LUMA-GIS Thesis 9

Development and Implementation of Air Quality Data Mart for Ontario, Canada

A case study of air quality in Ontario using OLAP tool

Samira Muhammad

2010
Dept. of Physical Geography and Ecosystems Analysis
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

A Master thesis presented to

Department of Physical Geography and Ecosystem Analysis

Centre for Geographical Information Systems

of

LUND UNIVERSITY

by

SAMIRA MUHAMMAD

in partial fulfilment of the requirements

for the degree of Master in Geographical Information Science

Supervisor:

Dr. Helena Eriksson, Lund University

ACKNOWLEDGEMENTS

The author wishes to thank the supervisor Dr. Helena Eriksson at the department of Physical Geography and Ecosystems Analysis at Lund University Sweden for providing helpful comments, suggestions, edits and valuable time in completion of my Master's thesis.

All the teachers who have provided excellent teaching support and have been a source of information throughout the entire program, Dr. Micael Runnstrom, Dr. Lars Harrie, Dr. Petter Pilesjo, Ulrik Martensson, Roger Groth and Jean Nicolas Poussart.

Ontario Ministry of Environment for the availability of online air quality historical data and Tony Munoz from the Air Quality Assessment and Reporting Unit for providing extended information on instrumentation deployed at the air quality monitoring stations.

Last but not least I would like thank my husband and two beautiful kids for their continuous support and patience for which I am deeply grateful.

Samira Muhammad

ABSTRACT

Air Quality Data Mart

This thesis describes the development and implementation of Air Quality Data Mart for Ontario Canada using *Online Analytical Processing* (OLAP) tool. It is followed by a case study which presents comparisons of air quality between the urban and rural areas, peak and non-peak hours, working days and weekends for various cities in Ontario. The purpose of this study is to develop a user friendly tool for historical air quality data and evaluate the functionality of the tool by extracting the data across several dimensions. The data for air quality is available on the Ontario Ministry of Environment website for 43 monitoring stations across Ontario. This data is in the form of static *Hyper Text Markup Language* (HTML) pages which cannot be used for analytical purposes.

Air quality data mart was developed using open source OLAP. The database was designed using multidimensional modeling approach. OLAP server “Mondrian” was used as the presentation server whereas “Openi” client was used as an end user tool for this study. The different functions available in this data mart are: rollup, drill down and slice and dice the data across several dimensions such as time, location and pollutant.

The most important conclusion of this thesis is the successful implementation of an air quality data mart with the possibility to extract accurate historical air quality data. The data in the form of a data mart provides numerous advantages, where it can be analyzed according to the required analytical perspective for a given city/cities. The only drawback of having data in the form of a data mart is that, if the data is drilled down to the finest precision i.e. to the hour (depending on the number of dimensions selected) the resulting chart will be very crowded but the generated report will present a complete overview of the analysis.

Keywords: OLAP, Data mart, Multidimensional modeling, Air quality, Ambient Air Quality Criteria, Canadian Environment Sustainability Indicator.

POPULAR SUMMARY

Air quality data mart built in this study consists of historical air pollutant data for cities across Ontario. This data is available at the Ontario ministry of environment website from 2000-2007 for most of the cities. There is missing data for some of the cities and some pollutants as well. It is because the monitoring of those specific pollutants did not commence in or before 2000.

This data mart facilitates the user to extract historical data. The user does not need to know query language skills. With user friendly interface data analysis can be performed using drag and drop feature. The statistical functions implemented in this data mart are AVG, MIN and MAX. The database is designed in a form where it is possible to extract data for a specific time period. This data can be further filtered based on location and a given pollutant. The query output is in the form of a chart and a table. It is also possible to save the analysis which can be accessed again in future for referential purposes.

The air quality in Ontario was compared using parameters like weekday – weekend effect, peak hour – off peak hour, urban and rural areas etc. Most of these parameters were used for the cities of Toronto and Ottawa. The results generated by the data mart showed that the pollutant concentration levels in both cities surpassed the recommended guidelines, but there were less exceedance days monitored in Ottawa compared to Toronto.

This data mart lacks the component of visual maps. It would be interesting if this data mart had a choropleth map associated with the pollutant measurements. This would enable the user to visually comprehend the air quality data. However, if there are limitations in air quality data mart for Ontario, Canada at this point it is an indicator that spatial maps with spatial statistics can be implemented in future.

GLOSSARY

AAQC	Ambient Air Quality Criteria
AQUIS	Air Quality Information Systems
CCME	Canadian Council of Ministers of the Environment
CEPA	Canadian Environmental Protection Act
CESI	Canadian Environment Sustainability Indicator
CO	Carbon Monoxide
CWS	Canada Wide Standards
DM	Dimensional Modeling
DSS	Decision Support Systems
EMRB	Environmental Monitoring and Reporting Branch
EPA	Environmental Protection Act
ETL	Extract Transfer Load
HTML	Hyper Text Markup Language
ICAP	Illness Costs of Air Pollution
J2EE	Java 2 Enterprise Edition
JDBC	Java Data Base Connectivity
JRE	Java Runtime Environment
NO	Nitric Oxide
NO_2	Nitrogen Dioxide
NOx	Nitrogen Oxides
O_3	Ozone
OEE	Office of Energy Efficiency
OLAP	Online Analytical Processing
OMA	Ontario Medical Association
OMOE	Ontario Ministry of Environment
PM	Particulate Matter
POI	Point of Impingement
QA	Quality Assurance
QC	Quality Control
SO_2	Sulphur Dioxide
SOx	Sulphur Oxides
XML	eXtensive Markup Language

Contents

LIST OF FIGURES:

LIST OF TABLES

1. INTRODUCTION

Air pollution was once assumed to be a summer and urban area problem, with increasing population, traffic, industries and human related activities which unfortunately no longer seems to be true. It affects the human health all year round even in the rural areas of Canada.

When we hear the term pollution it often makes us think of water pollution, waste pollution, noise pollution etc. This is because we are able to sense it, in the form oil spills, toxic waste from the factories, noises from different sources like cars, trains, wind turbines etc. Air pollution is often regarded less harmful in nature by an average person whereas in reality it can silently affect the well-being of individuals. However, it is the human related activities that contribute towards air pollution the most.

Human related activities associated with air pollution involve household emissions burning of wood for heating or other purposes which results in smoke in the air, coal fired power plants, emissions from the tailpipes of the auto motor vehicles, diesel run trains, cargo ships, chemical industries etc.

Exposure to acute air pollution over the longer period of time has negative health impacts on individuals. Illnesses such as chronic respiratory problems, cough, common cold, flu, viral infections, bacterial infections, ear infections, sinus problems, wheezing, shortness of breath and asthma are all result of air pollution (Health Canada, 2008).

Every year, there are approximately 6000 additional deaths in Canada because of short term exposure to air pollution and research suggests that 69% of these deaths come in the form of cardio and cerebrovascular disease (Heart & Stroke Foundation, 2008).

Knowledge and awareness about air pollution is important. There are several non-profit and non-governmental organizations working towards awareness and the education of the general public (Evergreen Foundation, 2000). Ontario ministry of environment has historical air pollutant data available for the general public. The pollutant data is available for monitoring stations across Ontario. The data presented on this website is on an hourly basis for the selected pollutant and time frame. However this data is hard to interpret. It could cause confusion if an individual is unaware of the acceptable *Ambient Air Quality Criteria* (AAQC). This issue can be handled if the data is stored in the form of a “Data Warehouse” or “Data Marts”.

Data warehouses are databases which are designed in a manner to increase accessibility to a wide range of data coming from different sources. The main objective of a data warehouse is providing accurate, timely and useful information to business analysts and decision makers. It is not a standalone product but a group of products coming from different sources. e.g. human resources, marketing, finance, sales, payroll and customers etc.)

Although data warehouses are an impressive proposition for any organization, which may yield positive results but technically as well as financially it is an enormous task to achieve. Therefore it is not uncommon to see organizations initially start managing, organizing, smaller data sets or data stores. These data stores are known as “Data Marts”.

A data mart is “*a small single subject data warehouse subset that provides decision support to a small group of people*” (Rob & Coronel, 2009). Data obtained from larger data warehouses with specific functionality can be termed as data mart. Building and implementation of a data mart is an attainable task rather than an ongoing process. The benefits of implementing a data mart instead of a data warehouse are, rapid and consistent query response times, it is also easier to address the issues and essential data requirements on a smaller scale rather than on a large scale project.

Online analytical processing (OLAP) tools support specialized form of analysis. They are widely used when analytical navigation of the data is required. For example, decision making, time series, statistical and trend analysis. The data is organized in a multidimensional form, where it is possible for a user or an analyst to examine data from multiple perspectives.

Data stored in multidimensional form has a set of advantages. Data analysis techniques used with multidimensional data have, enhanced data presentations where the data is presented in the form tables, charts, reports etc. Data aggregation and classification is the most important feature in an OLAP tool, where it possible to drill down, roll up and slice and dice the data across different dimensions and levels of aggregation. It is important to keep the end user interface of an OLAP tool user friendly where data extraction and data analysis tasks are simpler to perform.

The aim of this study is twofold:

1. To develop and implement air quality data mart for Ontario, Canada using Online Analytical Processing (OLAP) tool, making it possible for the user to present statistics and perform simple analyses regarding atmospheric pollution in a well defined area.

2. To determine the working functionality of the model by various comparisons. Specific objectives of this aim are

    a. To determine for 2005 the average annual concentration levels of SO_2, CO, NO_2, O_3 and $PM_{2.5}$ for the cities of Toronto, Ottawa, Hamilton, London and Windsor.

    b. A comparison of smog elements (O_3 and $PM_{2.5}$) between the urban cities (Toronto, Ottawa, Hamilton, London, Windsor) and rural areas (Sudbury, Thunder Bay, Sault Ste Marie, Sarnia, Chatham) for 2005.

    c. A comparison between Toronto and Ottawa, of monthly maximum concentration levels of O_3 and $PM_{2.5}$ during the “working days” and “weekends” for the year 2007.

    d. A comparison between Toronto and Ottawa, for maximum concentration levels of O_3 and $PM_{2.5}$ during the “Peak Hours” and “Non Peak Hours” for the period of April 1st to Sept 30 in 2007.

e. Trend analysis for O_3 and $PM_{2.5}$ for the city of Toronto and Ottawa (2003-2007).

f. Trend analysis (2000-2007) of ground level ozone for the city of Toronto and Ottawa.

This thesis consists of four main chapters after the introduction. The main focus of this thesis is, to develop, implement Air Quality Data Mart and evaluate the working functionality of the tool by presenting a few statistical analyses. In Chapter 2 (Theory) the focus is on what kind of air pollutants are monitored and what are the health risks of being exposed to these pollutants for longer periods of time. The policies and standards in place to control the pollutant levels are discussed. Chapter 3(Methods) includes two parts, The first is the introduction to the study area, database design, air quality data mart architecture, specific approaches for the migration and implementation, The second part of this chapter includes the data analysis where a comparative study is done between the city of Toronto and Ottawa across several dimensions. In Chapter 4 (Results) from the analysis are presented. In Chapter 5 (Discussion) revolves around the data analysis techniques, results achieved shortcomings of the data, and future directions are touched upon.

2. THEORY

2.1 Decision support systems

In recent years, with increased use of web applications and overwhelming availability of raw data collected through operational systems, the focus of many companies is to build a system that provides better data handling, management, storage and manipulation. Data organized in a way that is easy to retrieve and manipulate helps companies in better understanding their customer needs and makes grounds for rational decision making usually known as *Decision Support Systems* (DSS).

2.1.1 Data warehouses

The objective of a decision support system in a complex business environment is to facilitate crucial decision making where requirements vary with individual client needs. As the range of information and complexity is increased, the process of retrieving relevant data becomes complicated as well. Thus a new method of storing data was developed known as “Data warehouse”.

According to Inmon (Rob & Coronel, 2009) the term data ware house is defined as “*an integrated, subject oriented, time variant, nonvolatile collection of data that provides support for decision making*”

Integrated, subject oriented, time variant and nonvolatile are all important aspects of a data warehouse. A brief description about each is described below.

<u>Integrated</u>

Integrated characteristic of a data warehouse is that the data is stored for the entire organization in a consolidated database coming from different sources or departments. The underlying problem with integrated data warehouses is that, the data formats should be the same throughout the entire organization. It may seem like an achievable task but very often different departments store data slightly differently. For example if an entity needs to be stored as a “true” or “false” value some departments may prefer to save it as “0” or “1” or may be “T” or “F” etc. Slight variances of data formats can cause problems when retrieving data.

It is essential for the entire organization to conform to a common format that is acceptable throughout the organization. It is a time consuming task but once achieved it facilitates in better decision making. Integrated data warehouses also provide better understanding of business needs and client requirements which leads to better strategic approaches.

<u>Subject Oriented</u>

This feature of the data warehouse as the name suggests, the data is stored in an organized manner. The organization and summarization of the data is by department, by topic, by the number of sales, by customer, by product etc. These types of data warehouses prove to be more efficient when performing data retrieval, report generation and decision making tasks. Data organized in a subject oriented manner also proves to be beneficial when information is required from a certain perspective. For example businesses may require, number of sales from a specific location, or determine the product in high demand. This form of organized data will fulfill the end user's queries.

Time Variant

The third important feature of a data warehouse is time variant. Each record or transaction in the data warehouse is allocated with a time-id which cannot be deleted or updated. These data warehouses are generally very helpful when doing comparative analysis (time series) or trend analysis. For example a company would like to compare its business activity over the period of several years, or comparison between the current year and the year before. This type of information presents a performance history in the years past and facilitates improved decision support activities.

Nonvolatile

Nonvolatile data warehouses as the terms suggests are "untouchable". The data in the warehouse once stored is not deleted or modified it becomes historical. In conventional databases the data is updated once the transaction has been completed, in case of data warehouses the data is continually added and never removed. Data warehouses prove to be beneficial when a summary report of a business is required based product type, number of sales, sales per region, or purchase orders by customers and so on.

Before the implementation of a data warehouse, it is crucial to determine what purpose the data warehouse would serve. Often time's larger and complex businesses suffer from information management problem. There may be several underlying reasons for these sorts of problems, i.e. data coming from different operational systems, restrictive or limited access to data, data formats, platform compatibility issues etc. This can make information retrieval and decision support tasks extremely difficult.

It is very important to take into consideration that data warehouse is not a single product but an environment, which consists of four major components. Each component is briefly discussed below.

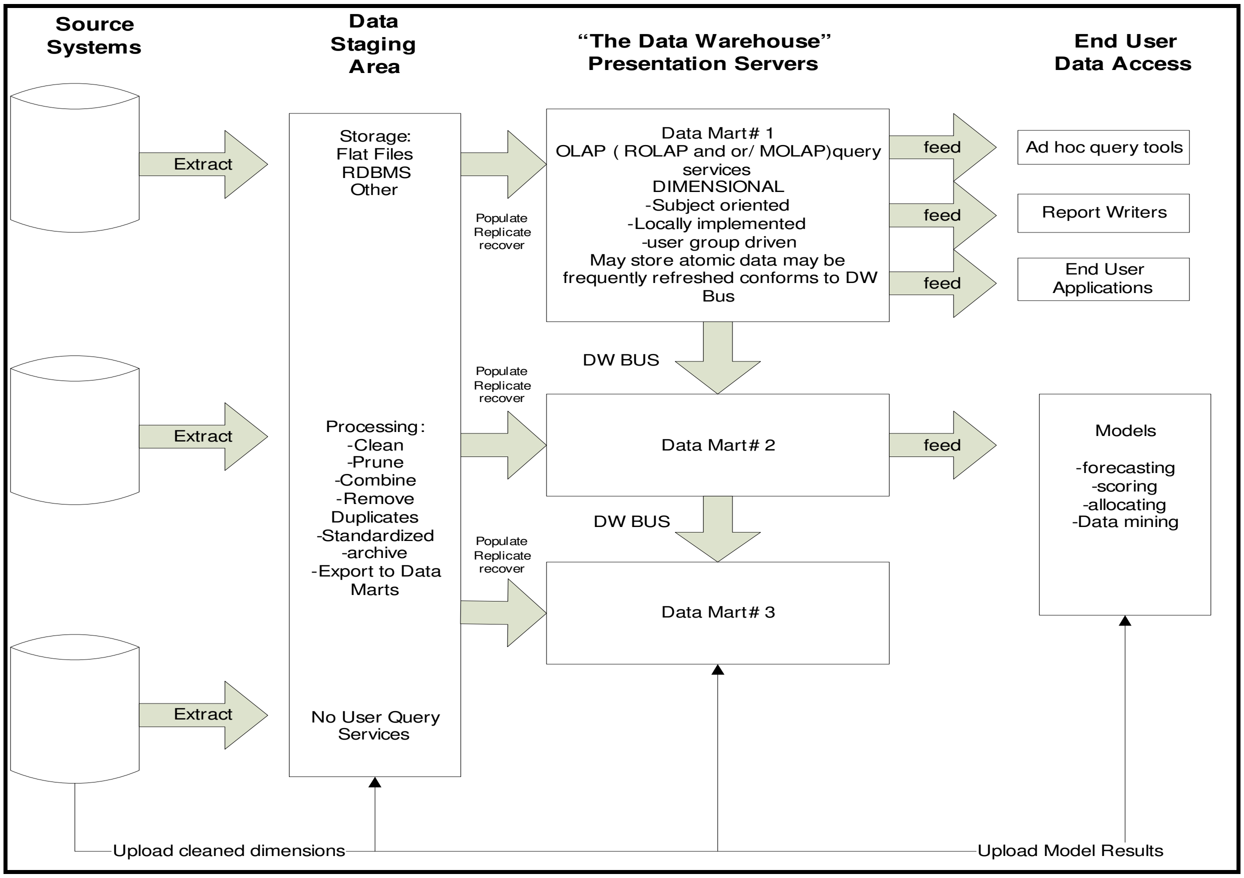


Figure 1: **Adopted from: The Data Warehouse Lifecycle Toolkit - Components of a Data warehouse Architecture (Kimball, 1998,p.15).**

Source Systems

A source system is the first major component of a data warehouse. They are transactional systems of an organization and are often called legacy systems in a mainframe environment. The main objective of source systems is timely availability of data. Queries against these systems are highly rigid in nature, usually a part of normal transaction flow. Source systems maintain limited historical data and are not queried in a broad and unexpected ways as the data warehouses.

Data Staging Area

This is the second major architectural component of a data warehouse, where the data is stored, transformed and prepared to be loaded into the data warehouse. It is an important phase between the source systems and the presentation server where tasks such as cleaning, removing duplicates or invalid values, data formatting, field formatting, reformatting data structures etc. are done. At this stage no query and presentation services are provided.

Presentation Server

It is the third component in the architecture of a data warehouse. Presentation server is the machine where, data is organized and stored for the end users. The data is dimensional and can be queried with respect to one or more dimensions based on the user or application need. A dimensional model consists of fact table and dimensions table. Where fact table contains measurements of an organization and dimensional tables contain attributes or fields that can be used for grouping data.

End User Tools

End user tools are used for querying, analyzing and presenting data. A minimum requirement of these tools is to provide users with data access, presentation of data in the form of spreadsheets or graphs and user friendly interface where the data can be queried and presented according to one or more dimensions.

2.1.2 Air quality data mart architecture and implementation

In previous section (2.1.1) a brief introduction of data warehouse architecture was presented. In this section the architecture of air quality data mart for Ontario Canada is presented. The implementation process consists of four major components. Source systems, data staging area, presentation server and end user data access (figure2).

