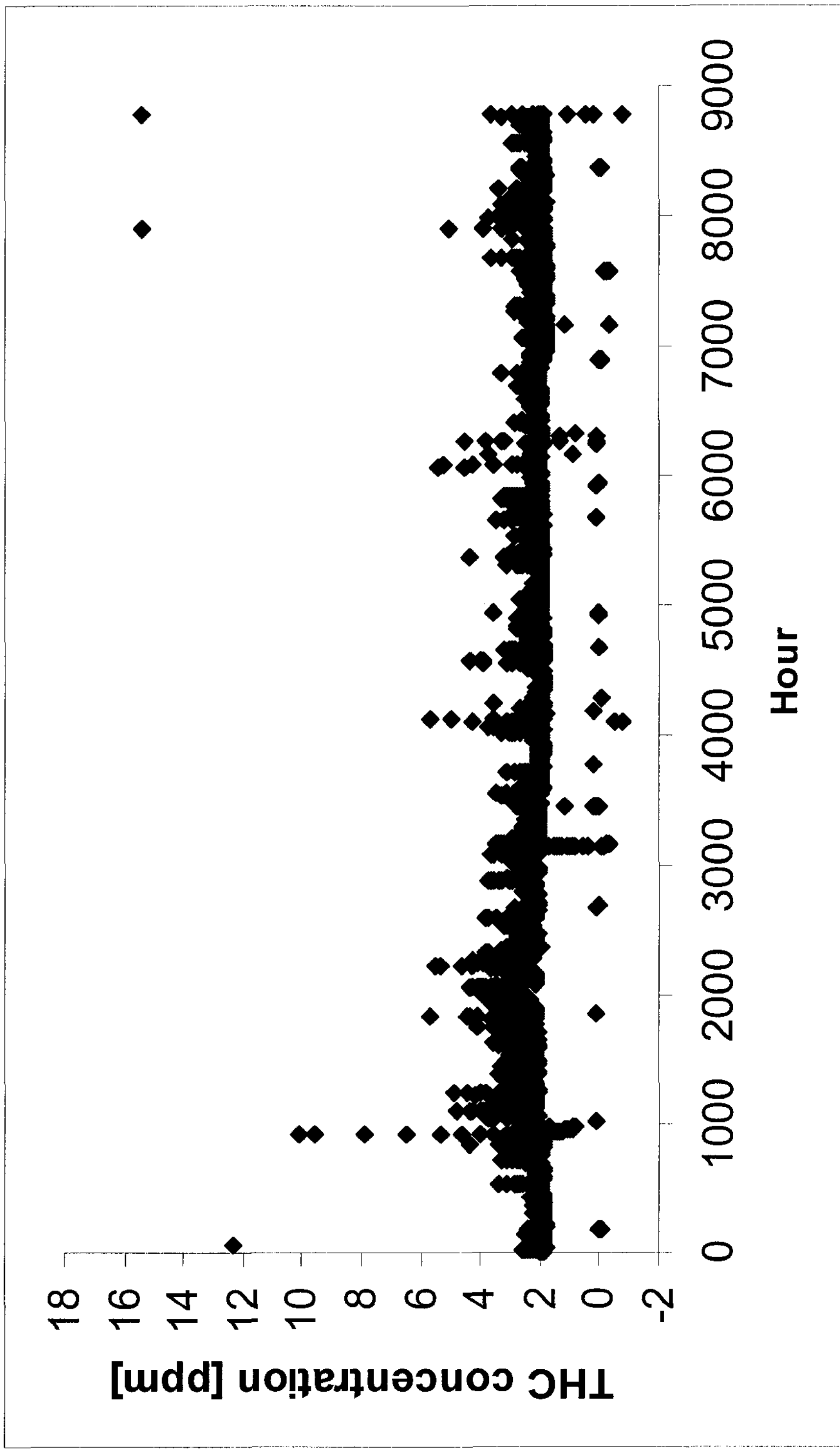


Figure A16. THC [ppm] versus hour plot for AMS #13.

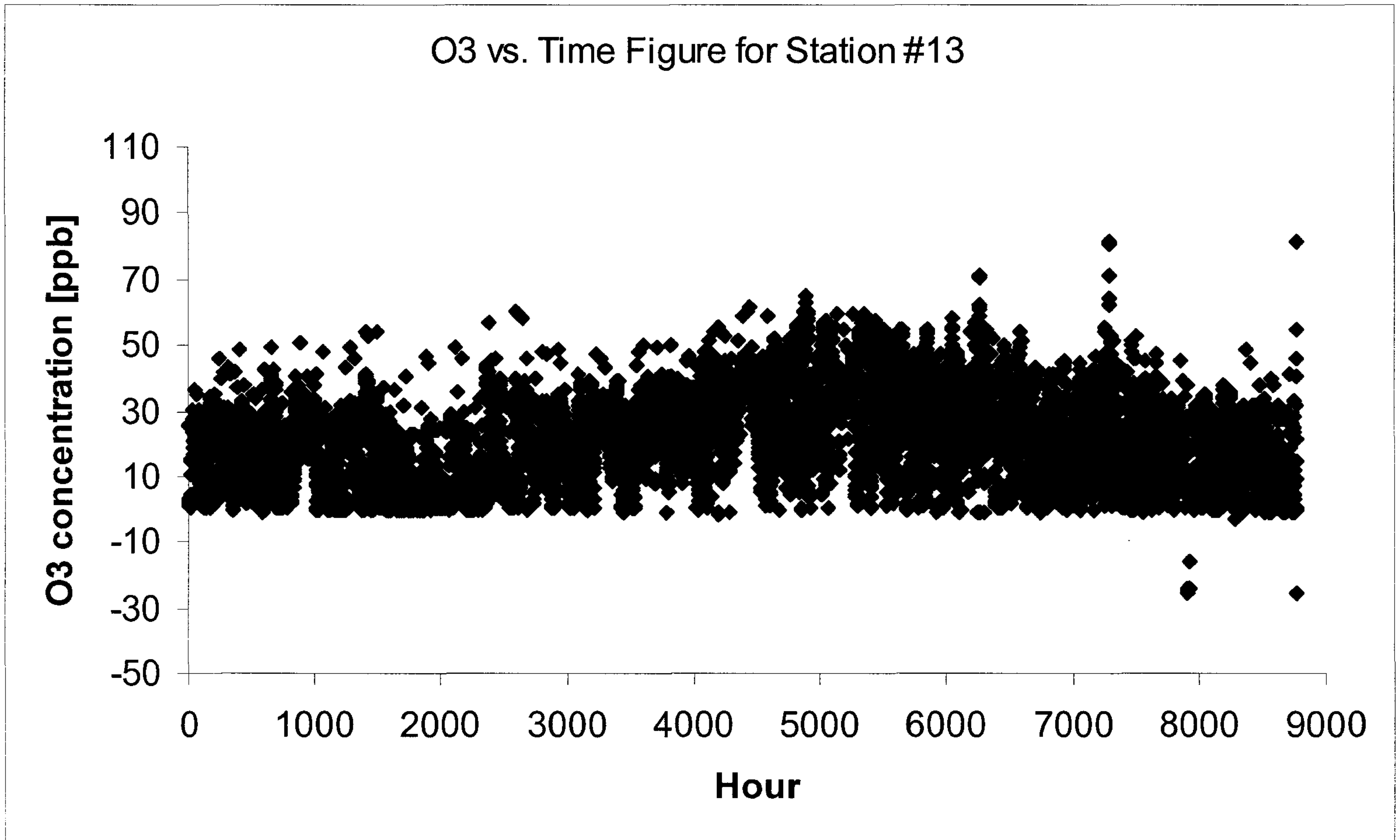


Figure A17. O_3 [ppb] versus hour plot for AMS #13.

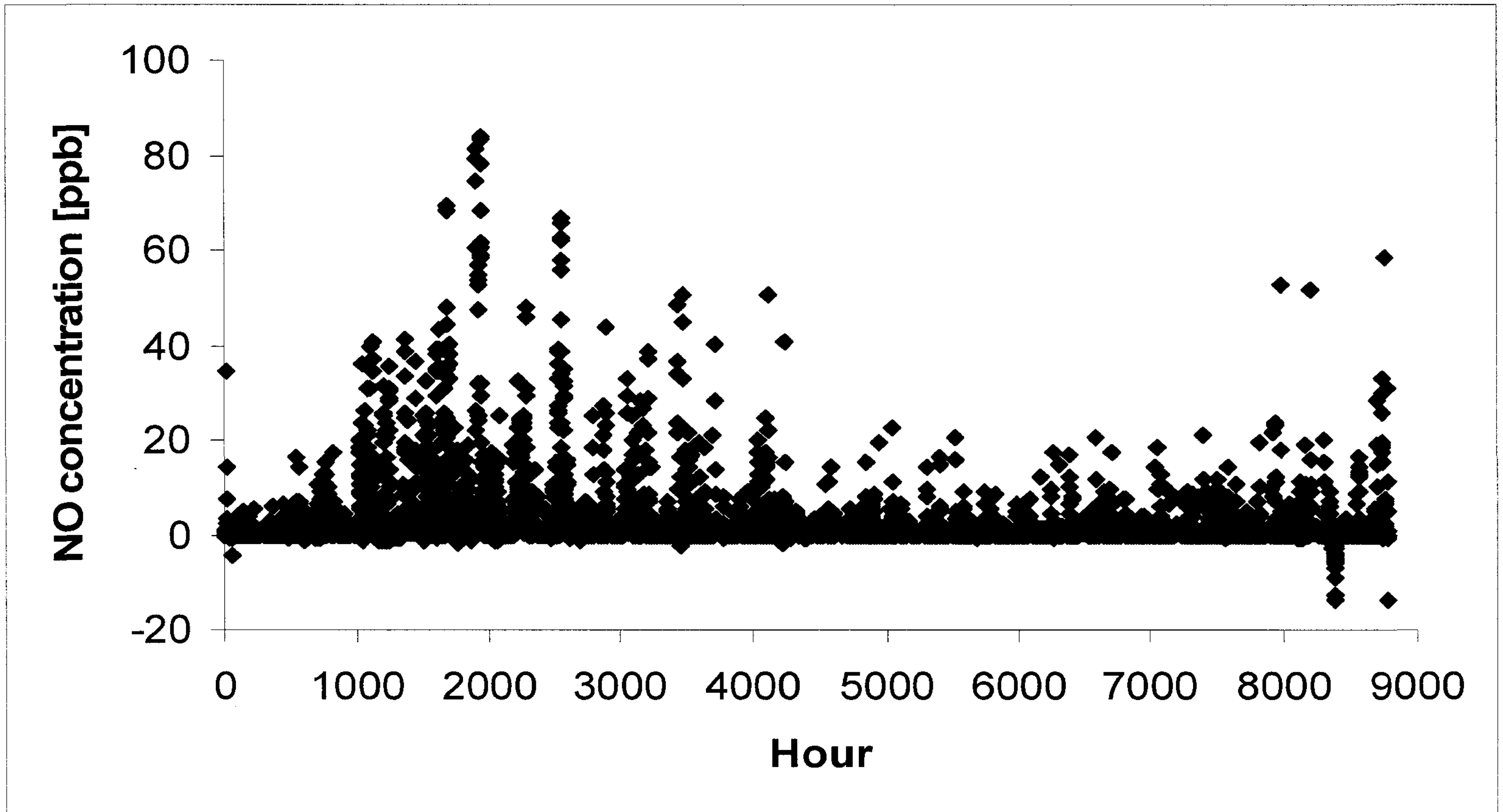


Figure A18. NO [ppb] versus hour plot for AMS #13.