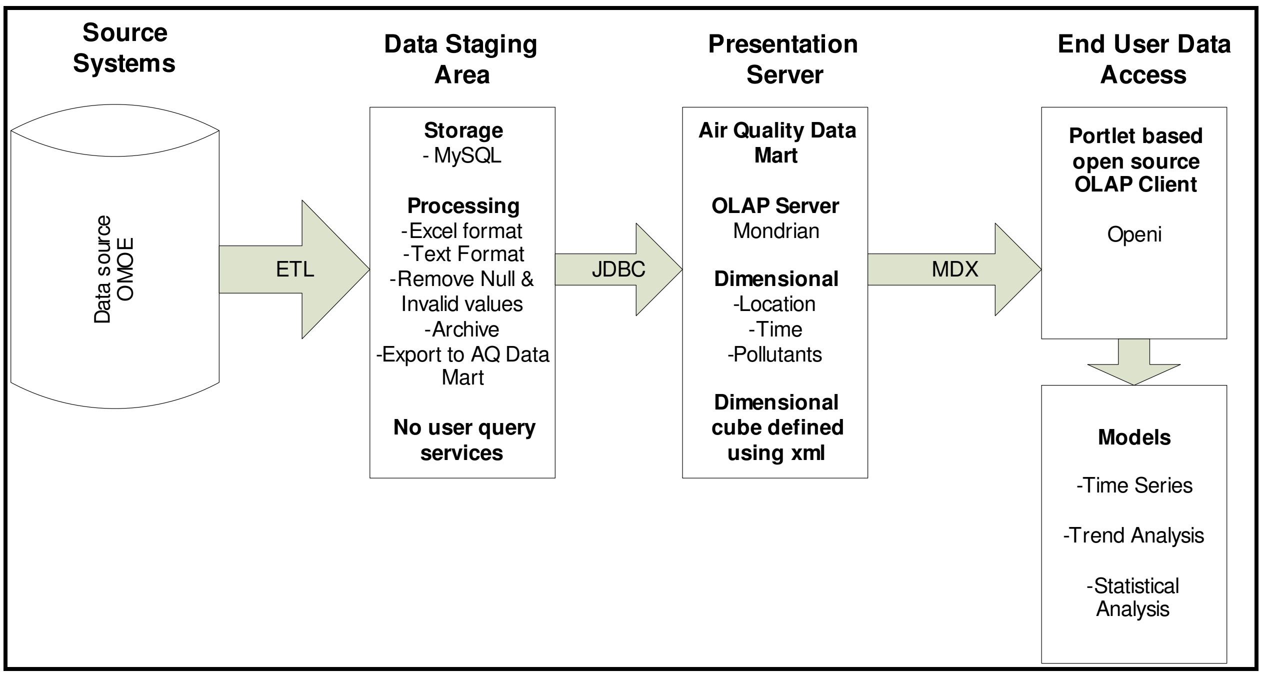


Figure 2: Architectural components of air quality data mart for Ontario Canada.

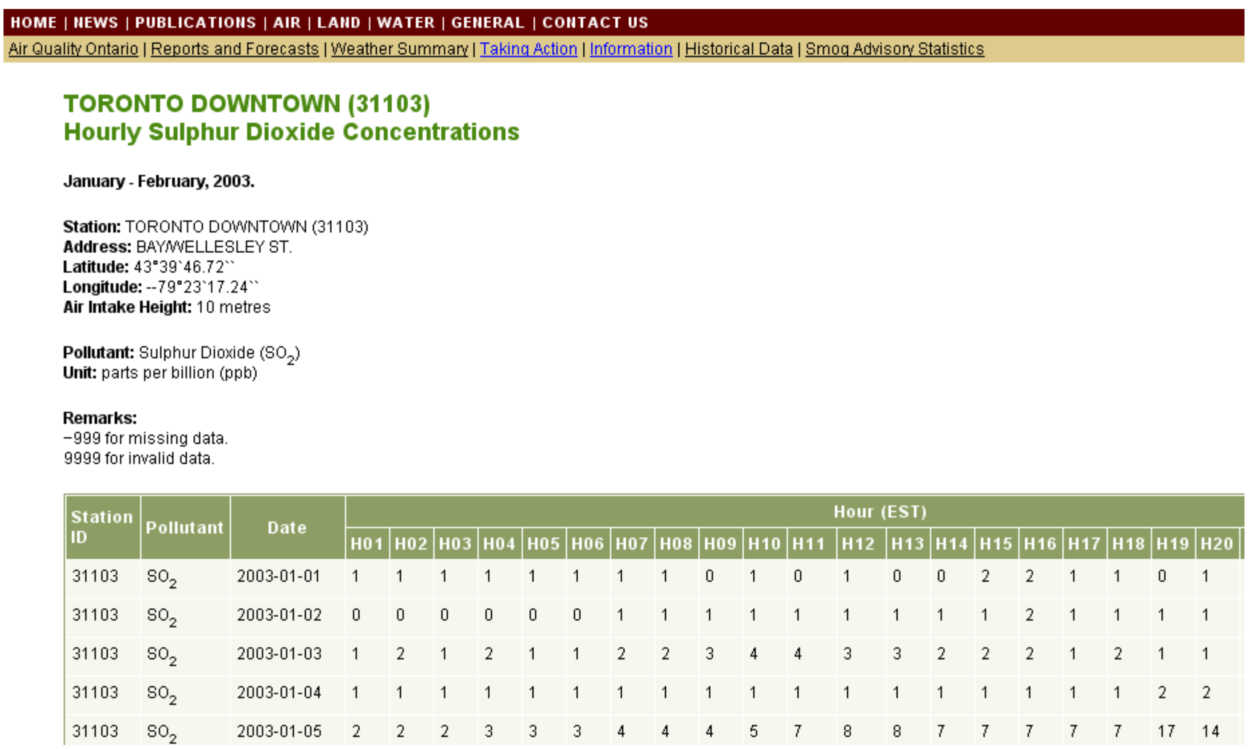
Source Systems

Ontario Ministry of Environment is the sole source system for air quality data mart. The data is available for different cities in Ontario dating from 2000 - 2007. There are several cities which do not have the data starting from the year 2000, this is mainly because air monitoring did not begin by the year 2000 but started a few years later. Similarly not all sites have the data for all pollutants mentioned above initially but gradually a few years later started the monitoring for most of the pollutants. The historical air pollutant data can be extracted by defining the categories in the search form (figure 3), e.g. station name, year, pollutant, start month, end month, output type etc. The data is in hourly format for the selected time period (figure4).

Search Historical Air Quality Pollutant Data

Search Historical Air Quality Pollutant Data	
Data Use*:	Academic Research
Station Name*:	Toronto Downtown
Year*:	2003
Pollutant*:	Sulphur Dioxide (SO2)
Start Month:	January
End Month:	December
Output Type:	HTML
	search

Figure 3: Search form for historical air quality pollutant data.



HOME | NEWS | PUBLICATIONS | AIR | LAND | WATER | GENERAL | CONTACT US

Air Quality Ontario | Reports and Forecasts | Weather Summary | Taking Action | Information | Historical Data | Smog Advisory Statistics

TORONTO DOWNTOWN (31103)
Hourly Sulphur Dioxide Concentrations

January - February, 2003.

Station: TORONTO DOWNTOWN (31103)
Address: BAY/WELLESLEY ST.
Latitude: 43°39`46.72``
Longitude: --79°23`17.24``
Air Intake Height: 10 metres

Pollutant: Sulphur Dioxide (SO_2)
Unit: parts per billion (ppb)

Remarks:
−999 for missing data.
9999 for invalid data.

Station ID	Pollutant	Date	Hour (EST)																			
			H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20
31103	SO_2	2003-01-01	1	1	1	1	1	1	1	1	0	1	0	1	0	0	2	2	1	1	0	1
31103	SO_2	2003-01-02	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	2	1	1	1	1
31103	SO_2	2003-01-03	1	2	1	2	1	1	2	2	3	4	4	3	3	2	2	2	1	2	1	1
31103	SO_2	2003-01-04	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	2
31103	SO_2	2003-01-05	2	2	2	3	3	3	4	4	4	5	7	8	8	7	7	7	7	7	17	14

Figure 4: Query results for historical air quality pollutant data.

2.2 Air Quality in Ontario

In this section the properties of different pollutants and their adverse health impacts are presented. Instruments used to collect data for individual pollutants, its precision and site locations across the province of Ontario are discussed and model used to interpolate spatial distribution of pollution are presented. A brief description of government policies in place is provided. *Ambient Air Quality Criteria* (AAQC) for contaminants investigated in this study is presented. Preventive measures to control air pollution will be discussed at the end of this section.

2.2.1 Air Pollutants

The study focuses on the following pollutants: Sulphur Dioxide(SO_2), Nitrogen Oxides (NOx), Carbon Monoxide (CO), Particulate Matter ($PM_{2.5}$), and Ozone (O_3). The data is monitored, managed and available at the Ontario Ministry of Environment (Ontario Ministry of Environment, 2010).

Sulphur Dioxide (SO_2)

Sulphur dioxide belongs to the family of sulphur oxide gases (SO_x). SO_2 reacts with water vapor in the air to form acids or acid rain which has been an issue of concern since the 1970's in many countries including Canada (Environment Canada, 2004). It is formed from the sulphur contained in raw materials such as coal, oil and metal-containing ores during combustion and refining processes. Developed countries are working towards removing the sulphur from motor fuels in the refining process in order to limit emissions. The sintering process used in metal smelting, which involves roasting metal sulfide ores in a stream of air can also be a major mechanism of sulphur dioxide production.

Previous studies suggest, exposure to sulphur dioxide tend to have short and long term effects on human health. Short term impacts include respiratory problems like shortness of breath or wheezing. It does not only impact individuals with asthma but normal young individuals have been impacted by this pollutant. Long term effects of sulphur dioxide include morbidity of lung function and in some cases mortality. Observational studies in Hong Kong suggest that reducing the sulphur dioxide levels will decrease the negative impacts on individuals and the environment (World Health Organization, 2005).

Carbon Monoxide (CO)

Carbon monoxide is a colorless, odorless and tasteless gas. It is of a major concern because, it is formed due to the incomplete combustion of carbon containing fuels. Carbon monoxide is primarily released from transportation sources such as cars, buses and off road vehicles. Other sources of carbon monoxide emissions include residential wood heating (wood stoves, fire places), wood and aluminum industries.

Exposure to carbon monoxide can have serious health impacts including respiratory diseases, visual perception, manual dexterity, learning functions etc. Since carbon monoxide gas is known to be poisonous gas long term exposure can lead to mortality (Environment Canada, 2004).

Nitrogen Oxides (NOx)

The process of combustion converts the nitrogen in fuels to oxides of nitrogen. Coal is the chief source of nitrogen whereas oil and gas contain far less amounts of nitrogen. At high temperatures nitrogen and oxygen in the atmosphere combine to form nitrogen oxides. Majority of nitrogen oxide formed through this process are released in the form of nitric oxide.

$$N_2 + O_2 \rightarrow 2\,NO$$

Nitric oxide is a colorless, odorless, and tasteless gas. It is produced naturally in the soil through biological nitrification and de-nitrification process. Major sources of nitric oxide emissions are gasoline/ diesel vehicles, electric generating stations, industrial boilers and home space heating. Previous studies also suggest that nitric oxide concentrations levels peak during the morning rush hour and rapidly deplete by oxidation to nitrogen dioxide (NO_2).

$$2\,NO + O_2 \rightarrow 2\,NO_2$$

Nitrogen dioxide is a colored gas that varies from yellow to brown depending on its concentration in the atmosphere. It has a pungent, irritating odor. Nitrogen dioxide is formed by the photochemical oxidation of nitric oxide; it peaks during the mid-morning hours in urban and suburban areas (Godish, 2004). It is considered as a secondary pollutant. When sunlight is absorbed by nitrogen dioxide it can be broken down to nitric oxide and oxygen atom. The oxygen atom then reacts with the oxygen molecule to form ozone (O_3). The process is evident during the day time hours and gradually fades out during the evening hours.

The impact of nitrogen dioxide on human health is of a great concern. It is released in the environment in the form of gas. Human health is compromised when this gas is inhaled from outdoor sources or indoor sources. Industrial manufacturing of nitric acids for explosives and welding are considered as outdoor sources, whereas tobacco smoking and the use of oil and gas fired source are considered as indoor sources.

According to air quality guideline global update 2005 a study conducted for *Airway Inflammation and Responsiveness*, results indicate that, exposure to lower concentrations (0.05 – 1.00 ppm) of nitrogen dioxide resulted in the antioxidant defences, uric acid and ascorbic acid being depleted in human bronchoalveolar lavage (BAL) fluid. The BAL fluid is a saline wash of the airways and air sacs for the recovery of inflammatory cells. Exposure to nitrogen dioxide concentrations more than 1.0ppm show changes in pulmonary functions. Individuals with asthma conditions are more susceptible to the impacts of nitrogen oxides. Furthermore a recent Canadian study conducted, examined the relationship between preterm birth and intrauterine growth retardation among singleton live births and it was determined that, nitrogen dioxide exposure during the last month of pregnancy was a factor in preterm birth and intrauterine growth retardation was a result of exposure to nitrogen dioxide in the first month of pregnancy.

Ozone (O_3)

Ozone is considered as a secondary pollutant because it is formed when two primary pollutants react in sunlight and stagnant air. The two primary pollutants resulting in the formation of ozone are nitrogen oxides (NOx) and volatile organic compounds (VOC's). Ozone is a colorless and highly irritating gas that forms just above the earth's surface (Environment Canada, 2004).

The main sources of ozone emissions are due to natural sources (coniferous forests) as well as human related activities (coal fired power plants, motor vehicles, residential wood combustion). Acute exposure to ozone has negative impacts on human health as well as vegetation. In humans it can impair the body ability to defend itself against infection. Other effects of ozone exposure are pulmonary function changes and respiratory responses which results in increased hospital admissions and emergency room visits. Severe impacts include increased asthma days and in some cases premature mortality. According to Environment Canada ozone not only affects the human health but also has an impact on the environment. It is known to decrease the productivity of some crops, injure flowers, shrubs and may contribute to forests decline in some parts of Canada.

Canada began giving serious consideration to the need to manage ground-level ozone in the mid 1980's. Additional ozone management agreements and programs continued to be established over the period from 1990 to date. The most recent major agreements addressing ground-level ozone are the Canada-wide Standards endorsed by Canadian Council of Ministers of the Environment (CCME) in 2000 (Environment Canada, 2004).

Particulate Matter (PM)

A Canada Wide Standard (CWS) for fine particulate matter PM (PM less than 2.5 microns in diameter) was developed and endorsed by the CCME in 2000. *(PM)* is generally known as airborne particles that are found in the form of solid and liquid. It includes aerosols, dust, fly ash, pollens, smoke and fumes. They vary in shapes, it can be found in simple spheres, crystalline cubes and even in irregular shapes. The shape of the PM is dependent on the chemical composition or the process by which it has formed. Size is an important characteristic for any particulate matter as different sizes impact humans differently. PM_{10} is (less than 10 microns in diameter) $PM_{2.5}$ is (less than 2.5 microns in diameter).

PM_{10} are inhalable particles that are small enough and can enter through the nasal passage and penetrate to the thoracic region. $PM_{2.5}$ consists of ultra fine particles that can even pass through the lungs and affect other organs. Studies have shown a link between PM and its effect on asthmatics. Other adverse health impacts include inflammation, respiratory tract infections, lung diseases etc. It has also been suggested that "chronic exposure to PM has a direct effect on the heart, potentially through uptake of particles into the blood or through the release of chemical components from PM into the circulation that affect either cardiac function or autonomic control of the cardiovascular system" (World Health Organization, 2005).

Particulate matter occurs in the atmosphere in two different ways. Primary particles are released in the atmosphere through natural (volcanoes, forest fires) and man-made sources (transportation, waste disposal, dust from roadways) etc. Secondary particles are a result of chemical reactions in the air involving other gases and aerosol particles. Ontario faces air pollution every year mostly during the summer season, due to secondary particles when people tend to travel more resulting in increased emissions from vehicles and use of air conditioning. This causes "Smog" it is sometimes visible in the form of yellow brown haze over the cities. Smog comprises of ground level ozone and particulate matter.

2.2.2 Monitoring air quality in Ontario

Air quality monitoring in Ontario is managed by the *Environmental monitoring and reporting branch* (EMRB). There are different instruments in place for continuous monitoring of each pollutant; this placement of instruments at different locations is known as the *Ambient network*. In 2007 Ambient network comprised of 139 continuous monitoring instruments at 40 different sites (Appendix G). These instruments are capable of recording data up to the minute. A list of instruments used to measure each pollutant is given below

Quality control (QC) and *Quality assurance* (QA) procedures are in place to determine data anomalies in order to make corrections in a timely manner. A telemetry system is used by the data analysts and station operators to verify the instrument precision. The monitoring instruments used across Ontario have been standardized to Thermo Electron Corporation Analyzers, this is done in order to streamline parts inventory and leverage common hardware used within each analyzer.

The air quality data available at the OMOE website is available by hour whereas it has been stated above that, the instruments collect air quality measurements every minute. Ambient air monitoring instruments have the capability to produce an average measurement for every hour. The data obtained from these instruments is stored in the ministry's *Air quality information systems* (AQUIS). Each pollutant has a predetermined concentration range based on the historical data; values outside the range are flagged for further investigation.

In this study hourly data is used in order to monitor the trends of different pollutants at given locations. This sort of data is generally helpful for the analysts in order to perform statistical and trend analysis and prepare reports accordingly. However, the general public needs to access information regarding air quality for their current location or for a different city in order to make travel plans or schedule prolonged outdoor activities. To attain this goal an indicator is used to determine the quality of air that can be best understood by the general public, this is known as the *Air quality index* (AQI).

Air quality index is based on real time pollutant measurements comprising of sulphur dioxide, ozone, nitrogen dioxide, total reduced sulphur compounds, carbon monoxide and fine particulate matter. At the end of each hour the pollutant concentration is converted to a number ranging from zero and upwards (Table 1). The pollutant with the highest number at a given hour becomes the AQI reading. The AQI reading changes as the quality of air or pollutant concentration is changed. The air quality is very good when the AQI readings are between 0-15 whereas readings above 100 determine very poor air quality (Table1).

Table 1: Adopted from Ontario ministry of environment, air quality index

AQI	DESCRIPTION	REPRESENTING COLOUR
0-15	Very Good	Blue
16-31	Good	Green
32-49	Moderate	Yellow
50-99	Poor	Orange
100+	Very Poor	Red

2.2.3 Policies and guidelines

The government of Ontario has several regulations and guidelines made under the *Environmental protection act* (EPA) in place to improve and prevent air pollution and its impacts. The regulation places limits on the total emissions from a facility. It requires that, "*where a facility emits a contaminant into the air from one or more sources, the concentration in the atmosphere resulting from that contaminant at any point of impingement (POI) must be less than the standard prescribed in the regulation"*[6]

Ontario MOE has a complete list of pollutants (approx 339) and the acceptable standards, generally known as the *Ambient air quality criteria* (AAQC). The scope of this study is to monitor the trends of few pollutants and not all. Table 2 below lists the acceptable standards of the pollutants used in this study.

Table 2: Acceptable standards of pollutants

CONTAMINANT NAME	AAQC $\mu g/m^3$ (microgram per cubic meter)	AVERAGING TIME	LIMITING EFFECT
Carbon Monoxide	36,200 (30ppm)	1 Hour	Health
	15,700 (13ppm)	8 Hour	Health
Nitrogen Dioxide	200 (0.10ppm)	24 Hour	Health
	400 (0.20ppm)	1 Hour	Health
Ozone	165 (0.080ppm)	1 Hour	Health
Particulate Matter	30	24 Hour	
Sulphur Dioxide	690 (0.25ppm)	1 Hour	Health &Vegetation
	275 (0.10ppm)	24 Hour	Health & Vegetation

2.2.4 Modelling the spatial distribution of pollution

Air pollution science does not deal with the effects of emissions from individual sources but with the aggregated effects. In order to better understand the processes behind different kinds of pollution and to be able to model the spatial distribution of pollution the sources have been categorized into three different categories.

- Point sources i.e. power stations, industrial sites etc.
- Line sources i.e. railway lines, road networks etc.
- Area sources, emission sources in this category are not major contributors to air pollution but when combined together do have an impact on the quality of air, for example boilers are used by most of the homes for heating purposes but individual boiler is not as harmful compared to a grid of residential area with boilers (World Health Organization, 2005).

The ministry of environment recommends the use of air dispersion models for the sources of emissions i.e. (power plants, chemical factories, industries etc). OMOE has three major approved air dispersion models. They are as follows

- AERMOD: used for more sophisticated modeling analyses
- ISC PRIME: used to assess potential for contamination of building air intakes
- SCREEN3: used for screening analyses.

Alternative models are used for site specific consideration, if conditions warrant their use. Approval from the MOE is required before the use of alternative models. Alternative models include

- CALPUFF: used for modeling long range transport up to 300km (land-sea or lake).
- CAL3QHCR: used for roadway's and modeling of intersections.
- SDM: (Shoreline Dispersion Model) used for emission sources located on the shorelines.
- CALINE – 4: used for roadways including intersection.

The use of the above mentioned models is the practice to determine and maintain the emission concentrations according to the ambient air quality criteria. However there may be other climatic factors that determine the spatial distribution of pollution. For instance wind speed and wind direction can have a significant impact on the quality of air not only around the emission site but several kilometers away from it. In 2000 the Ministry of Environment conducted a Clarkson Airshed Study (CAS) to address the concerns of the general public and stakeholders regarding the air quality in that area. In this study wind speed and wind direction were taken into consideration and correlated with pollution concentration. This was done using Wind Rose and Pollution Rose. Wind Rose is a graphical tool showing the distribution of winds from each compass direction at any given location. The long bars illustrate the origin of wind direction and different colors of the bars illustrate the wind speed. Pollution Rose is similar to Wind Rose except that it also determines the average pollutant concentration for each specific wind sector. It was concluded in this study that majority of air borne $PM_{2.5}$ measured in Southern Ontario is of transboundary origin. More than 50 % of fine particulate matter in Ontario originated from the U.S. (Ontario Ministry of Environment, 2006)

2.2.5 Pollution prevention

In Canada the provincial governments are primarily responsible for the prevention of air pollution. The *Canadian environmental protection act* (CEPA) is the principal act for the regulation of environmental contaminants. It was first passed as a law in 1988. Its focus is pollution prevention and the protection of the environment and human health in order to contribute to sustainable development (Health Canada, 2008).