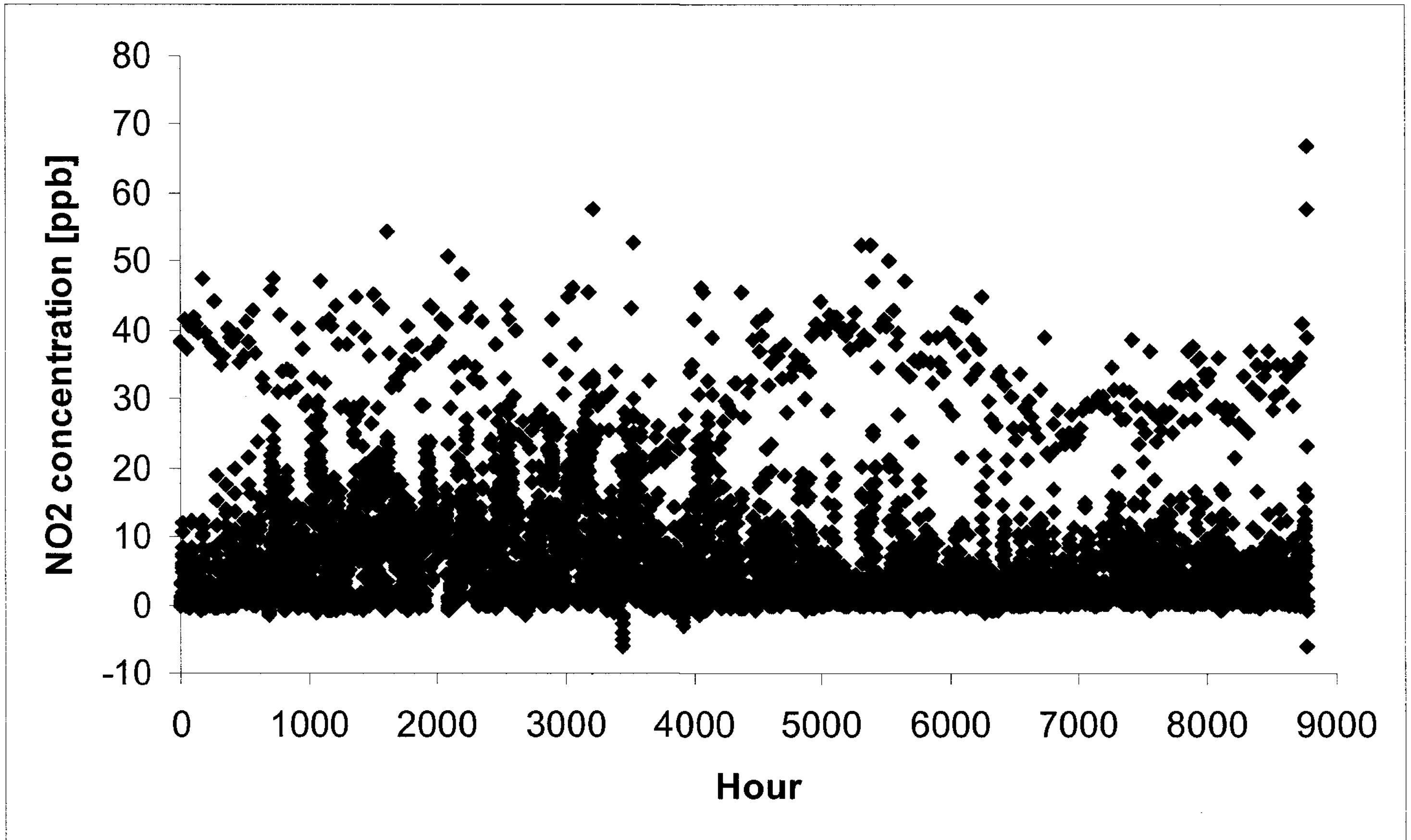


Figure A19. NO_2 [ppb] versus hour plot for AMS #13.

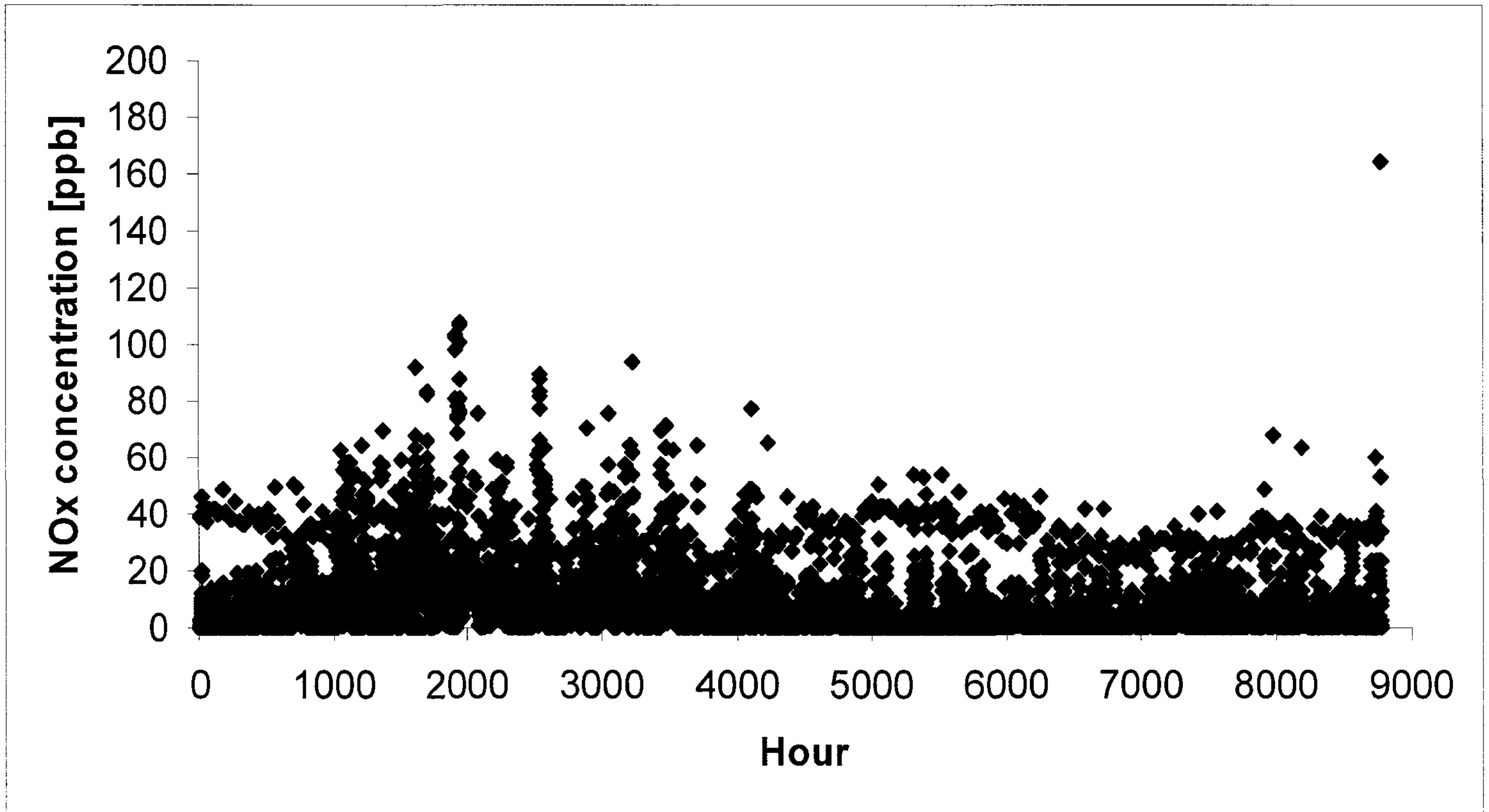


Figure A20. NO_x [ppb] versus hour plot for AMS #13.

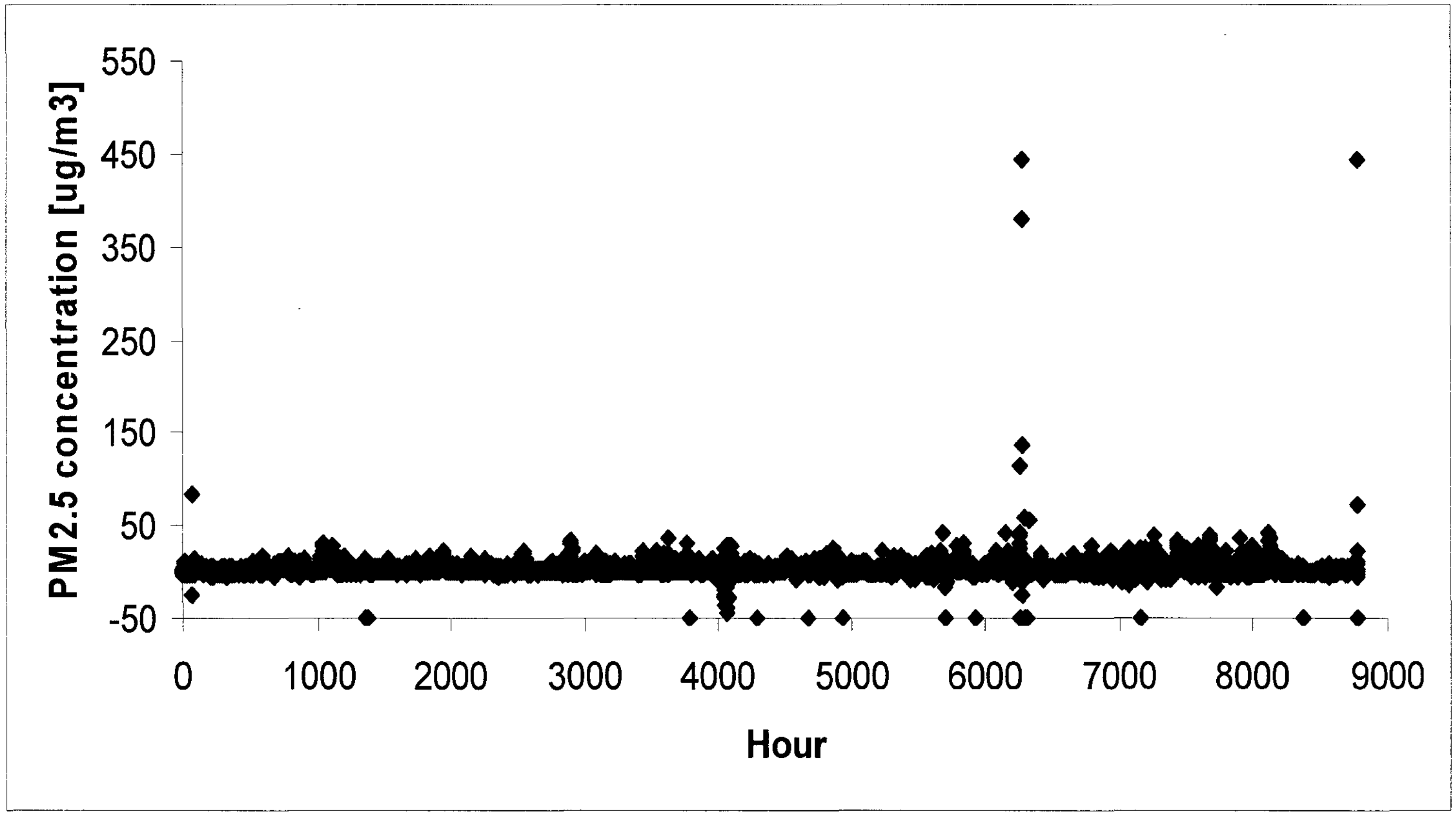


Figure A21. $PM_{2.5}$ [$\mu g/m^3$] versus hour plot for AMS #13.

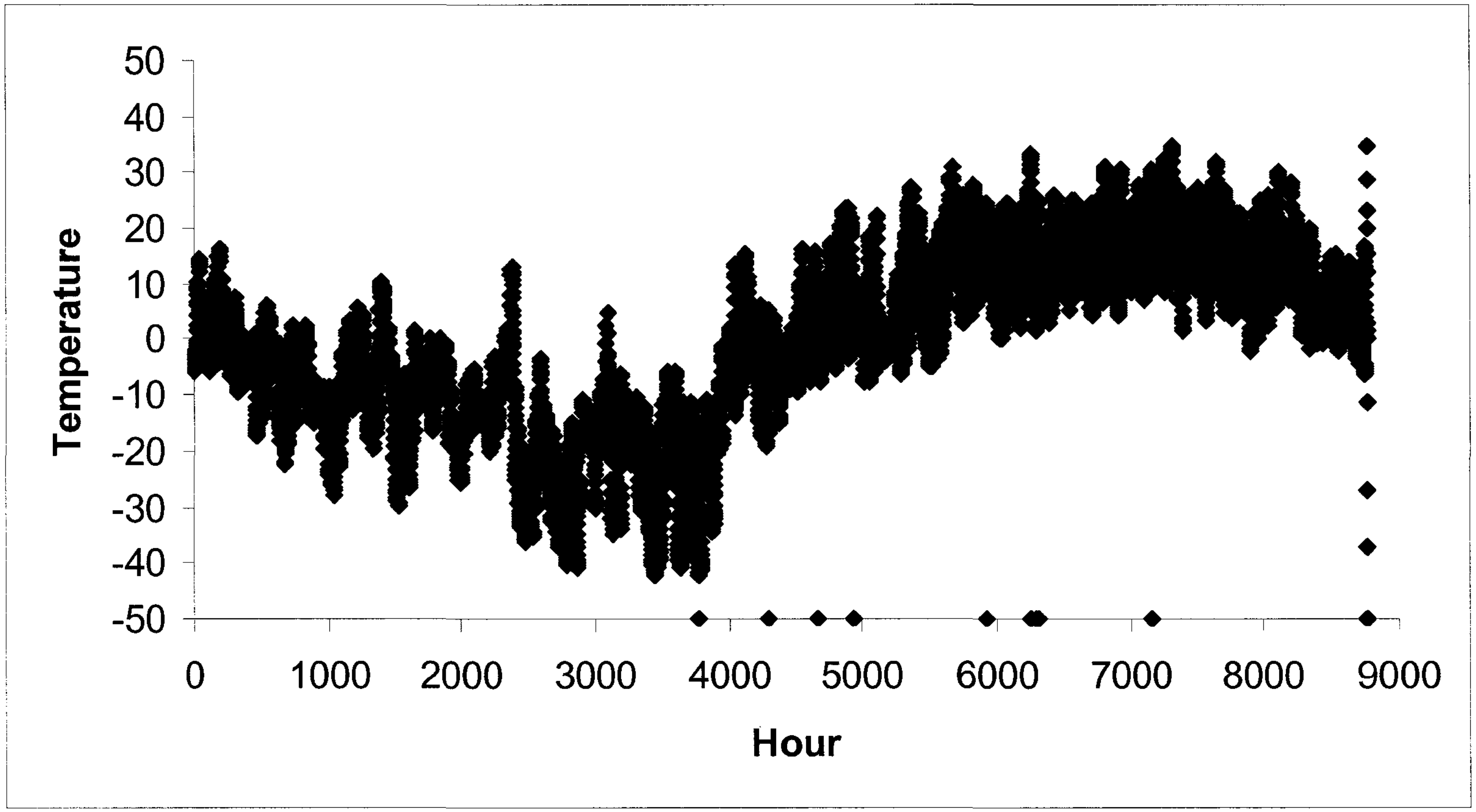


Figure A22. Temperature [°C] versus hour plot for AMS #13.