Pollution prevention is most effective when governments, industries and individuals are taking responsible actions in reducing air emissions. Some of the methods to improve air quality are discussed below.

Cleaner Fuels

The government of Canada has diverted its focus from plastics since the increasing oil prices and environmental concerns it is spending millions of dollars in research funding to find alternatives and create green fuels and green chemicals. Avoiding chemical products, aerosols, oil based paints, sprays and using less toxic alternatives and water based paints are better options (Ontario ministry of environment, 2008).

Drive Clean

Regular maintenance of the car is essential; a well maintained and well tuned car pollutes less. Drive Clean is Ontario's mandatory vehicle emissions inspection and maintenance program. All owners of the vehicles have to comply and undergo an emission test in order to identify any emissions problems and have them repaired. The Drive Clean program is a way to ensure and safeguard our health and protect our environment (Ontario ministry of environment, 2008).

Wood Burning

Since the late 1970's up until few years ago wood stoves and furnaces were popular methods of warming residential places. Wood burning appliances compromises the indoor air quality, it is known that the use of these appliance emit carbon monoxide, nitrogen oxides, sulphur dioxide and various volatile organic compounds.

An alternative to wood fireplace, the gas fireplaces are a new trend across North America. They are inexpensive, energy efficient and maintenance free. They are considered as comparatively less carbon monoxide and particulate emitters. However no energy source is completely environmentally friendly, but proper installation and usage will result in less energy consumption and less emissions (Ontario ministry of environment, 2008).

Electricity Consumption

The province of Ontario has wide variations of summer and winter temperatures. The far north has coolest temperatures whereas the south east and west are the warmest during the summer months. These varying temperatures all year round result in increased usage of air conditioning or heating. The Ontario ministry of environment recommends that increasing the temperature by a few degrees will help improve air quality (Ontario ministry of environment, 2008).

Alternative Fuels

The office of Energy Efficiency (OEE) in Canada is the centre of excellence for energy conservation, energy efficiency and alternative fuels information. It is playing an important leadership role in helping Canadians reduce energy costs and making the environment healthier. The OEE offers grants, incentives and awards if the eco- Energy programs are availed by the general public. Below are some of the Office of Energy Efficiency programs;

- eco-Energy for Biofuels: Production of renewable alternative sources to gasoline & diesel

- eco-Energy for Retrofit: provides grants to homeowners, businesses and industries to invest in energy and pollution saving upgrades.
- eco-Energy for Buildings and Houses: Installation of energy efficient equipments and products.
- eco-Energy Efficiency and Regulations: energy efficiency regulation in place for energy efficient equipments and products.
- eco-Energy for Industry: develop and use methods and industrial processes that are more energy efficient.
- eco-Energy for Personal Vehicles: provides Canadians with tips on buying, driving and maintaining energy efficient vehicles (Ontario ministry of environment, 2008).

3. METHODS

3.1 Study area

Canada is the second largest country by total area in the world. It extends approximately from latitude 42′ 16’ N to 74′ 43’ and longitude 52′ 45’ W to 135′ 4’ W (figure5). It has ten provinces (Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Ontario, Prince Edward Island, Quebec, Saskatchewan) and 3 territories (Northwest Territories, Nunavut, Yukon Territory). The capital city of Canada is Ottawa.

Canada resembles the US in its market-oriented economic system, pattern of production, and affluent living standards. Canada is the US's largest foreign supplier of energy, including oil, gas, uranium, and electric power. During 2007, Canada enjoyed good economic growth, moderate inflation, and the lowest unemployment rate in more than three decades (The world fact book, 2007).

Canada exports commodities such as wood pulp, crude petroleum, natural gas, auto motor vehicles and its parts, chemicals, plastics etc. Canada is also an importer majority of import goods approximately 54% come from U.S and some from China and Mexico (The world fact book, 2007).

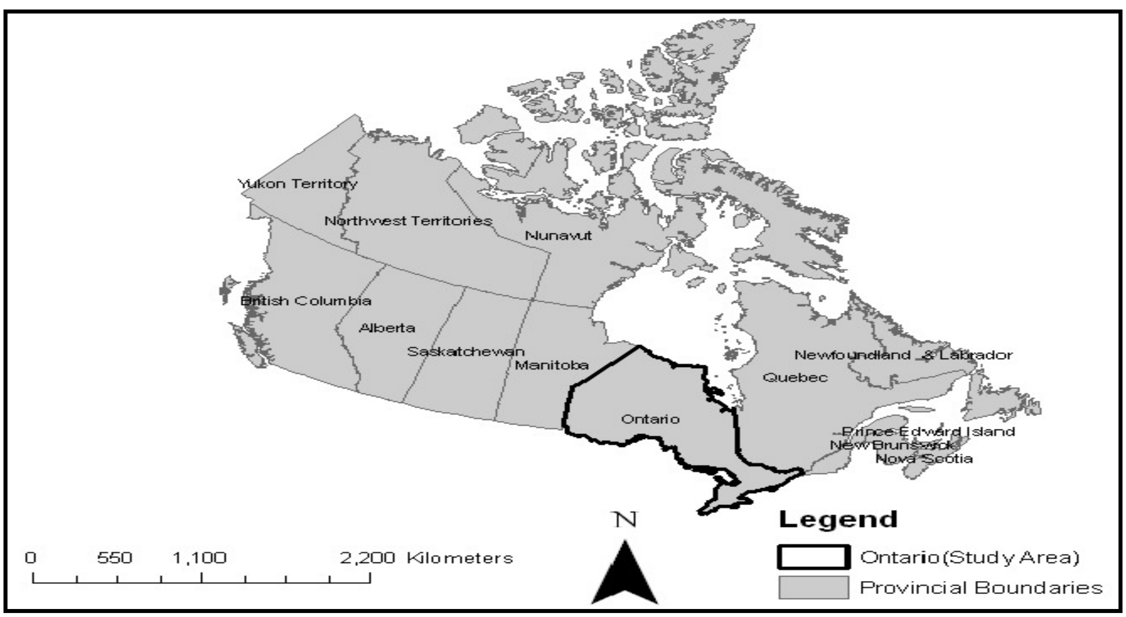


Figure 5: Canada, ten provinces and three territories (Yukon, Northwest, and Nunavut.

The study area focuses on the cities in the province Ontario, which is the second largest province by area in Canada. It extends approximately from latitude 42′ 16’ N to 49′ 48’N and longitude 74′ 45’ W to 94′ 22’ W (figure 5). It is one of the most populated provinces in the Canada. The province of Manitoba is located on the west and Province of Quebec on the east side. Ontario shares its borders with the Unites States through the Great Lakes including (Lake Superior, Lake Erie, Lake Huron and Lake Ontario).

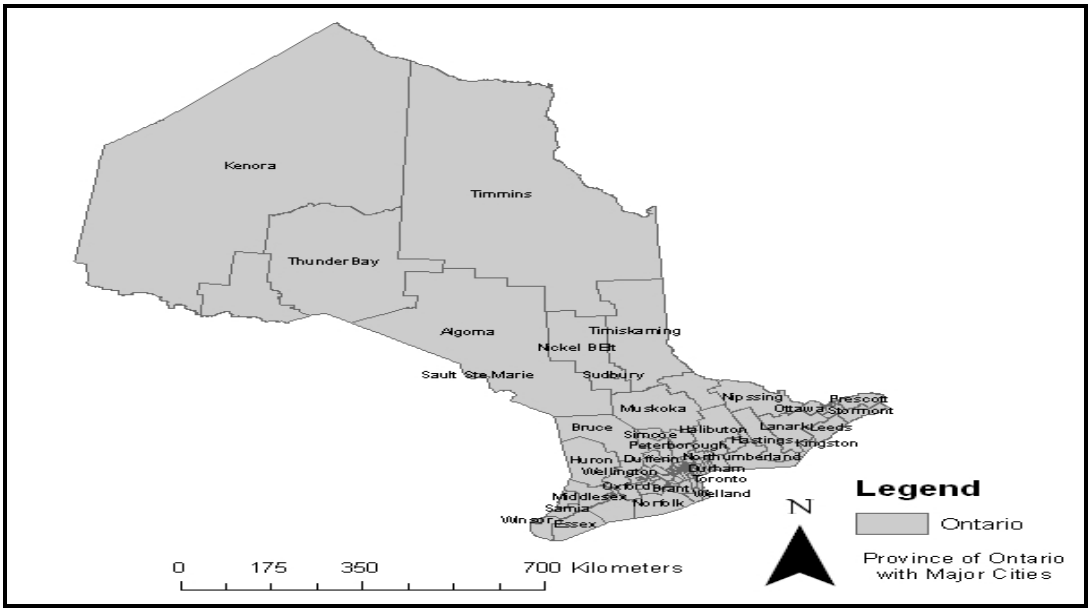


Figure 6: Province of Ontario with major cities.

<u>Climate</u>

The province of Ontario (figure 6) is divided into three regions Southwestern Ontario, Southern Ontario and Northern Ontario.

Southwestern Ontario includes cities of Windsor, London, Hamilton and Toronto. It has a moderate humid climate where summers are hot and humid and winters are cold. Snow storms are more frequent in the months of January and February. Spring and fall temperatures are pleasant. Annual precipitation ranges approximately 75 – 100 cm.

Southern Ontario including cities of Ottawa, Sudbury and North Bay have relatively shorter summer seasons compared to southwestern Ontario. The region has rough, cold long winters because of Lake Huron and Lake Superior on its east. Some of the snow belt areas of this region receive heavy snow approximately 300 cm.

Northern Ontario with no major cities has, frigid cold winters and temperatures around -40°C are very common. The summer season is short with cool to warm temperatures and low humidity levels. The region is covered with snow during the months of October – May (Weather Network, 2009).

<u>Population</u>

The population of Ontario has been on the rise since 1991. It is Canada's only province which has seen a growth rate increase of 6.6% that is above the national growth rate increase of 5.4%. According to Citizenship and Immigration Canada, international immigration is the main reason for Ontario's growing population trend since 1991. During the year 2001 – 2006 approximately

600,000 immigrants settled in this province. According to the 2006 census the population of Ontario was accounted for 12,160,282 inhabitants that is two fifths or 38.5% of Canada's population.

Table 3 below illustrates the populations for the provinces of Canada during 1956 – 2006. A span of 50 years, where it is fairly evident that there has been a growing trend in population in all provinces but most of the increase can be seen in the province of Ontario (Statistics Canada, 2009).

Table 3: Population of Canada, provinces in the last 50 years:

Population of Canada, provinces and territories in the last 50 years														
Year of census	Canada	N.L.	P.E.I.	N.S.	N.B.	Que.	Ont.	Man.	Sask.	Alta.	B.C.	Y.T.	N.W.T.	Nvt.
	in thousands													
Sources: Statistics Canada, censuses of population, 1956 to 2006														
1956	16,081	415	99	695	555	4,628	5,405	850	881	1,123	1,398	12	19	...
1961	18,238	458	105	737	598	5,259	6,236	922	925	1,332	1,629	15	23	...
1966	20,015	493	109	756	617	5,781	6,961	963	955	1,463	1,874	14	29	...
1971	21,568	522	112	789	635	6,028	7,703	988	926	1,628	2,185	18	35	...
1976	22,993	558	118	829	677	6,234	8,264	1,022	921	1,838	2,467	22	43	...
1981	24,343	568	123	847	696	6,438	8,625	1,026	968	2,238	2,744	23	46	...
1986	25,309	568	127	873	709	6,532	9,102	1,063	1,010	2,366	2,883	24	52	...
1991	27,297	568	130	900	724	6,896	10,085	1,092	989	2,546	3,282	28	36	21
1996	28,847	552	135	909	738	7,139	10,754	1,114	990	2,697	3,725	31	40	25
2001	30,007	513	135	908	729	7,237	11,410	1,120	979	2,975	3,908	29	37	27
2006	31,613	505	136	913	730	7,546	12,160	1,148	968	3,290	4,113	30	41	29

Source: Statistics Canada, N.L (Newfoundland and Labrador), P.E.I (Prince Edward Island), N.S (Nova Scotia), N.B (New Brunswick), Que (Quebec), Ont (Ontario), Man (Manitoba), Sask (Saskatchewan), Alta (Alberta), B.C (British Columbia), Y.T. (Yukon Territory), N.W.T. (Northwest Territories), Nvt.(Nunavut). (Column in Pink, Province of Ontario Population).

3.2 Implementation

3.2.1 Data staging area

In this study all the data originates from the Environment Canada website. First a form based query is executed on the website (figure 3) to produce the data from a year on a particular location and for a particular pollutant. This data (figure 4) is copied from the html file to an excel spreadsheet. Then, it is copied from the spreadsheet to the text file. Unfortunately, this two step process could not be avoided since copying directly from the html page to the text file creates some undesirable characters in the text file. The text files are saved with a particular naming convention to identify the contents of the file. This is important for the *Extract transfer load* (ETL) program to read these files. The ETL program is a very small utility program written in Java. It reads the file in a particular folder specified in the program. It parses the contents of the files and then uses the *Java database connectivity* (JDBC) to insert the data into the tables as shown in the following code snippet.

```
String sql = "insert into tblpollution values (" + id + ", '" + station + "', '" + pollutant + "', " + hr + ")";
```

Here the variable *id* is created by concatenating the year, month, day and hour of the pollution measurement. This is obviously not unique for each record and it is also used to relate the pollution measurement records with the time dimension. *station* and *pollutant* are obvious while *hr* is the level of the pollution.

Once the data has been processed, removing all missing and invalid values as this could cause a problem when performing aggregation functions such as roll up. The clean data is then uploaded in MySQL database.

<u>MySQL Database</u>

MySQL is an open source database developed by a Swedish company MySQL AB (MySQL, 2009). It has a large number of installations due to the free availability, stability of the code and ease of use. This project uses the latest version of the MySQL community server. The installation process is very straightforward. The setup program downloaded from mysql.com not only creates the directory structure for the programs and data files but it also creates a windows service. This service basically allows starting and stopping the database. MySQL also comes with a few graphical tools for database management.

<u>Database Design</u>

Database design is arguably the most critical phase in this project. Accurate database design is the basis for the final product to meet the user and systems requirements. During the designing process two models were taken into consideration, logical star schema where, all the objects are mapped and relationships are defined. The second is the database schema where, data types and how the data will be stored are defined. Each of these two models is described in detail below.

Logical Model

In this study the logical model has been defined using *Dimensional modeling* (DM) approach. A dimensional model consists of one table with multi part key or composite key known as the fact table and a set of minor tables known as the dimension tables. Each dimension table has a primary key that references to the multi part key in the fact table. These characteristics of a data model form a shape of a star and are known as star schemas.

Star Schema

A star schema is data modeling technique used to model multi dimensional data. It consists of facts, dimensions, attributes and attribute hierarchies. In this study the table pollution is a fact table. The dimensions are table location, table time and table pollutants. The attributes and attributes hierarchies for table time are year, month, day and hour. The attribute hierarchies for table location are country, province, city.

When using multi dimensional approach it is important to store the attributes in the dimension table with an "id" (station_id, time_id, pollutant_id) which helps in identifying every slice in a cube. This is a highlighting feature in an OLAP tool where aggregation functions are applied on the data. The data can be aggregated at a higher level using the "roll up" function or at a finer level using "drill down" function. It is also possible to determine single fact value of a given dimension using the "slice and dice" function.

The logical star schema of air quality data mat for Ontario Canada is given below.

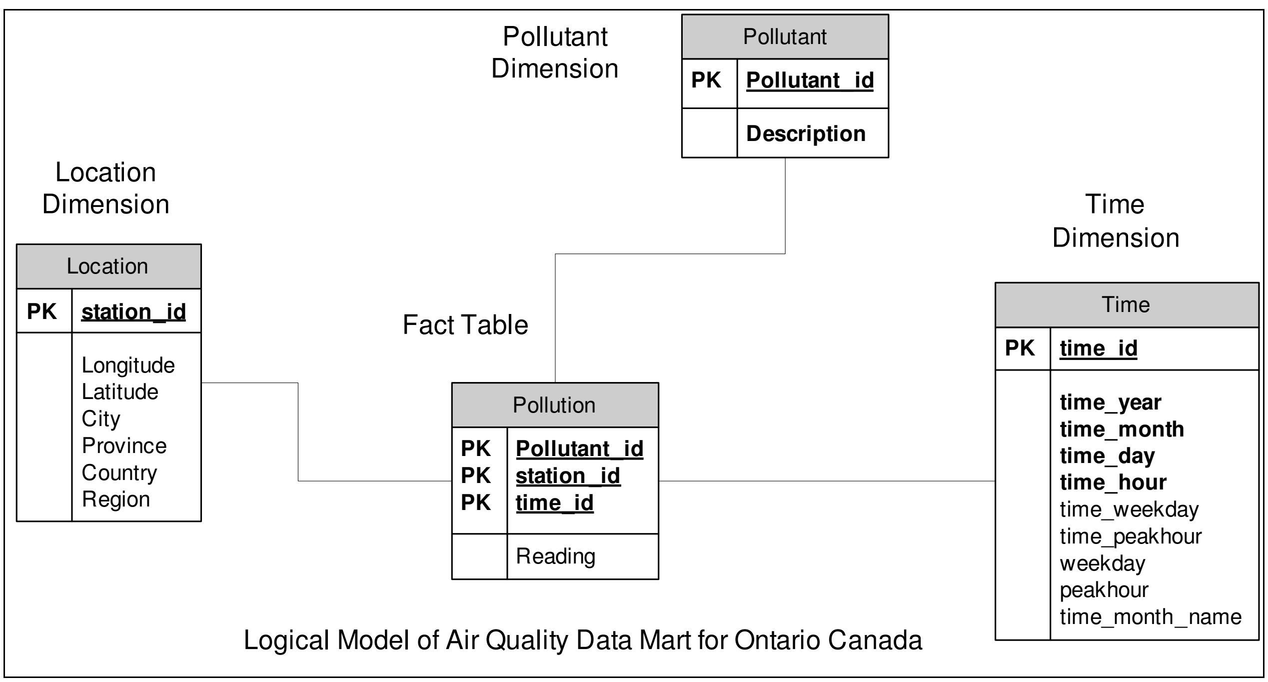


Figure 7: Logical Star Schema of air quality data mart for Ontario Canada.