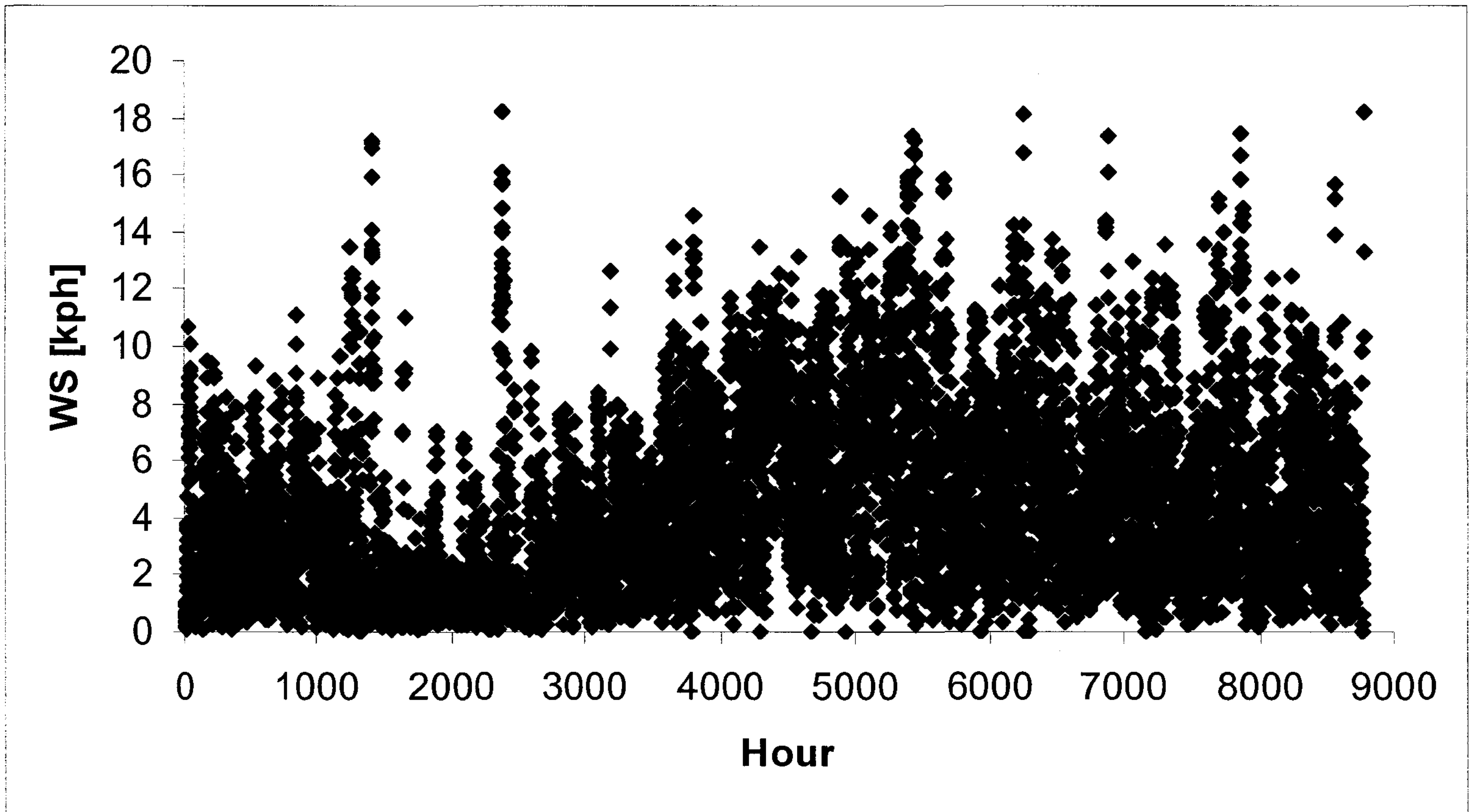


Figure A23. WS [kph] versus hour plot for AMS #13.

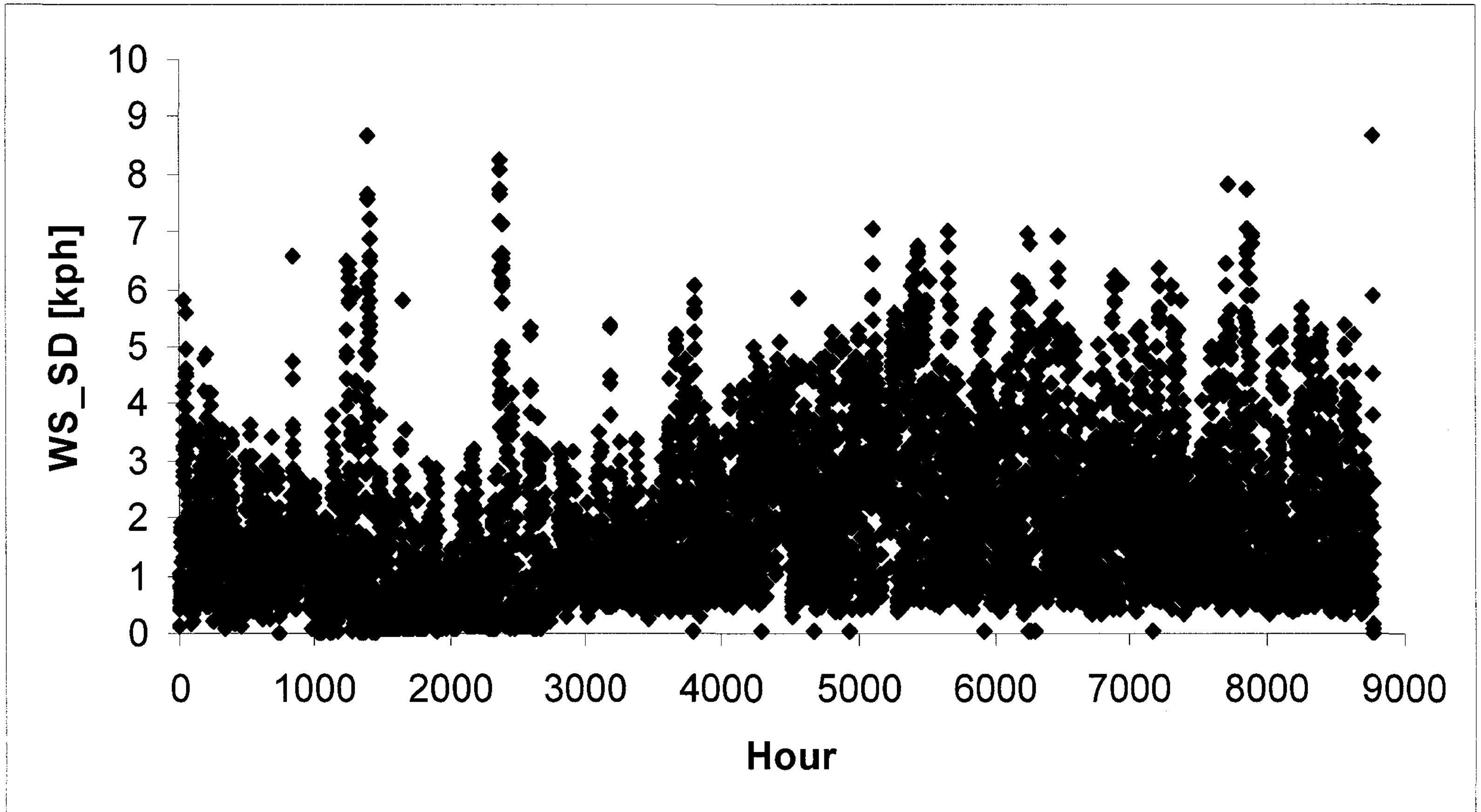


Figure A24. WS_SD [kph] versus hour plot for AMS #13.

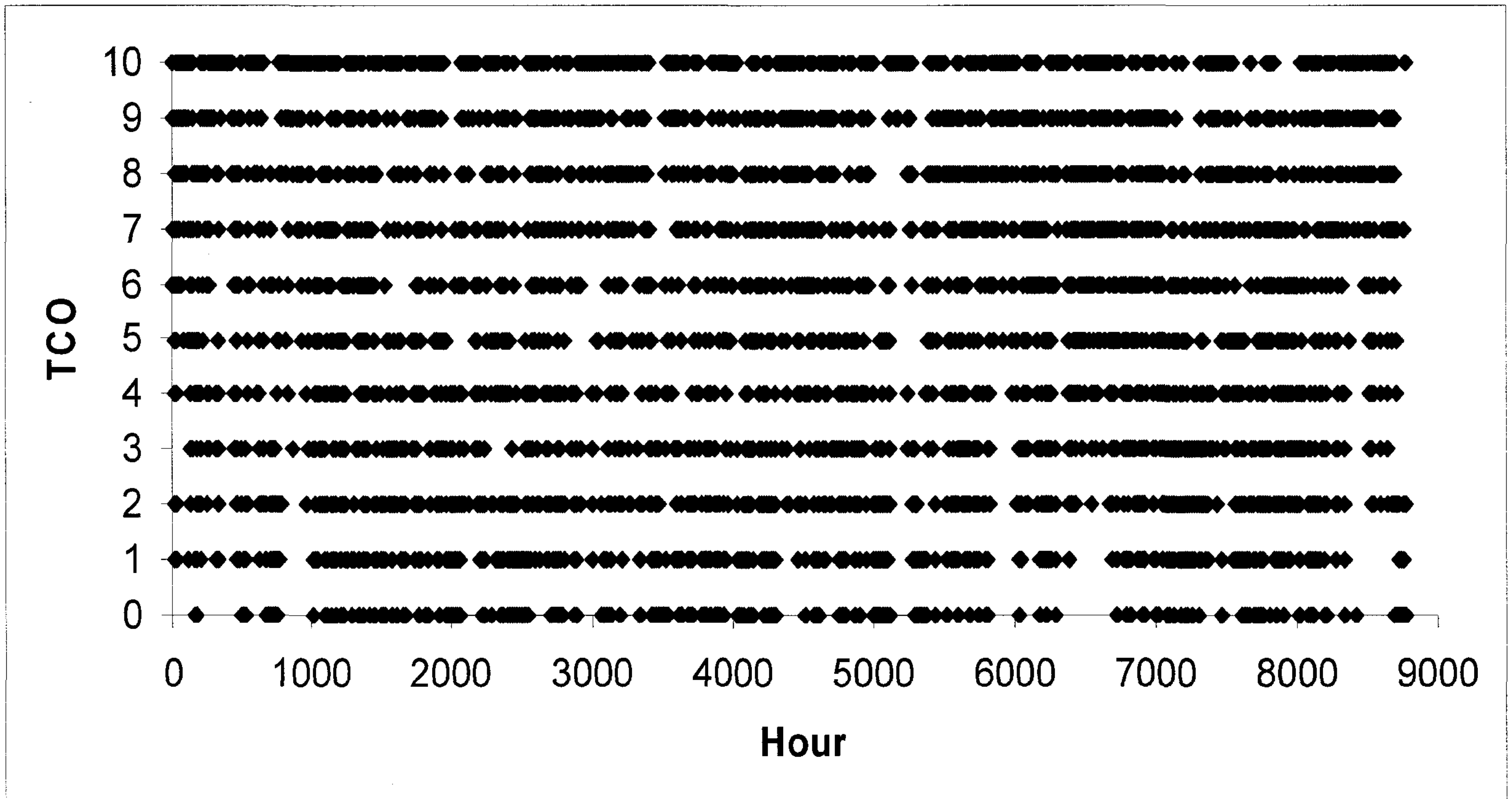


Figure A25. TCO vs. hour plot for Fort McMurray airport meteorological monitoring station.