Database Schema

A database schema deals with the finer details of a database structure. It defines the structure of the tables, what data types will be used, the length of entity in an attribute whether the entity accepts null values or if it is set as a primary or foreign key etc. In air quality data mart four tables have been defined *location, pollutant, pollution and time*. The structure of the tables is given in Table 4

Table 4: Database schema of air quality data mart for Ontario Canada

Table/Column Name	Data Type	Not Null	Primary Key	Comments
Tbl-Location				
Station_id	Varchar (6)	**x**	**x**	Station Number in a city
Longitude	Varchar (20)			
Latitude	Varchar (20)			
City	Varchar (30)			Name of the city where station is located
Province	Varchar (15)			Accepts Initials for Province
Country	Varchar (20)			Currently one value added "Canada"
Region	Varchar (20)			Can aggregate countries based on regions
Tbl Pollutant				
Pollutant_id	Varchar (5)	**x**	**x**	Values in chemical formula
Description	Varchar (45)	**x**		Full name of the pollutant
Tbl Pollution				
Time_id	Integer	**x**	**x**	Takes concatenated values in year, month, day, hour format i.e.(200404294)
Station_id	Varchar (6)	**x**	**x**	Values allocated by the OMOE according to geographic distribution of cities.
Pollutant_id	Varchar (5)	**x**	**x**	Accepts values in chemical formula
Reading	Double			Pollution measurements in parts per billion (ppb)

Tbl-time				
Time_id	Integer	**x**	**x**	Takes concatenated values in year, month, day, hour format i.e.(200404294)
Time_year	Integer	**x**		Accepts 4 digit values i.e. 2004
Time_month	Integer	**x**		Takes values 1 – 12
Time_day	Integer	**x**		Takes values 1 – 31
Time_hour	Integer	**x**		
Time_weekday	Varchar (45)			Takes values as “weekend or working day”
Time_peakhour	Varchar (5)			Takes values as “High or Low”
Weekday	Integer			Accepts values as “0” for weekend and “1” for working day
Peakhour	Integer			Accepts values as “0” for non peak hour and “1” for peak hour
Time_month_name	Varchar(45)			Takes values in short form Jan – Dec

3.2.2 Presentation Server

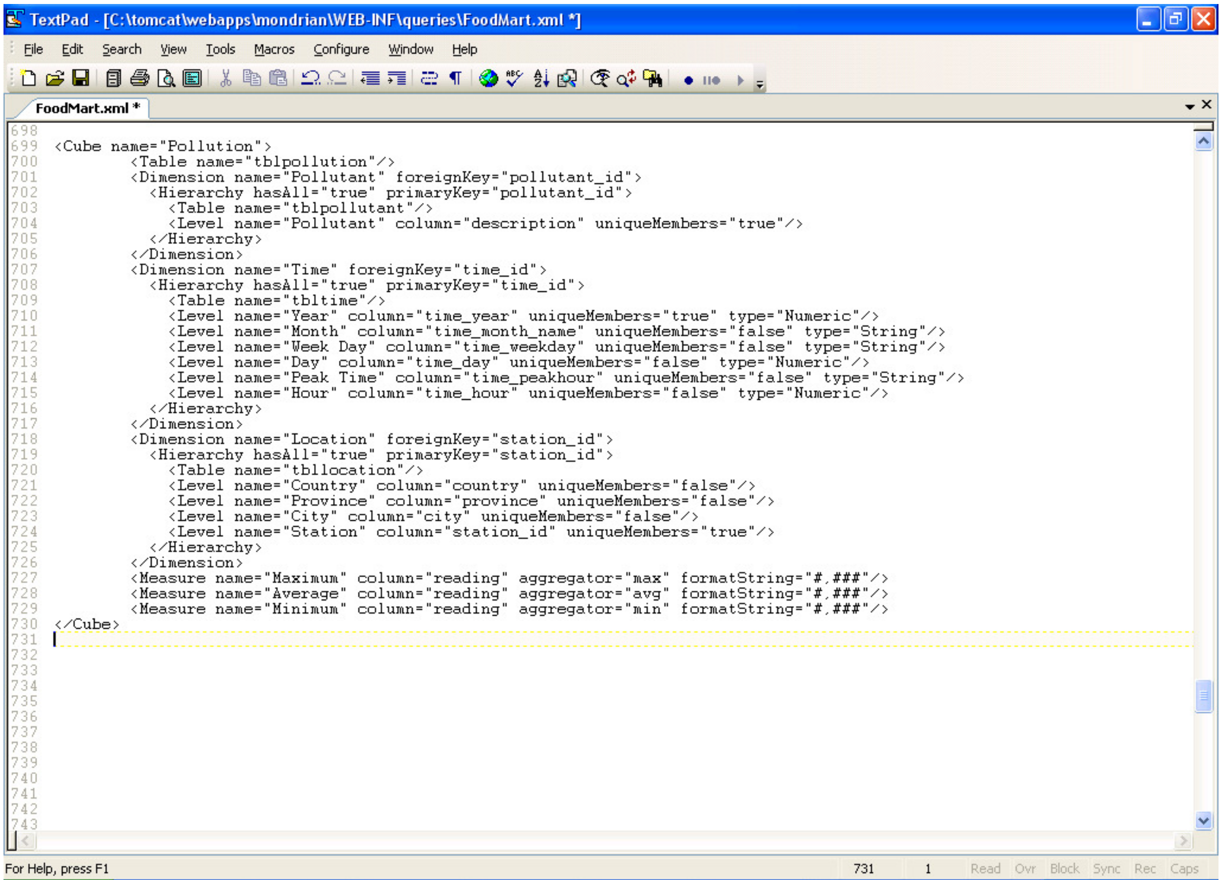
This section describes Java Application Server and OLAP server.

Application Server

Tomcat is a popular open source application server (Apache Tomcat, 2008). It is based on the Java framework for enterprise applications called J2EE. It is downloadable in a number of formats for different platforms. The project requires this application server because the OLAP server and client are J2EE based web applications and they need to be hosted on a J2EE complaint application server. Tomcat was a natural choice due to free availability and ease of use. The installation package creates a windows service which could be used to start and stop the application server. Tomcat installation requires that a Java Runtime Environment JRE is already available on the machine. Fortunately, latest windows installations already have java pre-installed.

<u>OLAP Server</u>

Mondrian is an open source OLAP server. It has gained a lot of attention recently. The stability of the code has encouraged a number of people to build applications based on Mondrian. Mondrian is a J2EE based web application so it requires an application server to deploy. The cube pollution is defined using *multidimensional expressions* (MDX) which is a dialect of *extensive markup language* (XML)




```xml
<Cube name="Pollution">
        <Table name="tblpollution"/>
        <Dimension name="Pollutant" foreignKey="pollutant_id">
          <Hierarchy hasAll="true" primaryKey="pollutant_id">
            <Table name="tblpollutant"/>
            <Level name="Pollutant" column="description" uniqueMembers="true"/>
          </Hierarchy>
        </Dimension>
        <Dimension name="Time" foreignKey="time_id">
          <Hierarchy hasAll="true" primaryKey="time_id">
            <Table name="tbltime"/>
            <Level name="Year" column="time_year" uniqueMembers="true" type="Numeric"/>
            <Level name="Month" column="time_month_name" uniqueMembers="false" type="String"/>
            <Level name="Week Day" column="time_weekday" uniqueMembers="false" type="String"/>
            <Level name="Day" column="time_day" uniqueMembers="false" type="Numeric"/>
            <Level name="Peak Time" column="time_peakhour" uniqueMembers="false" type="String"/>
            <Level name="Hour" column="time_hour" uniqueMembers="false" type="Numeric"/>
          </Hierarchy>
        </Dimension>
        <Dimension name="Location" foreignKey="station_id">
          <Hierarchy hasAll="true" primaryKey="station_id">
            <Table name="tbllocation"/>
            <Level name="Country" column="country" uniqueMembers="false"/>
            <Level name="Province" column="province" uniqueMembers="false"/>
            <Level name="City" column="city" uniqueMembers="false"/>
            <Level name="Station" column="station_id" uniqueMembers="true"/>
          </Hierarchy>
        </Dimension>
        <Measure name="Maximum" column="reading" aggregator="max" formatString="#,###"/>
        <Measure name="Average" column="reading" aggregator="avg" formatString="#,###"/>
        <Measure name="Minimum" column="reading" aggregator="min" formatString="#,###"/>
</Cube>
```

Figure 8: Cube definition for air quality data mart.

In this project the data cube is three dimensional. The data cubes can grow to n- number of dimensions and become hyper cubes. It is important to determine the dimensions of the data cube at beginning of the process because addition of an extra dimension means recreating the data cube entirely and the process can be very time consuming.

The programming code snippet has been shown in (figure 8) used in this project. The cube, table and dimension names are defined in tags with quotation marks. The primary and foreign keys are defined in the dimension tag. Every dimension, for example dimension ‘time’ has hierarchies such as year, month, day, hour etc. Each hierarchy also has its column name and data type defined as well.

Before closing the cube tag the aggregation functions are defined for the fact table in this case, average, minimum and maximum are defined for the pollutant measurements along with the data types.

The main feature of any data mart is to store dimensional data. This can be very helpful for the analysts to interpret data in many different ways. In air quality data mart the data is three dimensional.

***Location*:** Several air monitoring stations are located across the province of Ontario. Some cities have multiple air monitoring stations and most have only one per city.

Time: When analyzing data time is a crucial entity, to determine or investigate the events that occurred in a particular time frame. In air quality data mart the data can be found for the years 2000 – 2007. The data is not only available by year but it can be found at lower level of aggregation by months, days and by hour.

Pollutants: There are several pollutants regulated by the OMOE, but in this data mart the following pollutants (SO_2, NO, NO_x, NO_2, O_3, CO, $PM_{2.5}$) can be found. To monitor the trends over the number of years or comparative analysis between different studies, it is very important to have pollutant measurements. These measurements in air quality data mart are available for different pollutants for most cities and for most of the years. The reasons for any unavailable data are either the data collected was invalid or missing or the air quality monitoring did not start in those years.

3.2.3 End user tool

Mondrian also comes with a built-in OLAP client with the foodmart example. This client is based on another open source project jpivot. Initial experimentation with the built-in client OLAP revealed that this sample client application is not stable and it will not be able to handle large dataset as required by this project. That is why Openi was chosen as an OLAP client for this project. The end user will be able to use the OLAP client Openi (pronounced as open eye) which is an open source web application for OLAP reporting to perform analytical operations. The results are generated in the form of table, chart or reports.

The developmental phase of air quality data mart, reducing the complexity of creating and publishing reports was one of the main priorities of this project. It is intended to serve as a user friendly interface as well as provide models that can be used for data analysis.

3.3 CASE STUDY

To evaluate the working functionality of the tool a case study was performed for some of the major cities of Ontario. The study also tries to answer the differences in air quality between the urban and rural areas of Ontario. Most of the comparisons are done for the city of Toronto and Ottawa. The results are presented from the air quality data mart as well as from the *Canadian environmental sustainability indicator* (CESI) in order to verify the results achieved. The CESI indicator is used for sub objectives (a, b, e and f) sub objectives (b and c) are completely performed using air quality data mart.

To determine the concentration levels of SO_2, CO, NO_2, O_3 and $PM_{2.5}$ for the cities of Toronto, Ottawa, Hamilton, London and Windsor, the CESI air quality indicator approximates the population exposure to these pollutants. The indicator measures the maximum 24-hour average for SO_2, NO_2 and $PM_{2.5}$.. 8-hour maximum average for CO and1-hour maximum average for O_3. The calculation considers successive 1 hour, 8 hour, or 24 hour periods depending on the pollutants and their averaging times as shown in Table 4 and computes the average concentration ending on each hour. The highest of these averages is used for that day for a given station. The annual maximum average concentrations are calculated similarly, the maximum value during the year will be the representing value for a given year, station and pollutant.

Ground level ozone and fine particulate matter $PM_{2.5}$ are the key components of smog; it is a mixture of pollutants with ground level ozone as a main component. Smog is most apparent during the summer season with maximum ozone concentrations during the mid afternoon hours. The stationary air can trap pollutants and when in contact with sunlight create unacceptable levels of ground level ozone. Usually, ozone layer protects the earth from harmful ultraviolet rays but at ground level, ozone can be a harmful and a health hazard pollutant. In this study most of the comparisons and analysis revolve around these two pollutants.

The year 2005 was selected because all major cities started the monitoring of most pollutants by that time. All of the cities selected above had the selected pollutants and their readings for the year 2005. The urban and rural areas of Ontario were determined based on their population according to the most recent census of 2006. The peak hours of the day are assumed to be between 7 am – 9 am and 4 pm – 7 pm, rest of the hours is considered as non peak hours. This assumption is adopted from the traffic management practices in Canadian cities.

To use the CESI the data was organized in MySQL query browser. A “view” was created using *table pollution* where values for station_id, pollutant_id, and time_id can be extracted.

```
create view testtable as select * from tblpollution p where p.station_id = '31103' and p.pollutant_id = 'CO'
and p.time_id like '2007%'
```

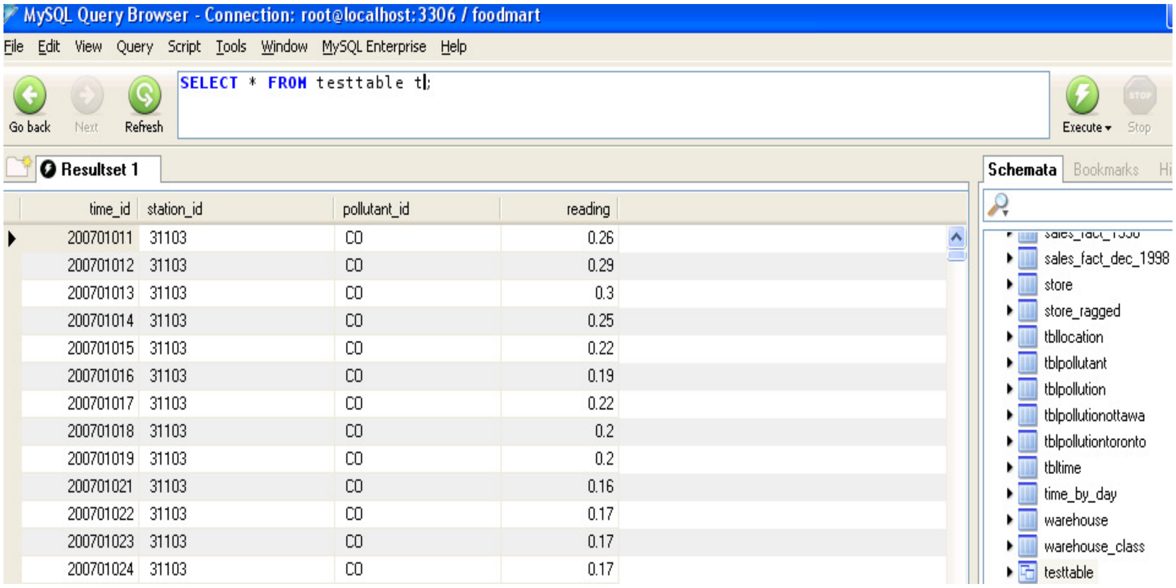


Figure 9: Query Result from the View in MySQL Query Browser.

Using the following query the data shown in figure 9 was further joined on *table time* to order the data by year, month, day and hour. The resultant data is shown in figure 10.

Select t.time_year, t.time_month, t.time_day, t.time_hour, p.reading from tbltime t left join testtable p on (t.time_id = p.time_id) where t.time_year = 2007 order by t.time_year, t.time_month, t.time_day, t.time_hour

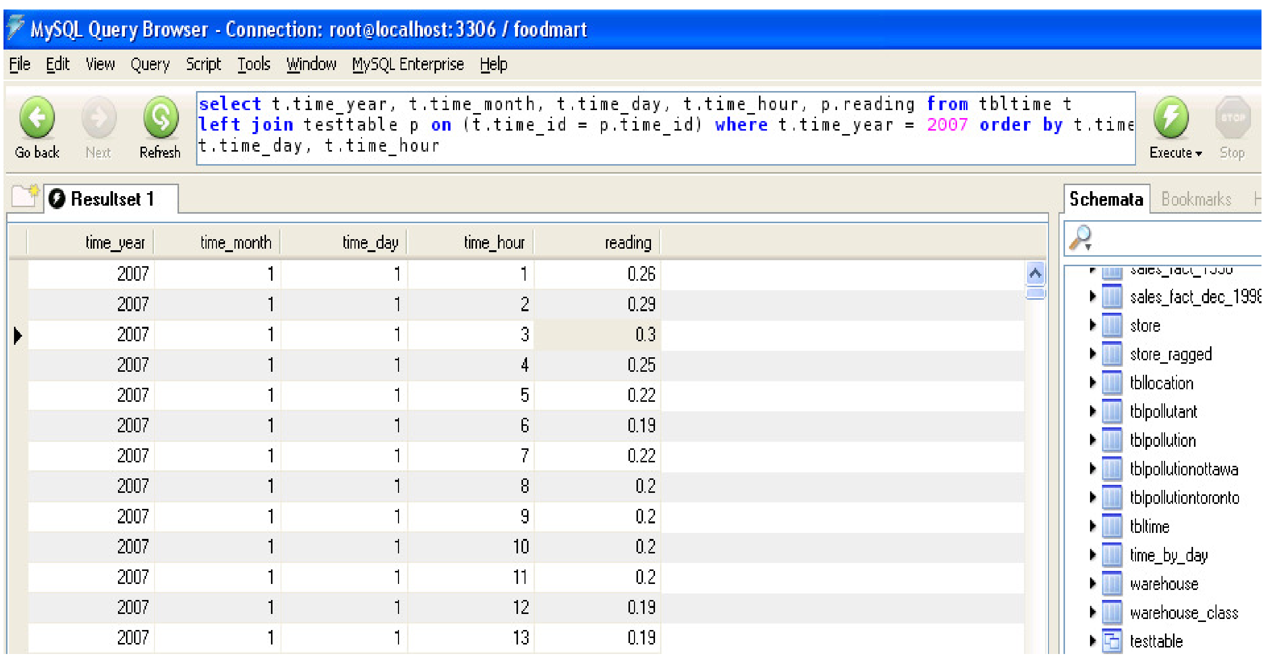


Figure 10: View joined with table time and ordered by year, month, day and hour.

The result set is then exported to MS Excel where 1- hour, 8- hour or 24- hour annual maximum averages are computed. This is done because air quality data mart has *Measures* defined as maximum, minimum and average and not as moving averages as shown in figure 8. The disadvantage of using moving averages rather than mathematical average is that the values are smoothed and graph created does not show sharp peak or low values.

A comparison of results will be provided for the results achieved using air quality data mart and the results achieved using the CESI. It is important to monitor the maximum pollution readings for the day, month or year in order to assess the true facts or readings when the pollution levels were the highest.

4. RESULTS

4.1 Development and implementation of the air quality data mart

Development and implementation was done using the air quality data mart architecture as described in the methods section (3.2.1) a few screen shots from “Air quality Data Mart for Ontario Canada” are presented below.

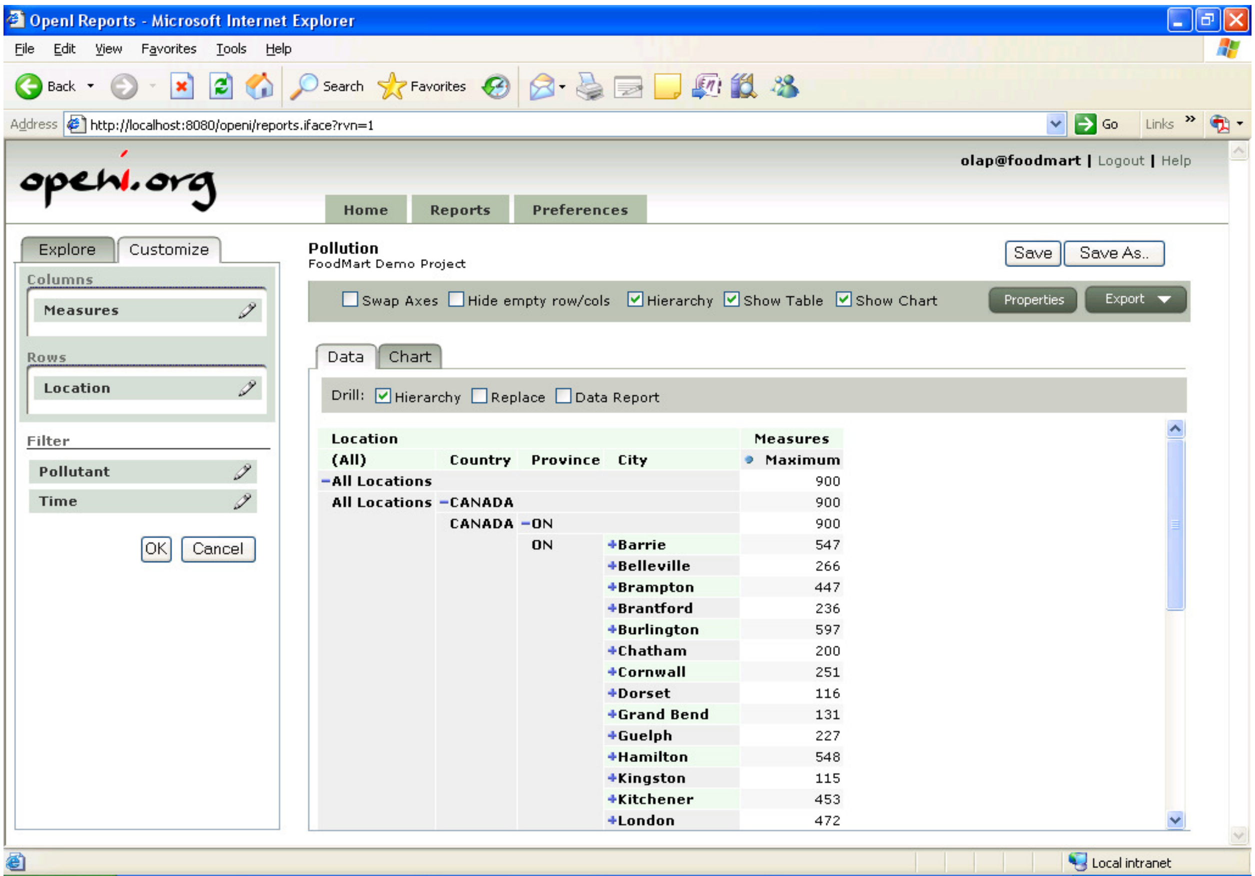


Figure 11: Location with aggregated pollution values.

In figure 11 on the left hand side the columns tab contains the fact table, dimension table is under the rows tab and filter tab contains additional dimension tables where the data can be filtered according to the user’s need. The rows and the filter tabs can be changed by drag and drop feature.

On the right hand of figure 11 the location values from the dimension table and maximum values from the fact table can be seen. These are aggregated values for each location.

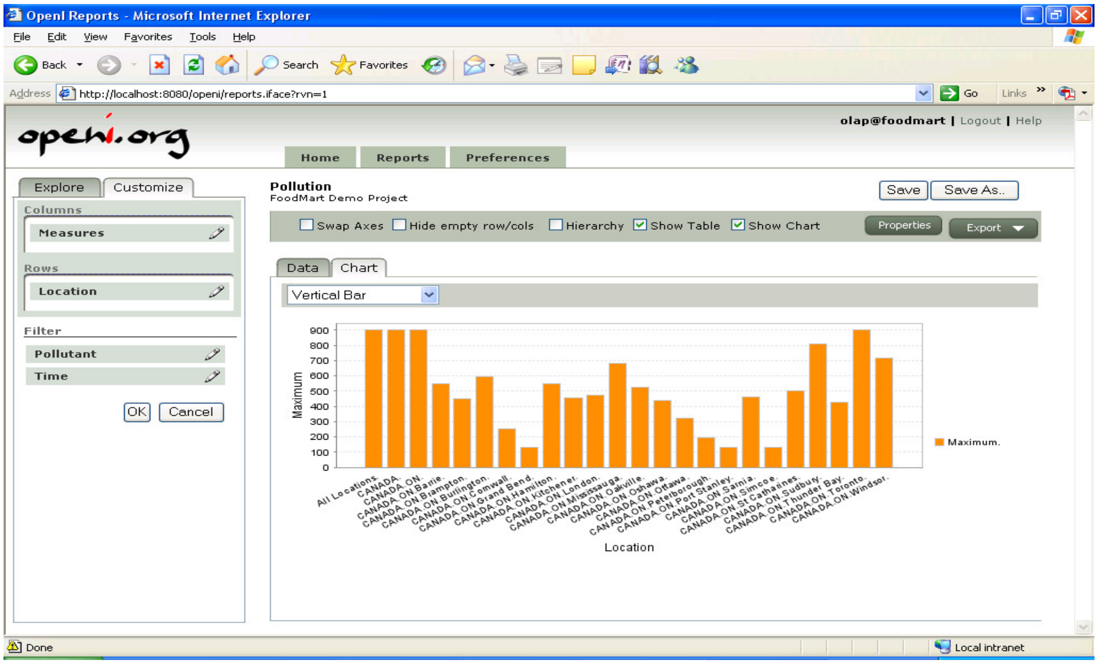


Figure 12: Chart of location with aggregated pollution values.

In figure 12 the same table in figure 11 can be illustrated in the form of chart, with location on the x-axis and measurements on the y-axis. On the right side of the screen there are two tabs the data and chart. Under the chart tab there is a drop down list where the style of the chart can be selected.

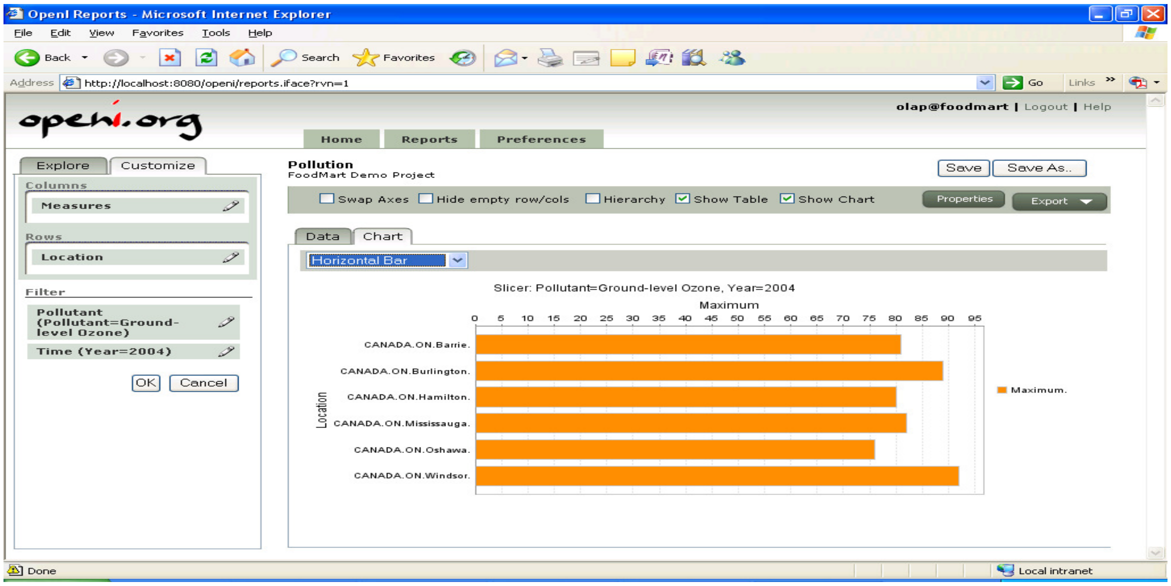


Figure 13: Data filtered on pollutant and time dimension.

In figure 13 the data has been filtered on pollutant “Ground level Ozone”. The filtering process has been applied for few cities in Ontario for the year 2004. This is done by clicking on the pen icon and selecting the desired location, time frame, pollutant and measurement (average, minimum, maximum)

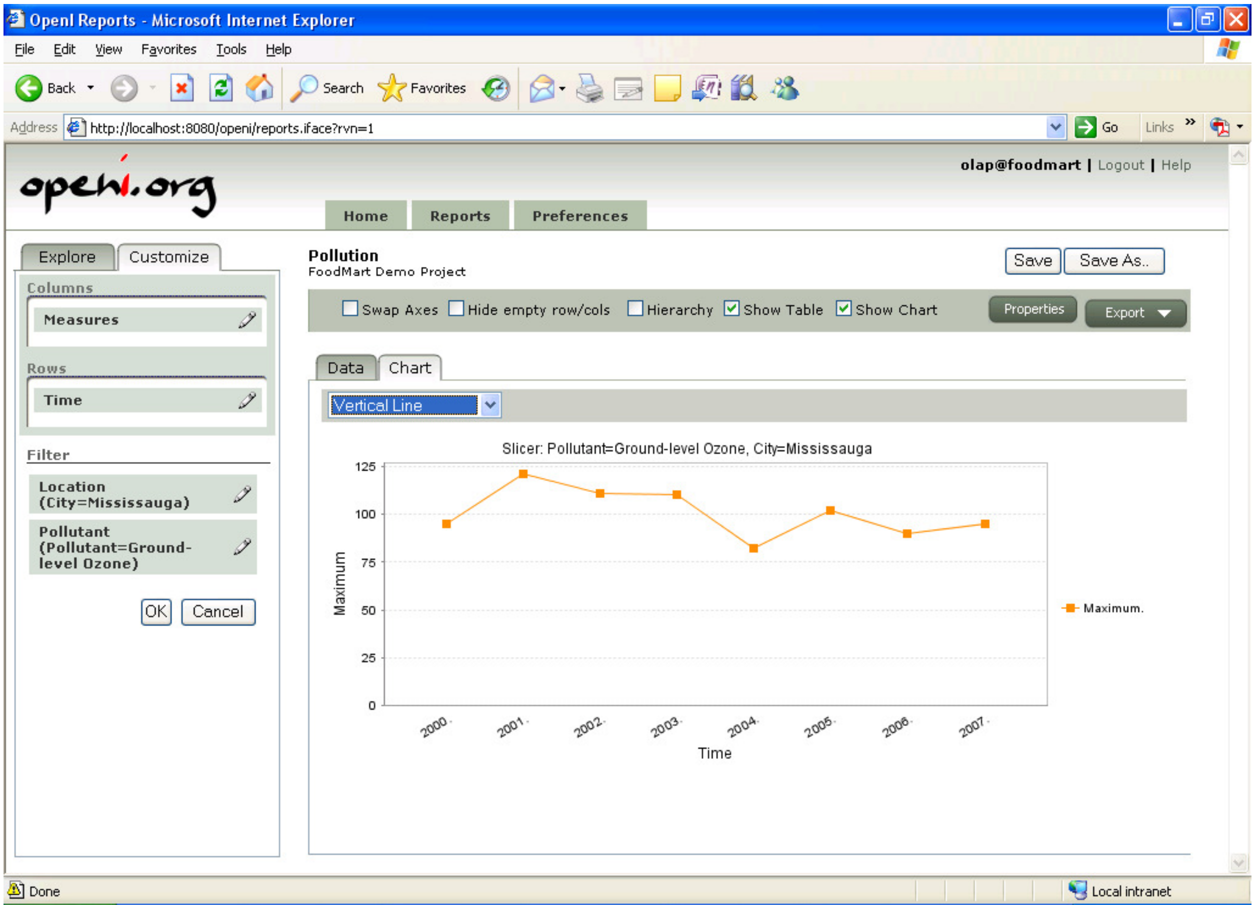


Figure 14: Time series for the city of Mississauga, pollutant ground level ozone.

The final screen shot figure 14 illustrates the level of ground level ozone for the city of Mississauga from the year 2000 – 2007. This chart can be used for modeling time series and comparative data analysis. It should be noted how the “time” dimension was dragged from the filter tab into the rows tab on the left side of the screen where, it is possible to select more than one year or location.

The analysis performed can be saved by clicking on the “save” or “save as” button on the top right side of the screen.

4.2 Validation of air quality data mart

Urban (Toronto, Ottawa, Hamilton, London and Windsor)
Rural (Sudbury, Thunder Bay, Sault Ste Marie, Sarnia, Chatham)

4.2.1 Average annual concentration levels of SO_2, CO, NO_2, O_3 and $PM_{2.5}$ for urban cities in 2005.

Two separate methods were used to determine the annual average concentration levels of SO_2, CO, NO_2, O_3 and $PM_{2.5}$. The CESI method using 24 hour average annual concentration for SO_2, NO_2, and $PM_{2.5}$ showed increased concentration levels of SO_2 in the city of Hamilton (station-id 29000) and Windsor (station- id 12008) (27.6 and 24.5 ppb) respectively. These levels do not exceed the AAQC of 100ppb. Using the air quality data mart the maximum values of SO_2 concentration were found to be exceeding the AAQC. The maximum values for city of Windsor were found to be 127 ppb. These values can be drilled down at a finer precision i.e. determine concentration values by month, day or hour the pollutant levels exceeded the AAQC.

The 24 hour average annual NO_2 concentration levels using the CESI show that the city of Hamilton having (190.3 ppb) exceeded the AAQC of 100ppb. Using air quality data mart the maximum NO_2 concentration levels for Hamilton was found to be 393 ppb more than three times the AAQC. None of the other major cities were found to surpassing the AAQC.

Using the CESI method to determine the 24 hour average annual $PM_{2.5}$ concentration levels, the results indicate that all cities exceeded the AAQC of 30µg/m^3. Windsor was found to have the highest concentration levels of (53.1 µg/m^3) followed by London (51.6 µg/m^3), Ottawa (51.2 µg/m^3), Toronto (47.7 µg/m^3) and Hamilton (45.8 µg/m^3). The maximum values of fine particulate matter were determined by using air quality data mart which showed that the concentration values exceeded more than 50µg/m^3 in the year 2005.

An 8-hour average annual concentration for carbon monoxide was calculated for some of the major cities of Ontario, the results show that none of the cities surpassed the AAQC of 13ppm. The results were then compared using the air quality data mart to determine if any of the cities showed maximum values more than the AAQC. It was found that all cities were within in the AAQC.

One hour average annual concentration for ground level ozone was determined using the CESI for the city of Hamilton, Windsor, Toronto, Ottawa and London. The results showed that all major cities exceeded the AAQC of 80 ppb. The highest levels were found for Windsor (107.5ppb) followed by Toronto (99.5ppb), London (91.0ppb) Hamilton (89.0ppb) and Ottawa (84.0ppb). Using the air quality data mart for maximum values it showed that, the values were not very different from CESI. The maximum values determined were, Windsor (108ppb), Toronto (100ppb), Hamilton (95ppb), London (93ppb) and Ottawa (84ppb).

A complete set of results can be found in Appendix A.

4.2.2 Comparison of smog elements (O_3 and $PM_{2.5}$) between the urban/rural cities for 2005.

A comparative study for smog elements was done for urban and rural areas of Ontario, using the CESI and air quality data mart. The results show that smog is not only an urban area problem but also a rural area issue. The rural areas namely "Chatham" and "Sarnia" had 1 hour average annual concentration of ground level ozone (114.0ppb) and (111.5ppb) respectively. The two urban cities Windsor and Toronto had the 1 hour average annual concentration of ground level ozone (107.5ppb) and (99.5ppb) respectively. The maximum values determined using air quality data mart showed almost similar results Chatham (116ppb) and Sarnia (115ppb).

A 24 hour average annual concentration for $PM_{2.5}$ was determined between the urban and rural areas of Ontario. The results indicate that all urban and rural areas exceed the AAQC 30µg/m^3. Sarnia was found to have the highest 24 hour average annual concentration of fine particulate matter 65.9µg/m^3 among the urban and rural areas followed by Windsor city53.130µg/m^3. The maximum values from air quality data mart indicate that Sarnia has the highest concentration levels of 114 µg/m^3 followed by Sault Ste Marie with concentration levels of 107 µg/m^3. This result does not resemble the results achieved using the CESI method, it is because all values have been averaged. The actual values to the hour precession can be determined using air quality data mart.

A complete set of results can be found in Appendix B.

4.2.3 Comparison between Toronto and Ottawa, of monthly maximum concentration levels of O_3 and $PM_{2.5}$ during the "working days" and "weekends" for 2007.

Monthly maximum concentration levels of smog elements (O_3 and $PM_{2.5}$) were determined for Toronto and Ottawa during the working day and weekends using only air quality data mart and not the CESI method. It was not considered appropriate to apply running averages with missing days in between.

The results show that the ground level ozone did not exceed the AAQC of 80 ppb during the weekends of year 2007 in Ottawa. The maximum O_3 values ranged between 33ppb – 68ppb. However during the working days Ottawa showed increased levels of O_3 mainly beginning of the smog season. The exceeded levels of O_3 were found in the month of May 90ppb and June 81ppb.

The ground level ozone did exceed the AAQC of 80ppb on the weekends of 2007 in Toronto during the months of June and July with 87ppb and 94ppb respectively. Ground level ozone concentrations were found to be highest during the working for most of the smog season which is considered to be between April-Sept, Toronto exceeded the AAQC of 80ppb during May – Sept but the maximum O_3 values recorded for Toronto during the working days were in the month of May 93ppb and June 97ppb.

Fine particulate matter is a mixture of other solid microscopic particles suspended in the air. Air quality data mart was used to determine the annual average concentration levels of $PM_{2.5}$ during the working days and weekends for the cities of Toronto and Ottawa. Results indicate that $PM_{2.5}$ concentration levels surpassed the AAQC of 30$\mu g/m^3$ on the weekends in the months of Mar 35$\mu g/m^3$ and Sept 42 $\mu g/m^3$ in Ottawa. However during the working days the $PM_{2.5}$concentration levels were found to be high ranging between (35$\mu g/m^3$ – 57$\mu g/m^3$) during the summer months (May – Sept).

Higher concentration levels of $PM_{2.5}$ were also recorded in the city of Toronto during the weekends. The values surpassed the AAQC of 30$\mu g/m^3$ in the months of June 41 $\mu g/m^3$, July and Sept 36$\mu g/m^3$. Results also indicate that the $PM_{2.5}$ concentration levels exceedances were recorded for most of the summer months (May- Sept) during the working days in Toronto. Highest levels were recorded in the month of May at 51$\mu g/m^3$.

All relevant charts from this section can be found in Appendix C.

4.2.4 Comparison between Toronto and Ottawa, for maximum concentration levels of O_3 and $PM_{2.5}$ during the “Peak Hours” and “Non Peak Hours” for the period of April 1st to Sept 30, 2007.

Maximum concentration levels of O_3 and $PM_{2.5}$ using air quality data mart were considered in this part of the analysis. Only “working days” were considered during April- September to determine the effect of peak and non peak hours during the smog season. Peak hours were set as 7am – 9am for the morning peak hours and 4pm – 7pm were considered as evening peak hours.

Results for Toronto (Table5) for $PM_{2.5}$ show that, there were 15 days during April – Sept that exceeded the AAQC of 30 $\mu g/m^3$ during the peak hours whereas 26 days during the non-peak hours. A similar pattern was determined for O3 with AAQC of 80ppb, where number of exceedance days was found to be of 6 during the peak hours and 7 days during the non-peak hours. The table 5 for Toronto and table 6 for Ottawa below, show the range of concentration levels for O_3 and $PM_{2.5}$ during the smog season (April – Sept) and number of exceedance days along with the AAQC for each pollutant.

Table 5: Number of exceedance days and range of concentration during peak &non-peak hours for Toronto

Toronto Station Id 31103		# of Exceedance Days During	Concentration Range AAQC 30 µg/m3	# of Exceedance Days During	Concentration Range AAQC 30 µg/m3
Fine Particulate Matter	Months	Peak Hours		Non Peak Hours	
	April	0	2 to 26	0	3 to 22
	May	4	3 to 51	6	4 to 51
	June	5	3 to 49	4	4 to 46
	July	1	2 to 39	6	4 to 37
	August	2	3 to 36	2	4 to 38
	September	3	3 to 48	8	4 to 50
Total # of Exceedance days		**15**		**26**	
Ground Level Ozone			AAQC 80ppb		
	April	0	15 to 58	0	24 to 70
	May	0	24 to 80	1	31 to 93
	June	2	19 to 94	2	31 to 97
	July	1	26 - 91	0	29 to 78
	August	2	27 to 87	2	34 to 87
	September	1	13 to 81	2	16 to 81
Total # of Exceedance days		**6**		**7**	

A similar pattern was found at the Ottawa downtown station id 51001(Table 6) where the number of exceedance days for $PM_{2.5}$ was found to be 10 during the peak hours and 14 days during the non-peak hours. The O_3 concentration levels were well within the AAQC of 80ppb for the entire smog season. 2-days of exceedance were found for both peak and non-peak hours. The tables in appendix D show which day exceeded the AAQC but these values can be further drilled down using air quality data mart to determine at what hour the concentration levels were surpassing the legal limits.

Table 6: **Number of exceedance days and range of concentration during peak &non-peak hours for Ottawa**

Ottawa Station Id 51001		# of Exceedance Days During	Concentration Range AAQC 30 µg/m3	# of Exceedance Days During	Concentration Range AAQC 30 µg/m3
Fine Particulate Matter	Months	Peak Hours		Non Peak Hours	
	April	0	1 to 24	0	2 to 24
	May	2	2 to 52	2	3 to 53
	June	3	2 to 48	3	2 to 57
	July	1	2 to 36	4	5 to 35
	August	1	1 to 35	1	3 to 33
	September	3	2 to 46	4	3 to 49
Total # of Exceedance days		**10**		**14**	
Ground Level Ozone			AAQC 80ppb		
	April	0	15 t 70	0	17 to 75
	May	1	26 to 84	2	29 to 90
	June	1	21 to 81	0	25 to 75
	July	0	21 to 64	0	27 to 72
	August	0	20 to 70	0	25 to 68
	September	0	11 to 79	0	14 to 78
Total # of Exceedance days		**2**		**2**	

.

4.2.5 Trend analysis for O_3 and $PM_{2.5}$ for the city of Toronto and Ottawa (2003-2007).

Patterns for ground level ozone and fine particulate matter were studied using the CESI method and air quality data mart. Results for Toronto indicate that using the one hour annual average maximum concentration for ozone, elevated levels were found for the year 2003 (111.5ppb). The city of Toronto did manage to bring the ozone concentration levels within the AAQC of 80ppb in 2004 but could not maintain it in the years ahead (2005- 2007) ranging between 91ppb-99.5ppb. Maximum concentration values were generated using air quality data mart which showed fairly similar results, highest O_3 concentration value recorded in the year 2003(115ppb). The concentration values were low in the year 2004 at 82ppb but still higher than the AAQC. This proves the advantage of using the maximum values rather than the average values which tend to smooth the true scenarios.

A similar pattern was found in the City of Ottawa, elevated levels in 2003 (101ppb), these concentration levels sharply dipped in the year 2004 (69ppb). However the O_3 concentration levels in the years ahead did not increase drastically but were fairly marginal ranging between 83.5ppb-88.5ppb. Maximum values were generated using the air quality data mart which showed similar results, increased concentration levels in 2004 at 104ppb. The year 2004 was well within the AAQC at 72 ppb, however this pattern did not continue and values in the following years ranged between 84ppb-90ppb.