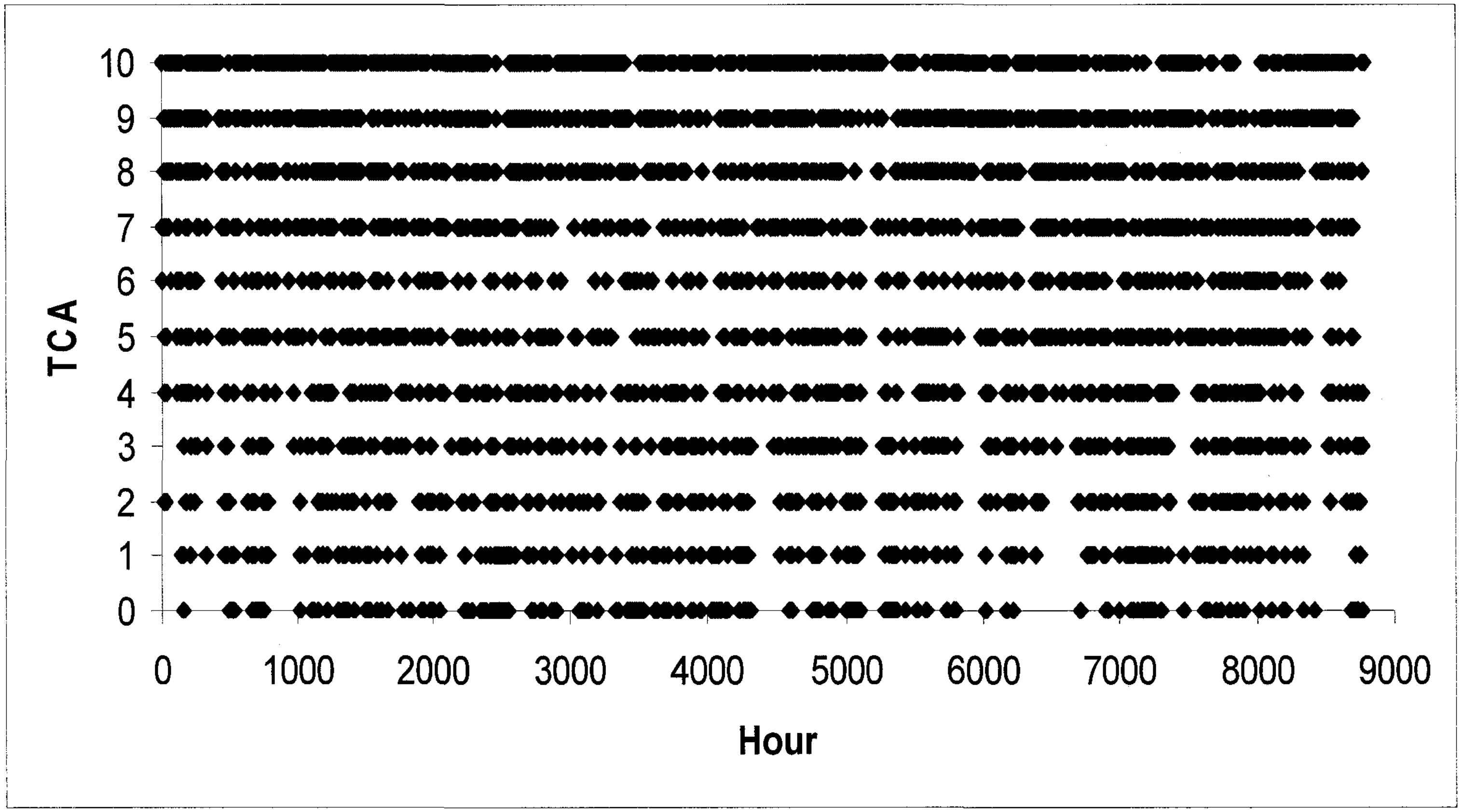


Figure A26. TCA vs. hour plot for Fort McMurray airport meteorological monitoring station.

APPENDIX D – HOURLY $PM_{2.5}$ PREDICTION AFTER $PM_{2.5}$ CONCENTRATION AT AMS #13 INCREASED BY 25%

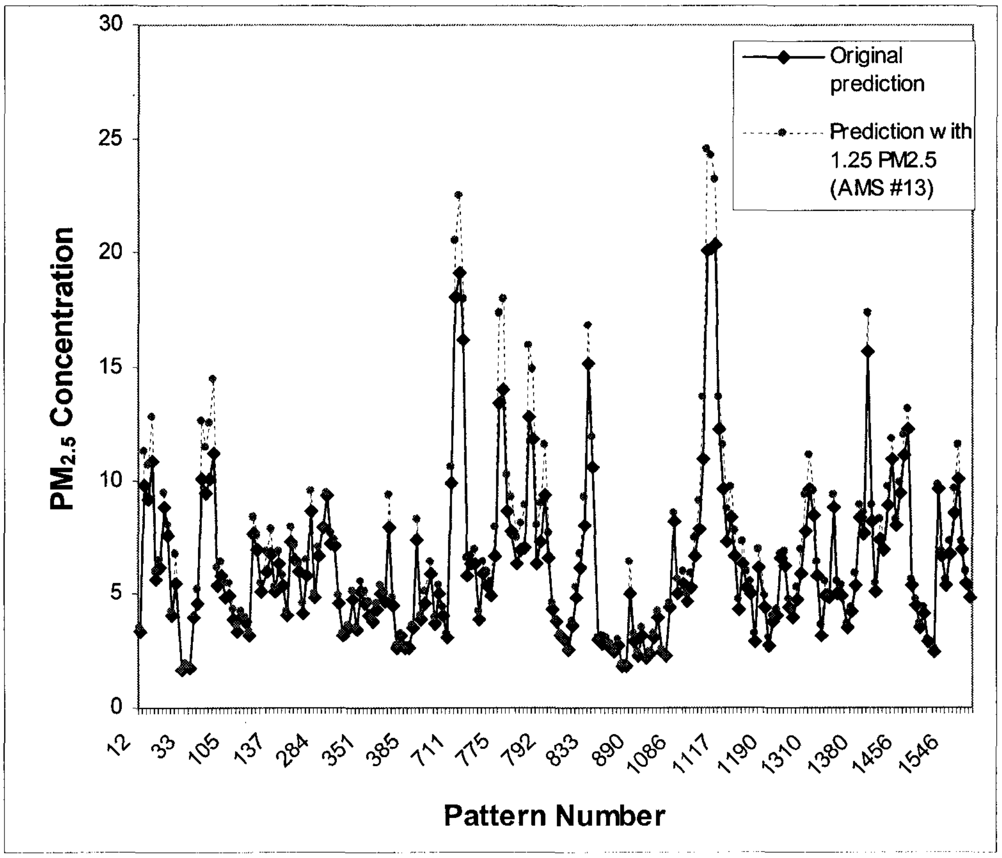


Figure A27. **$PM_{2.5}$ concentrations at Fort McKay station [$\mu g/m^3$] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 25% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.**

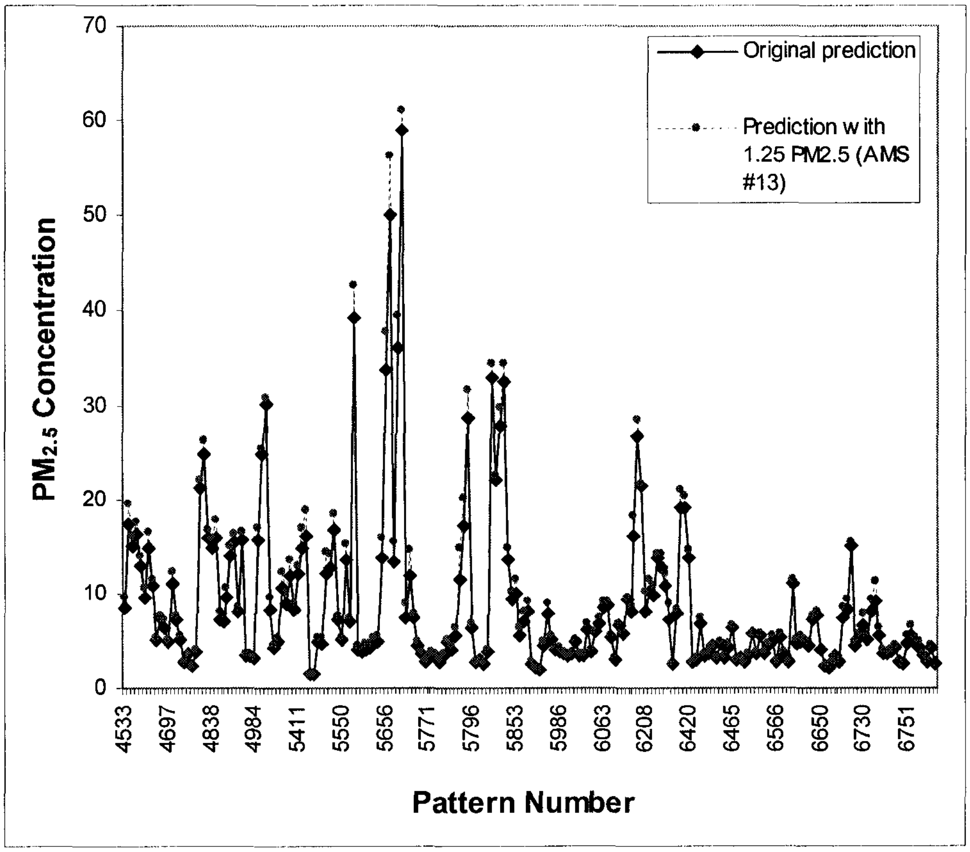


Figure A28. **$PM_{2.5}$ concentrations at Fort McKay station [μg/m^3] based on sensitivity analysis of ANN sub-model using category #1 data (205 data patterns) – original prediction versus simulated 25% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.**