Results for fine particulate matter using the 24 hour annual average concentration for both cities showed a mixed pattern but higher than the prescribed AAQC of 30$\mu g/m^3$. The values for Toronto ranged between 40.8 $\mu g/m^3$-51.8$\mu g/m^3$. The maximum $PM_{2.5}$ values generated using air quality data mart showed values ranging between (51$\mu g/m^3$– 65$\mu g/m^3$).

The results generated for $PM_{2.5}$ concentration levels for the city of Ottawa using 24 hour annual average showed that concentration values were well above the prescribed AAQC of 30 $\mu g/m^3$ ranging between (34.5 $\mu g/m^3$-51.2 $\mu g/m^3$) during the years 2003 – 2007. The maximum values ranged between (53 $\mu g/m^3$-60 $\mu g/m^3$).

All relevant charts and associated tables can be found in Appendix E.

4.2.6 Trend analysis (2000-2007) of ground level ozone concentrations for the city of Toronto and Ottawa

One hour annual average maximum for O_3 was studied for the years 2000-2007 using the running averages (CESI) method for the city of Toronto and Ottawa. Results for both cities indicate a mixed pattern, concentration values were found to be lower in the year 2000 peaked during the year 2003 and dipped sharply in the year 2004 within the AAQC of 80ppb and then rising again in the years ahead. However in Toronto the concentration values were not within the AAQC during the years 2000-2007 except for the year 2004, whereas in Ottawa the year of 2000 had 68.5 ppb and 69ppb in the year 2004 rest of the years exceeded the AAQC of 80ppb.

Similar results were generated using air quality data mart. Only maximum values were taken into consideration because with the use of plain average values the results did not indicate any pattern rather a flat line showing all values well within the limits of AAQC during the years of 2000-2007 for both cities.

Taking maximum values into consideration in air quality data mart the results present a better view of exceedances of annual O3 concentration levels. For the city of Toronto it was determined that all eight years of the study period exceeded the AAQC of 80ppb whereas using the CESI indicated that seven of the eight years exceeded the AAQC.

All relevant charts and associated tables can be found in Appendix F.

5. DISCUSSION

The implementation of air quality data mart consisted of four major components, data for specific pollutants for the period of 2000-2007 for all monitoring stations in Ontario were available at the Ontario Ministry of Environment website. Pre-processing of the data required html format files to be saved in Excel files and then to text files. This two step process could not be avoided because copying from html pages directly onto text files created some undesirable characters in the text file. This step was tedious and time consuming in all of the implementation process. Approximately 2500 files were created (43 monitoring stations, 7 pollutants and 8 years) and saved with special naming convention that identified the contents of the file in order for the ETL program to read it and insert data into tables. Missing data in html files was stored as -999 and invalid data as 9999, these values had to be removed before uploading the data in the database. This step was critical in order for the roll up function in the data mart to work properly. Finally, cleaned preprocessed data was uploaded in MySQL database (figure 2).

The database was designed using multi-dimensional modeling approach where, the table holding the actual concentrations of the pollutants known as the fact table (Pollution) is associated with the dimensions table (Time, Pollutant and Location), that add further perspectives to a given fact (figure 7).

Cube definition for air quality data mart is done within the Mondrian OLAP server. Primary keys, foreign keys, dimensions, attributes and its hierarchies were defined using XML (figure 8). All dimension tables have hierarchies with table “time” having the most hierarchy (year, month, day, hour, working day, weekend, peak hour and non peak hour), in order to analyze the pollutants with different perspectives. The fact table stores the hourly data which is high frequency and as a result generates a large number of records. This could be conveniently avoided by storing the aggregated summary data instead of hourly data. The summary data could have hierarchy like weekly, monthly or even yearly pollutant measurements. We could also store the corresponding minimum, maximum and average values. Though this will require significant pre-processing of data before loading the tables but it will also reduce the number of records. The reduction in number of records will increase the performance of ad-hoc queries and reports. Openi was used for the web based user interface. A web interface comes with a lot of additional costs like the web hosting and internet bandwidth. A desktop based thick client, not only reduces the cost of implementation but it could also create rich user interface which is not always possible using web interface. Since the desktop based user interface is only accessible within the network in an organization, it helps to eliminate the internet security threats. Lack of public access reduces the number of queries and improves the performance for internal users of the organization.

A case study was done using air quality data mart and Canadian Environment Sustainability Indicator that calculates moving averages on the pollutant concentration. Results for SO_2 concentration using the CESI did not show any of the cities exceeding the AAQC of 100ppb but the generation of annual maximum values using air quality data mart showed Windsor city exceeding the legalized criteria in the year 2005. SO_2 is known to have adverse effects on human health such as respiratory illness, cardiovascular diseases etc. The impact of SO_2 on the environment is at a much larger scale, damaging the trees, crops, acidification of soils, lakes and streams etc. A 67% of the SO2 emissions are contributed by the industries and the rest by residential, transportation, electricity and other miscellaneous sources in Ontario.

NO_2 concentrations were within the limits prescribed according the AAQC in all cities except Hamilton. NO_2 is considered a secondary pollutant and is released during the oxidation process. Transportation sector is a major contributor of NO_2 gases. Local, commercial or long range transports between the U.S and Canada can affect the regional air quality as air pollution does not respect any political boundaries. Using the CESI, results show Hamilton with 190.3ppb and maximum values generated by air quality data mart showed a 393ppb in the year 2005, these values are well above the 100ppb limits.

Results for $PM_{2.5}$ showed that all cities exceeded the AAQC of 30$\mu g/m^3$. $PM_{2.5}$ consists of ultra fine particles, which can be in the environment as a result of natural processes i.e. volcanoes or forest fires or due to chemical reactions in the atmosphere with other pollutants. Using the CESI the results indicate that Windsor had the highest concentration levels followed by London, Ottawa, Toronto and Hamilton. $PM_{2.5}$ is one of the smog components along with O_3 and volatile organic compounds.

Ground level ozone concentrations using the CESI and maximum's from the air quality data mart did show exceedance of the AAQC of 80 ppb in all cities. Windsor city was found to have the highest concentration followed by Toronto, Hamilton, London and Ottawa. O_3 emissions are a result of urbanization involving human activities such as use of vehicles, coal fired power plants, wood burning for residential purposes etc. O_3 emissions are not only known to impact the human health but damage crops and decline of forests are known due to O_3 emissions.

A comparison for O_3 and $PM_{2.5}$ between the urban and rural areas showed that air pollution is not only an urban area issue but also a rural area problem. The one hour average annual for O3 in the year 2005 indicated all but Thunder Bay (68ppb) surpassed the AAQC 80ppb. The highest concentrations were found in Chatham 114ppb. Similar results were found for $PM_{2.5}$ between the urban and rural areas. All cities exceeded the AAQC of 30$\mu g/m^3$, Sarnia was found to have the highest $PM_{2.5}$ concentration with 65.9$\mu g/m^3$. Maximum values from air quality data mart showed identical results, Chatham with highest O_3 concentrations and Sarnia with highest $PM_{2.5}$ concentrations.

O_3 concentrations during the weekends in Ottawa were within the 80ppb but increased during the working days, whereas in Toronto O_3 concentrations were above the recommended AAQC during the weekends and the working days. $PM_{2.5}$ concentrations were higher for both cities during the weekends and working days. These values were higher mostly during the summer months when higher temperatures, emissions from industrial sources, increased use of vehicles and wood burning practices during the outdoor activities i.e. camping can contribute to poor air quality.

O_3 and $PM_{2.5}$ concentrations were also compared between Toronto and Ottawa during the peak and non peak hours of working days for the year 2007. It was found out that concentration for both pollutants were higher during peak hours of the smog season (April 1st – Sept 30) for both cities. Working days were only taken into considerations and number of exceedance days was greater in Toronto compared to Ottawa. To further investigate a five year trend (2003-2007) was monitored for O_3 and $PM_{2.5}$ which illustrates that O_3 concentrations in Toronto have been above the AAQC during the five year period whereas in Ottawa concentration levels were higher in year 2003 and well below the AAQC in the year 2004. However this pattern could not be found in later years but the city did manage to keep them at marginal levels between 83ppb-88ppb approximately. $PM_{2.5}$ concentrations were higher for both cities and did not show an increasing or decreasing trend rather a mixed pattern, which forces this study to take meteorological factors into considerations i.e. temperature, wind speed, humidity, precipitation etc. However this was not the scope of this study but to implement a functional data mart with air quality data that can be used for detailed analytical purposes.

The integration of air quality data into a data mart using OLAP tool successfully generated considerably reliable analytical results. The overall purpose of designing and implementing air quality data mart for Ontario Canada was to make aggregated data available for all cities and certain regulated pollutants. Which would provide analysts, environmentalists and academics with data that can be used for statistical or trend analysis. The motivation for air quality data mart for Ontario Canada was led by a report released by OMA based on "Illness Cost of Air pollution". This report depicted the number of premature deaths, health care costs, hospital admissions and emergency room visits in the province of Ontario.

The decision to aggregate air quality data into a data mart using open source tools was based on several factors. Commercial products suffer from lack of information on the internet whereas open source tools have ample step by step implementation information. The user group's forums were found to be very helpful in the event of technical problems. Another reason in choosing an open source tool versus a commercial product was that the project can be implemented at a low cost and can be opted in countries where funding and resources are limited.

There were certain limitations using air quality data mart. The major issue was when maximum number of dimensions was selected. The resultant table provided complete information but the chart turned out to be much more cluttered.

Future Directions

Air quality data mart for Ontario Canada can be further enhanced by adding public health and meteorological data i.e. temperature, wind speed, wind direction, humidity, precipitation etc. This will not only yield historical air pollutant measurements but it would have the capabilities of conducting correlational studies against meteorological data. Furthermore it can be analyzed which diseases are prevalent and what age group is affected most. There is a need to integrate historical air quality data at spatial level to improve visual analytical capacity. It will not only provide the functionality of roll up and drill down of the data but also provide data visualization on a map. JMAP spatial OLAP is a commercial product that integrates geospatial dimensions with business intelligence. However in this project the target was to make a data mart that can be developed using open source tools and with limited funds. It can also be chosen as an option in countries where resources and funding is limited. Geo-Mondrian is an open source spatially enabled version of Mondrian OLAP server. It could not be used in this project because it is still an experimental project headed by Prof Thierry Badard at Laval University Quebec, Canada. To date there is no official release of this product. Air quality data mart in its current state lacks geospatial visualization component. It was mainly because of lack of expertise on how to implement roll up function of geometrical features. Once this task is achieved in future it will work as an effective tool for the experts. It will have the capabilities of performing statistical analysis and data visualization within the same interface. The results of the analysis can be viewed both in numerical as well as spatial form. This will be the focus of enhancement in future work. Figure 15 illustrates the future end user interface of *spatial air quality data mart* (SAQDM) for Ontario, Canada. It will be a union of OLAP and geospatial component. The developmental focus of SAQDM will be to provide user friendly interface. The user will be required to select the dimensions and measurements without having to know the query syntax. The results will be presented in table, chart and map at the same time. The SAQDM will have the ability to generate reports according to the user's data granularity needs. The reports will either be at a detailed level or summarized level to aid in comprehensive decision making.

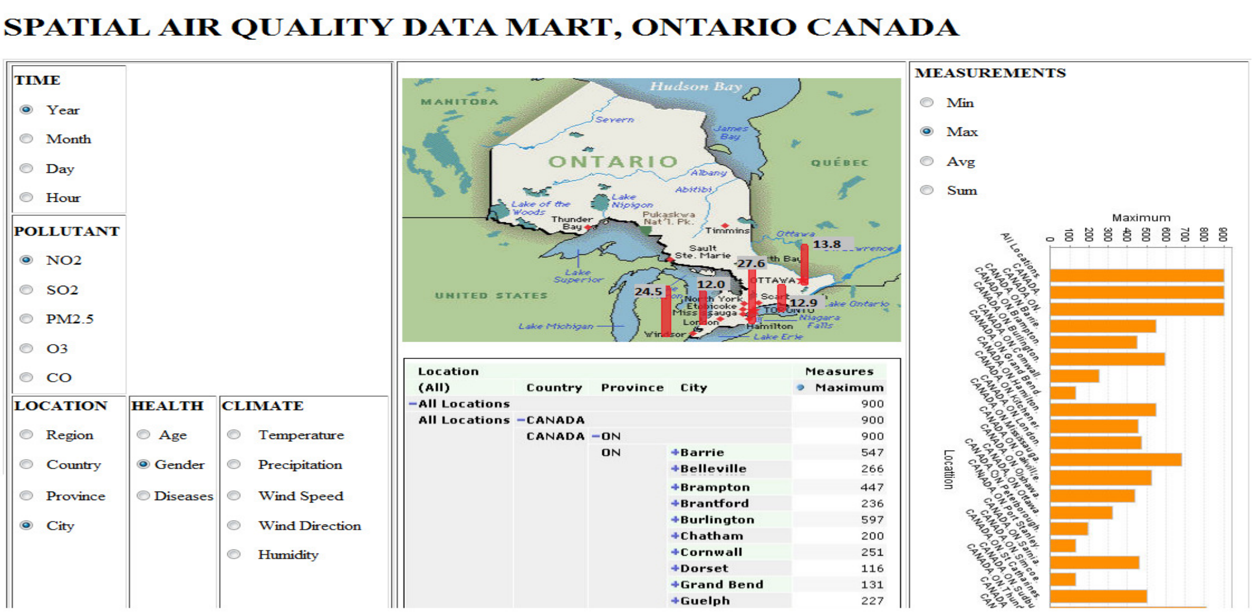


Figure 15 Future end user interface of spatial air quality data mart Ontario, Canada

CONCLUSIONS

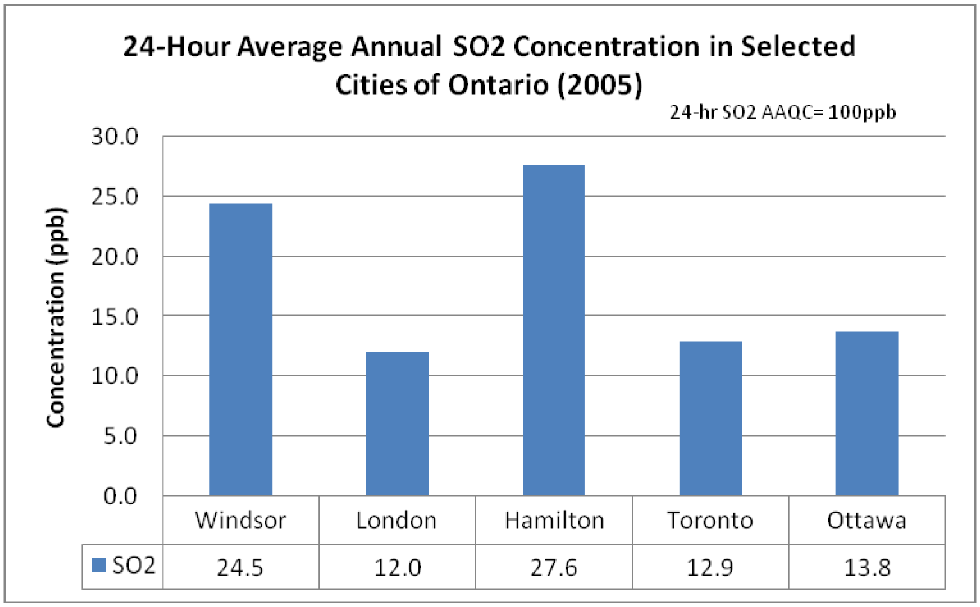
The design component of air quality data mart is achieved using star schema. This multidimensional modeling technique aids in fast retrieval of historical data. The data is saved in denormalized form. The queries are simpler to write as well as to understand by the user. However, air quality data mart is accessed by using OLAP client Openi an end user tool which provides the user with a friendly interface. The user does not need to have any query language skills, the analysis can be performed using the drag and drop feature.

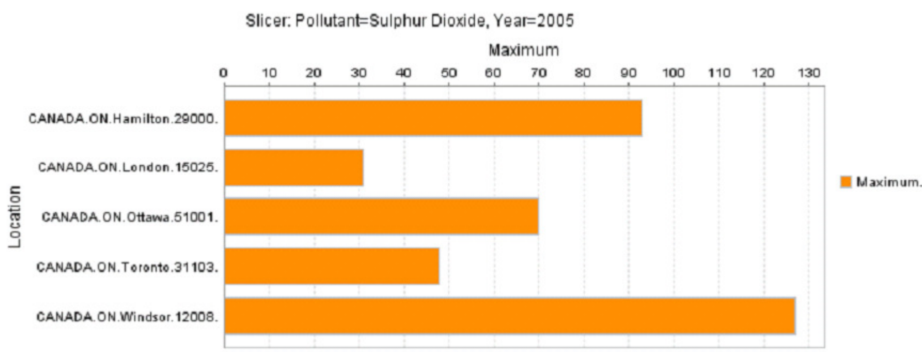
The case study is successfully performed for the two Ontario cities using air quality data mart. This data mart has the ability to roll up, drill down and slice and dice the data across several dimensions. The data granularity is up to the hour. The reports generated by air quality data mart include urban versus rural area effect, peak hour versus off peak hours effect, weekday versus weekend effect and trends of ground level ozone and particulate matter. The quantitative measures implemented in this study are AVG, MIN and MAX.

Air quality data mart at this point lacks the spatial maps component. It is easier to read and interpret data visually. If air quality data mart has some limitations, it is a pointer to future related work.

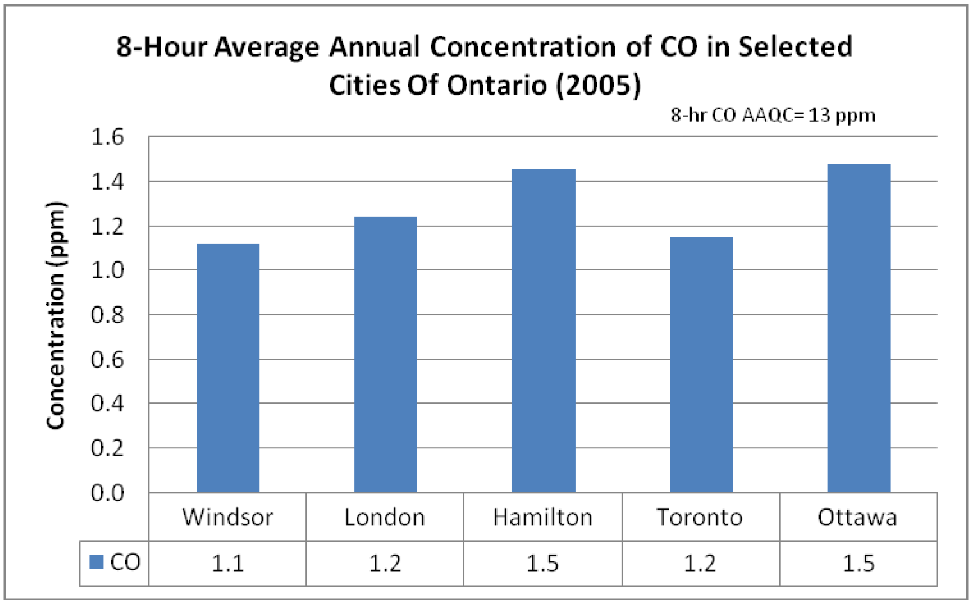
APPENDIX A

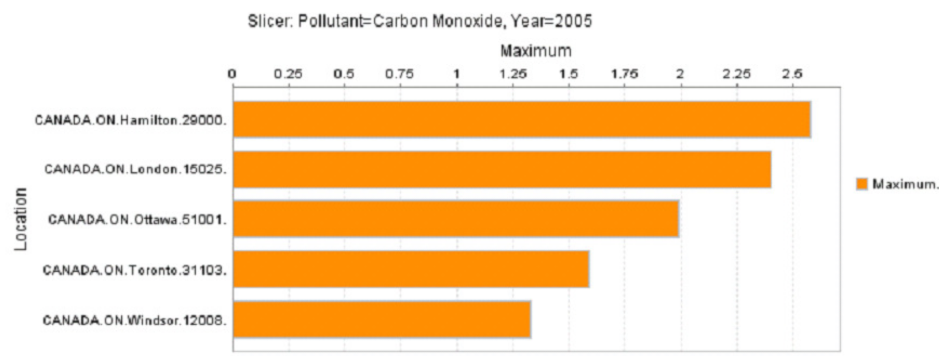
To determine for year 2005 the average annual concentration levels of SO_2, CO, NO_2, O_3 and $PM_{2.5}$ for the cities of Toronto, Ottawa, Hamilton, London and Windsor.



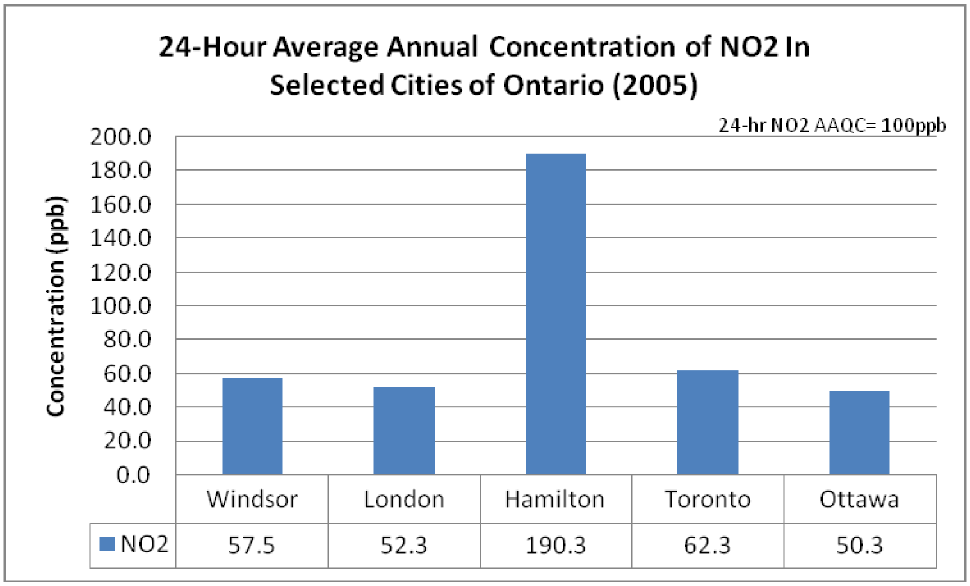


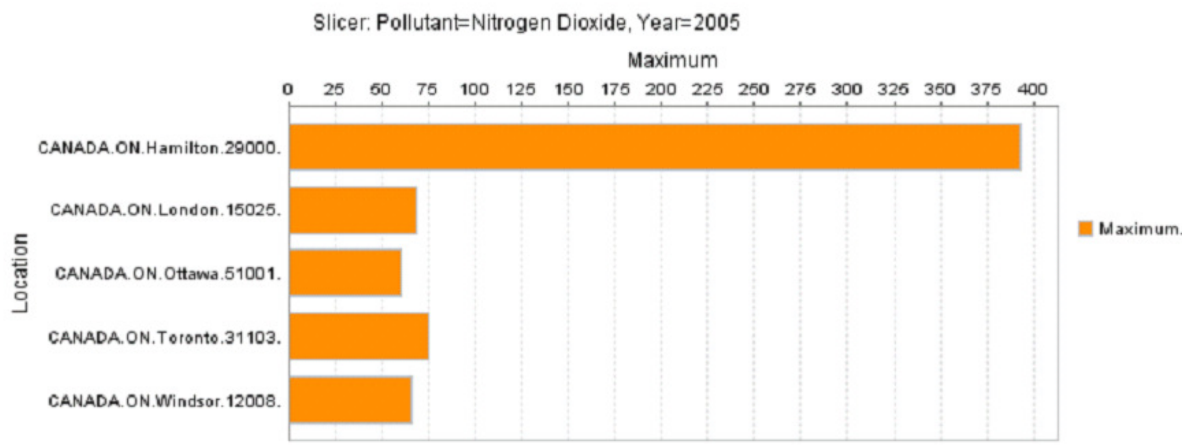
Location				Measures
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Hamilton**	**29000**	93
		London	**15025**	31
		Ottawa	**51001**	70
		Toronto	**31103**	48
		Windsor	**12008**	127



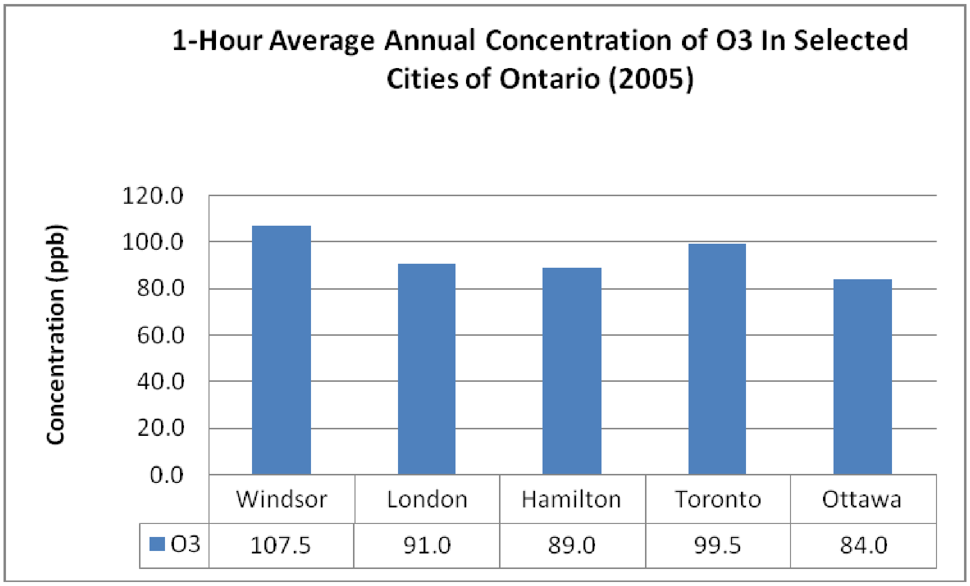


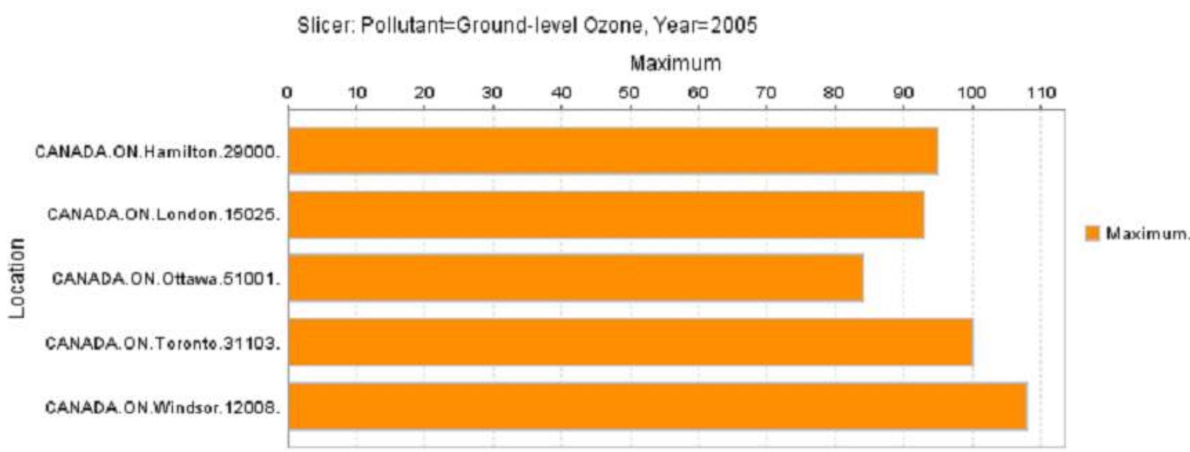
Location				**Measures**
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Hamilton**	**29000**	3
		London	**15025**	2
		Ottawa	**51001**	2
		Toronto	**31103**	2
		Windsor	**12008**	1



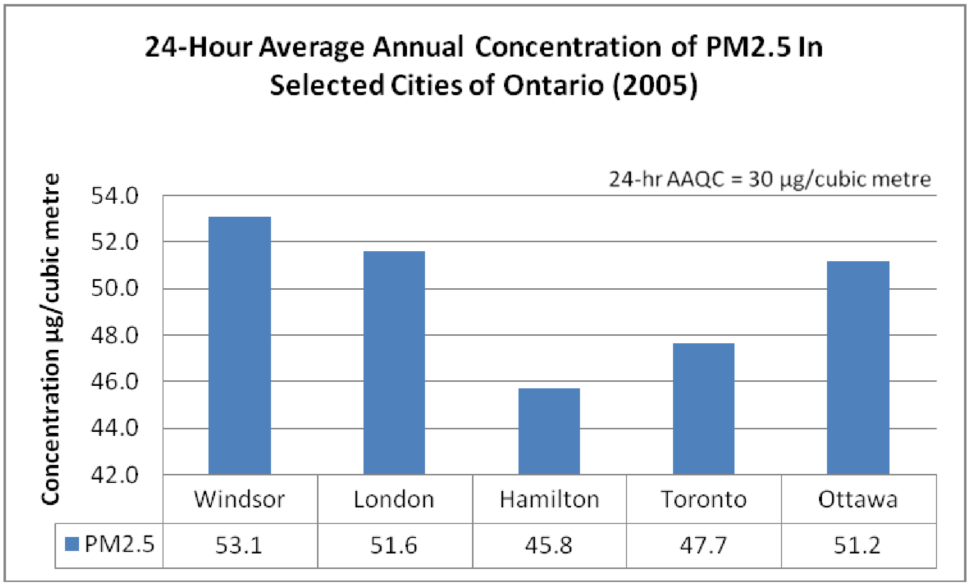


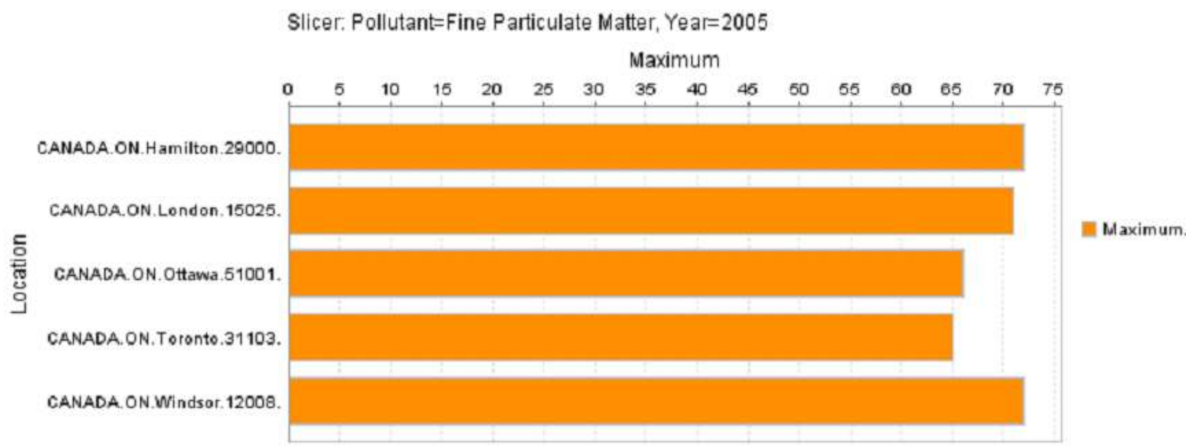
Location				Measures
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Hamilton**	**29000**	393
		London	**15025**	69
		Ottawa	**51001**	60
		Toronto	**31103**	75
		Windsor	**12008**	66





Location				**Measures**
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Hamilton**	**29000**	95
		London	**15025**	93
		Ottawa	**51001**	84
		Toronto	**31103**	100
		Windsor	**12008**	108

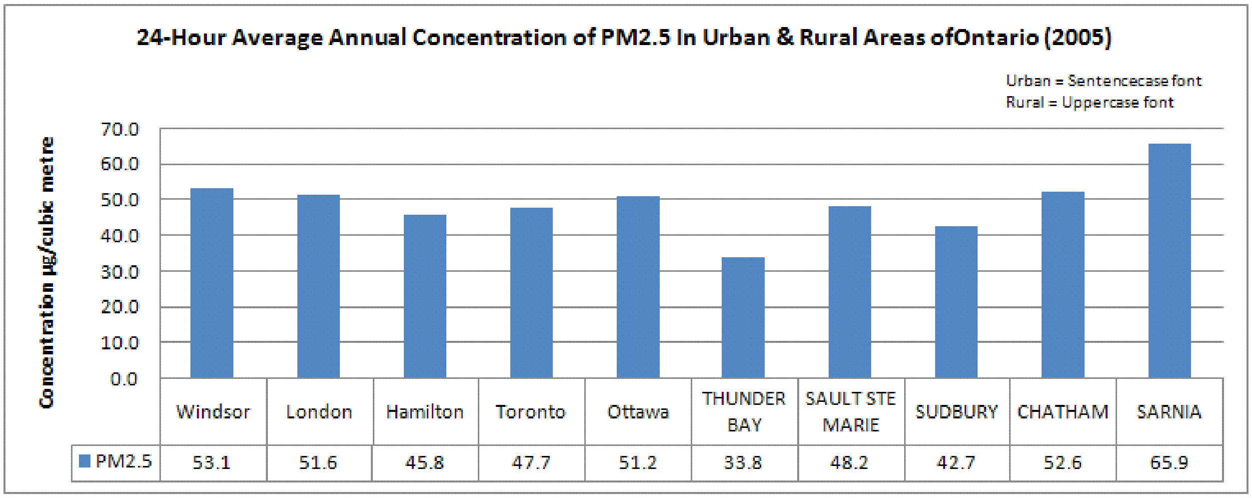


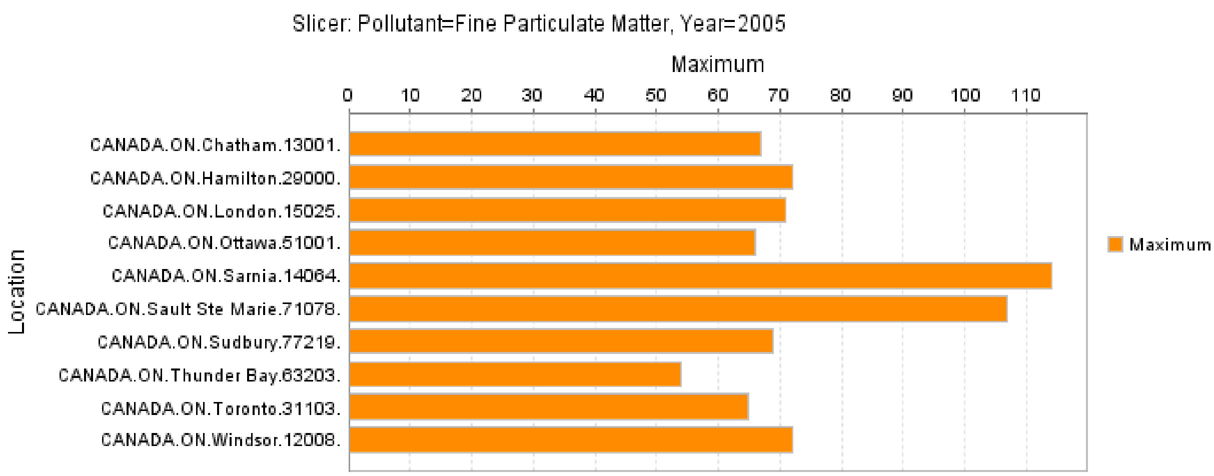


Location				Measures
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Hamilton**	**29000**	72
		London	**15025**	71
		Ottawa	**51001**	66
		Toronto	**31103**	65
		Windsor	**12008**	72

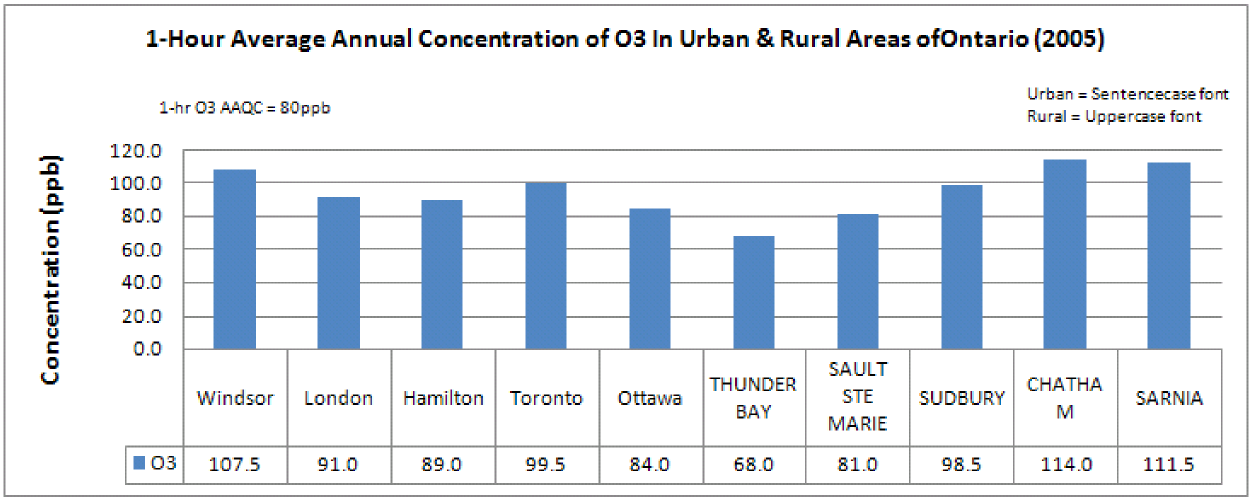
APPENDIX B

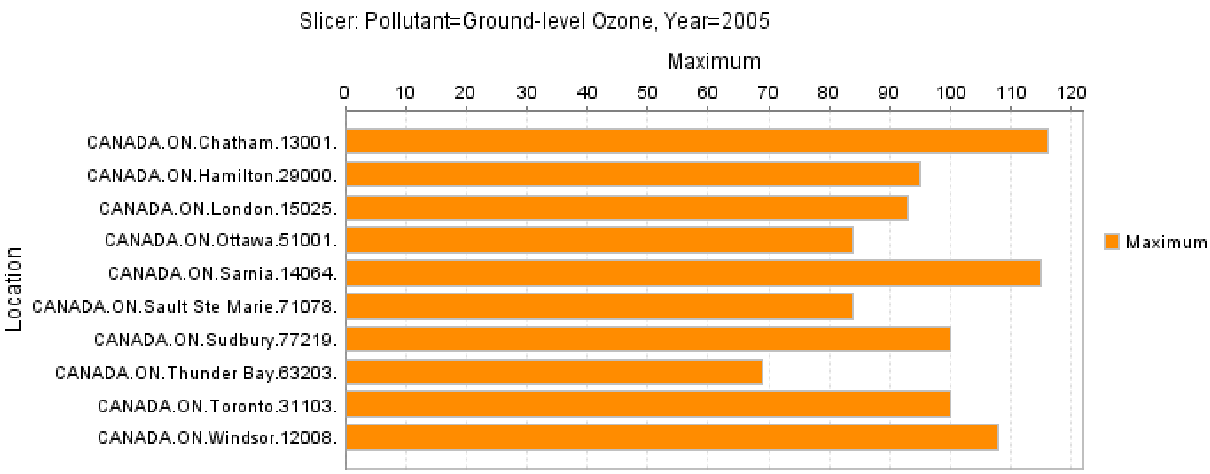
A comparison of smog elements (O_3 and $PM_{2.5}$) between the urban cities (Toronto, Ottawa, Hamilton, London, Windsor) and rural areas (Sudbury, Thunder Bay, Sault Ste Marie, Sarnia, Chatham) for the year 2005.