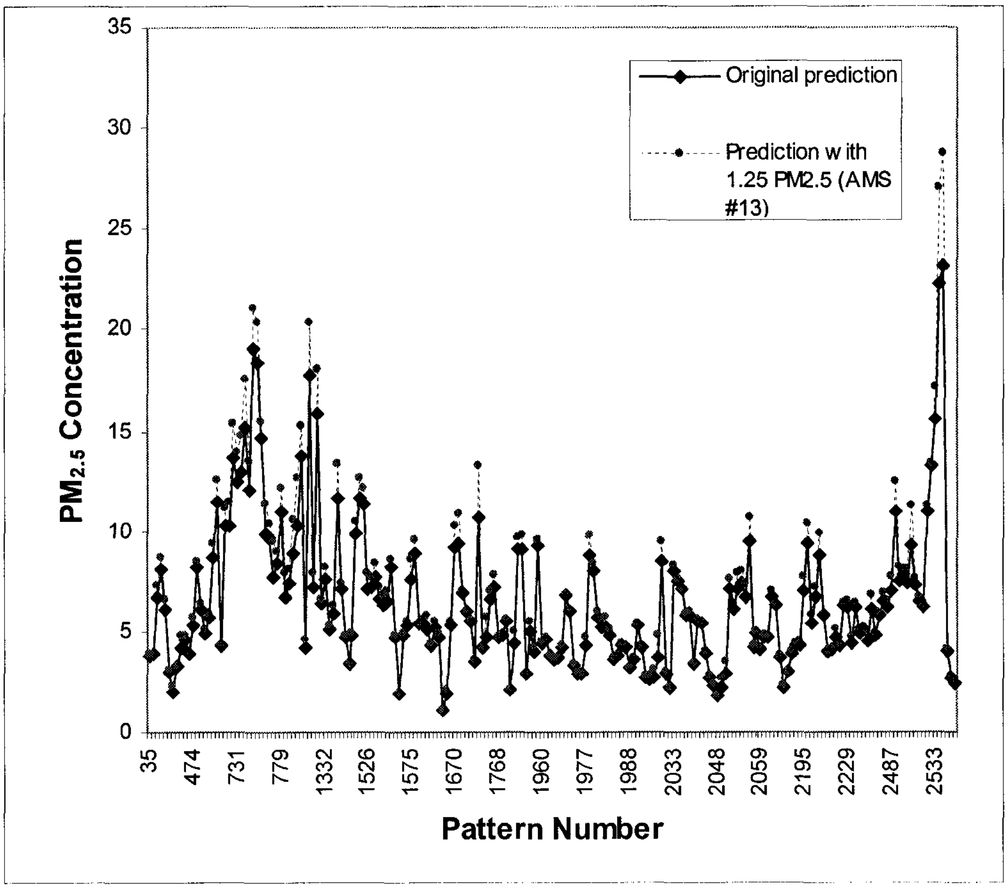


Figure A29. $PM_{2.5}$ concentrations at Fort McKay station [$\mu g/m^3$] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 25% increase in $PM_{2.5}$ concentrations at AMS #13 for typical conditions at AMS #1.

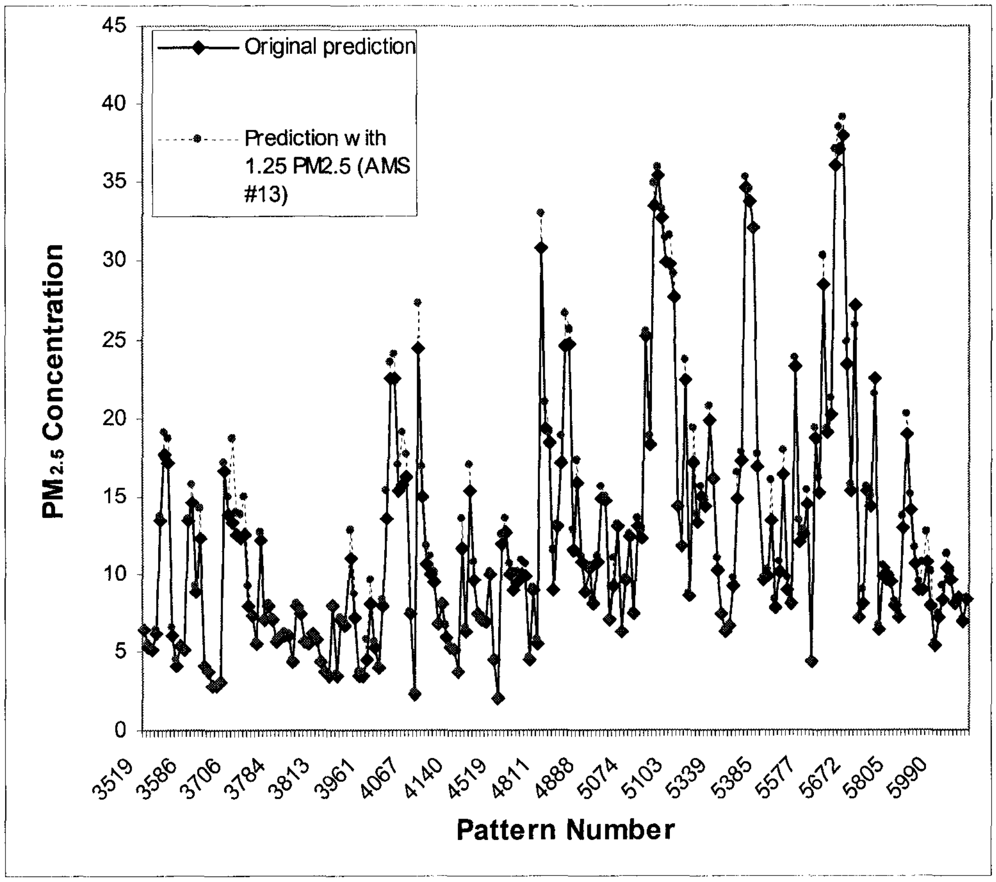


Figure A30. $PM_{2.5}$ concentrations at Fort McKay station [μg/m^3] based on sensitivity analysis of ANN sub-model using category #2 data (205 data patterns) – original prediction versus simulated 25% increase in $PM_{2.5}$ concentrations at AMS #13 for peak events at AMS #1.