Location				**Measures**
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Chatham**	**13001**	67
		Hamilton	**29000**	72
		London	**15025**	71
		Ottawa	**51001**	66
		Sarnia	**14064**	114
		Sault Ste Marie	**71078**	107
		Sudbury	**77219**	69
		Thunder Bay	**63203**	54
		Toronto	**31103**	65
		Windsor	**12008**	72

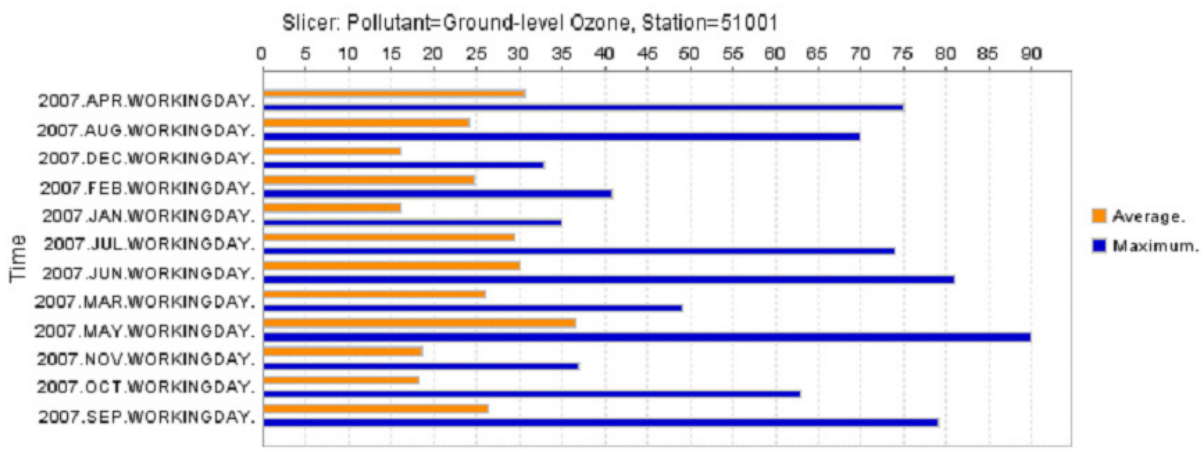


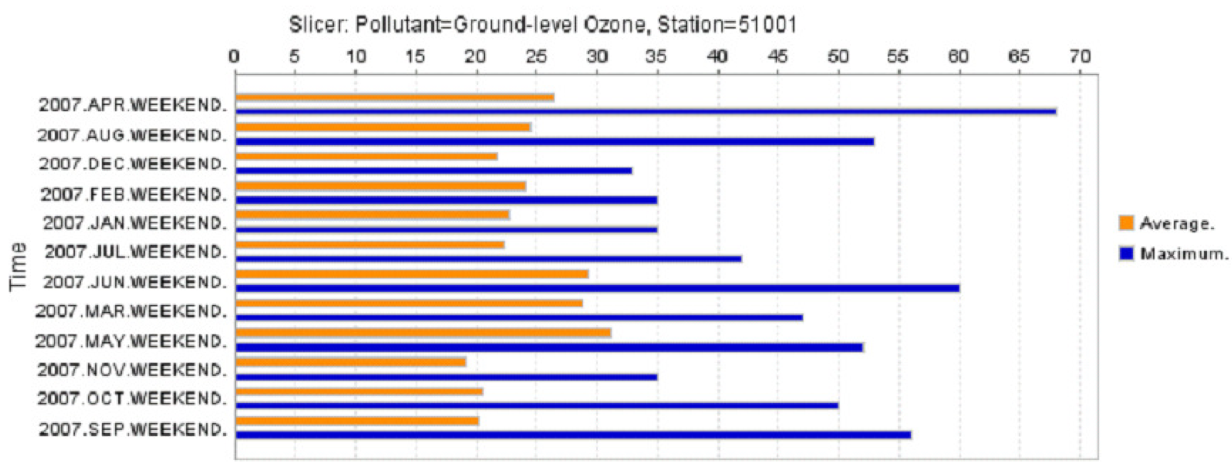


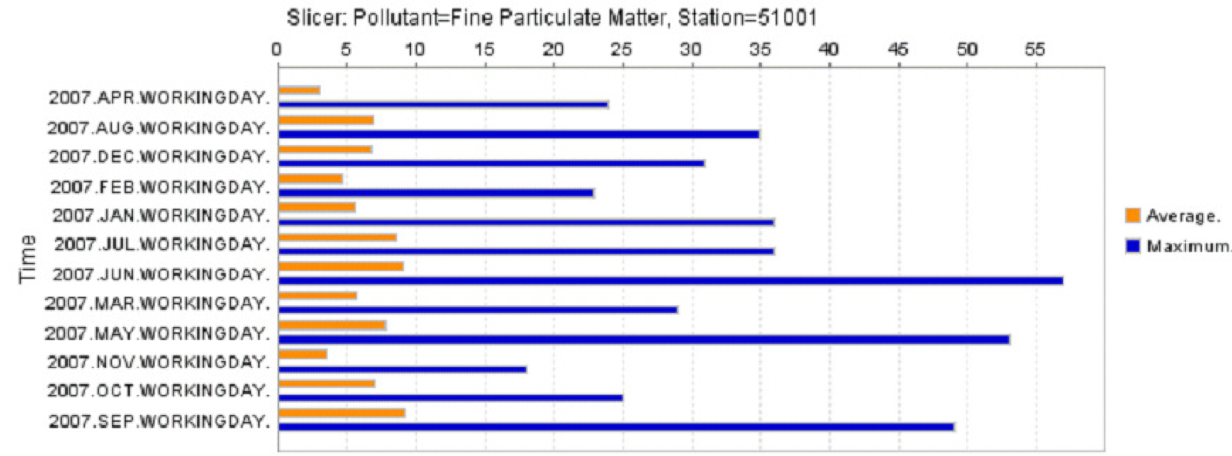
Location				**Measures**
Country	**Province**	**City**	**Station**	**Maximum**
CANADA	**ON**	**Chatham**	**13001**	116
		Hamilton	**29000**	95
		London	**15025**	93
		Ottawa	**51001**	84
		Sarnia	**14064**	115
		Sault Ste Marie	**71078**	84
		Sudbury	**77219**	100
		Thunder Bay	**63203**	69
		Toronto	**31103**	100
		Windsor	**12008**	108

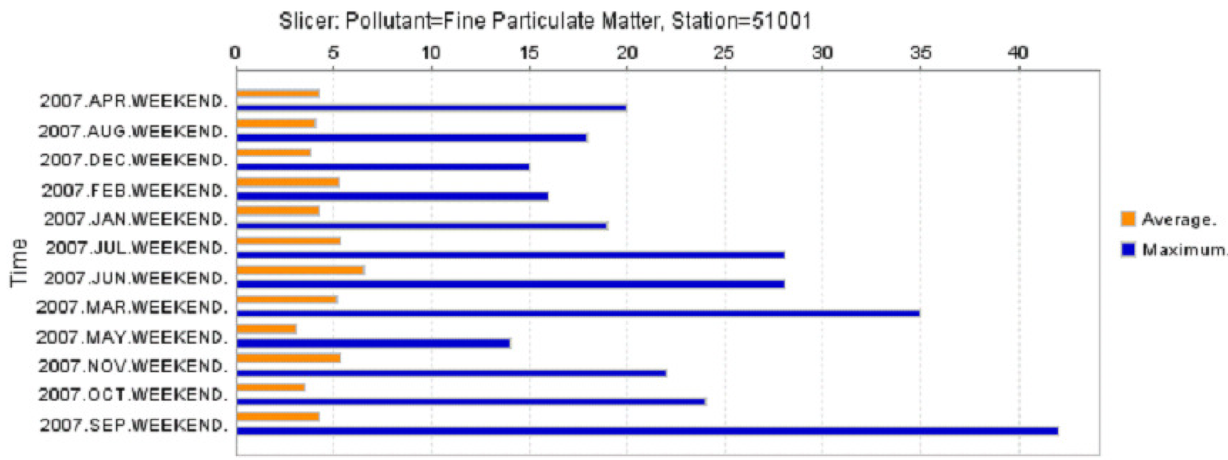
APPENDIX C

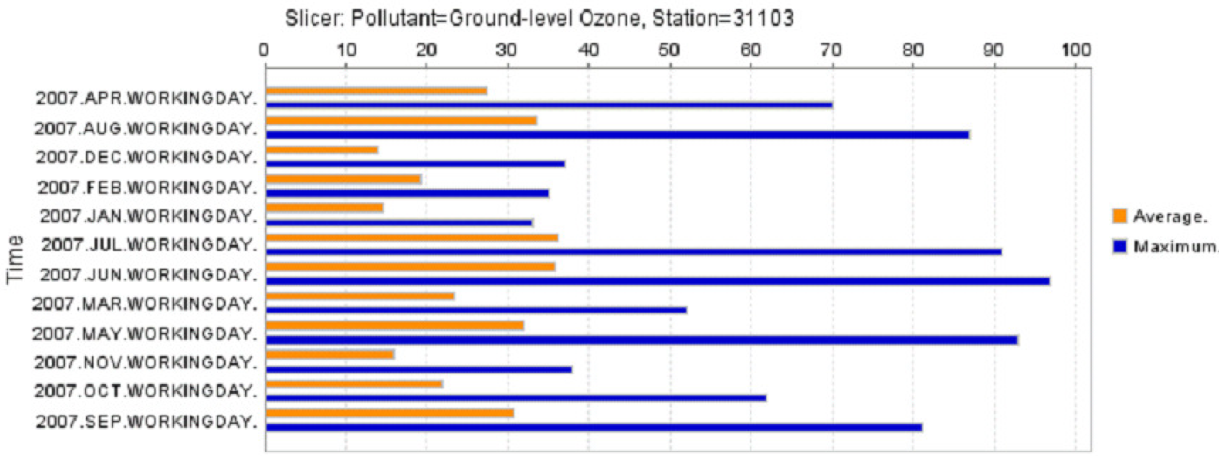
A comparison between Toronto and Ottawa, of monthly maximum concentration levels of O_3 and $PM_{2.5}$ during the "working days" and "weekends" for the year 2007.



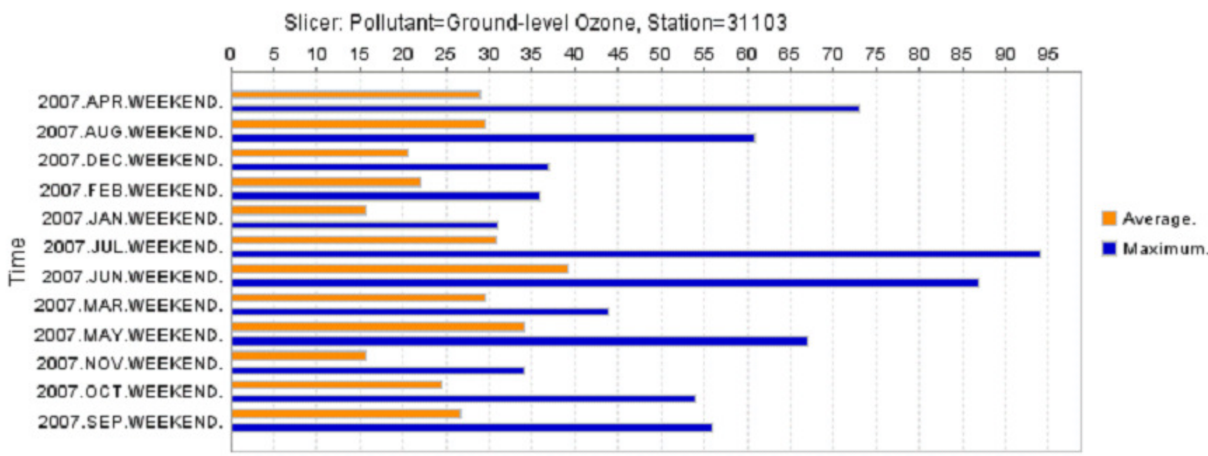



Slicer: Pollutant=Fine Particulate Matter, Station=51001
0 5 10 15 20 25 30 35 40 45 50 55
Time
2007.APR.WORKINGDAY.
2007.AUG.WORKINGDAY.
2007.DEC.WORKINGDAY.
2007.FEB.WORKINGDAY.
2007.JAN.WORKINGDAY.
2007.JUL.WORKINGDAY.
2007.JUN.WORKINGDAY.
2007.MAR.WORKINGDAY.
2007.MAY.WORKINGDAY.
2007.NOV.WORKINGDAY.
2007.OCT.WORKINGDAY.
2007.SEP.WORKINGDAY.
Average.
Maximum.

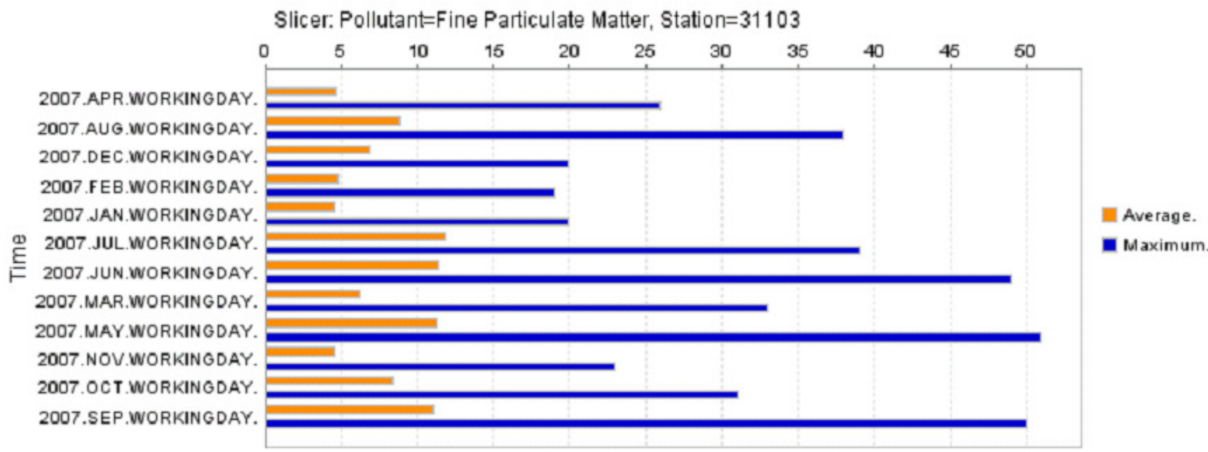

Slicer: Pollutant=Fine Particulate Matter, Station=51001
0 5 10 15 20 25 30 35 40
Time
2007.APR.WEEKEND.
2007.AUG.WEEKEND.
2007.DEC.WEEKEND.
2007.FEB.WEEKEND.
2007.JAN.WEEKEND.
2007.JUL.WEEKEND.
2007.JUN.WEEKEND.
2007.MAR.WEEKEND.
2007.MAY.WEEKEND.
2007.NOV.WEEKEND.
2007.OCT.WEEKEND.
2007.SEP.WEEKEND.
Average.
Maximum.

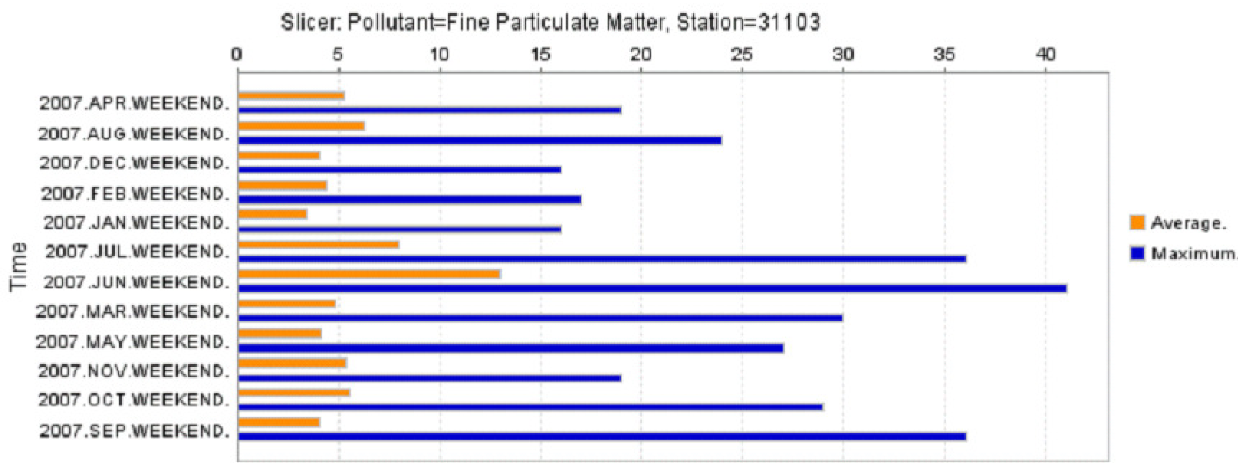


Slicer: Pollutant=Ground-level Ozone, Station=31103
0 10 20 30 40 50 60 70 80 90 100
2007.APR.WORKINGDAY.
2007.AUG.WORKINGDAY.
2007.DEC.WORKINGDAY.
2007.FEB.WORKINGDAY.
2007.JAN.WORKINGDAY.
2007.JUL.WORKINGDAY.
2007.JUN.WORKINGDAY.
2007.MAR.WORKINGDAY.
2007.MAY.WORKINGDAY.
2007.NOV.WORKINGDAY.
2007.OCT.WORKINGDAY.
2007.SEP.WORKINGDAY.
Time
Average.
Maximum.



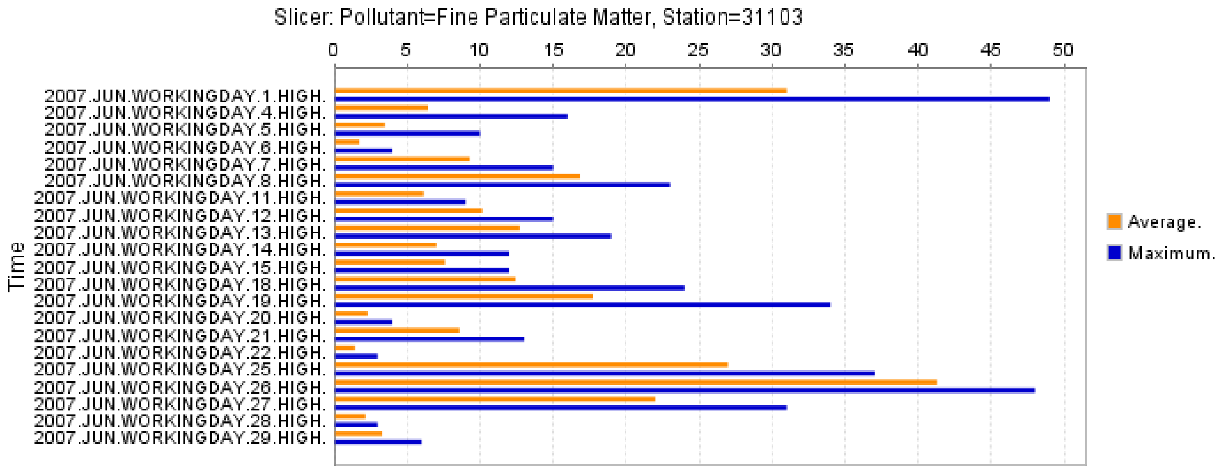
Slicer: Pollutant=Ground-level Ozone, Station=31103
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
2007.APR.WEEKEND.
2007.AUG.WEEKEND.
2007.DEC.WEEKEND.
2007.FEB.WEEKEND.
2007.JAN.WEEKEND.
2007.JUL.WEEKEND.
2007.JUN.WEEKEND.
2007.MAR.WEEKEND.
2007.MAY.WEEKEND.
2007.NOV.WEEKEND.
2007.OCT.WEEKEND.
2007.SEP.WEEKEND.
Time
Average.
Maximum.

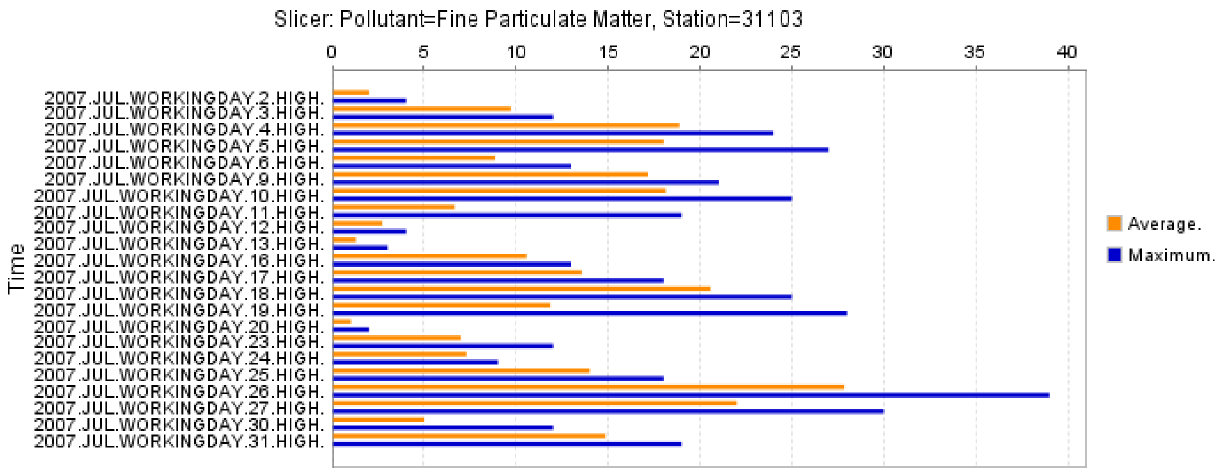

Slicer: Pollutant=Fine Particulate Matter, Station=31103
0
5
10
15
20
25
30
35
40
45
50
2007.APR.WORKINGDAY.
2007.AUG.WORKINGDAY.
2007.DEC.WORKINGDAY.
2007.FEB.WORKINGDAY.
2007.JAN.WORKINGDAY.
2007.JUL.WORKINGDAY.
2007.JUN.WORKINGDAY.
2007.MAR.WORKINGDAY.
2007.MAY.WORKINGDAY.
2007.NOV.WORKINGDAY.
2007.OCT.WORKINGDAY.
2007.SEP.WORKINGDAY.
Time
Average.
Maximum.

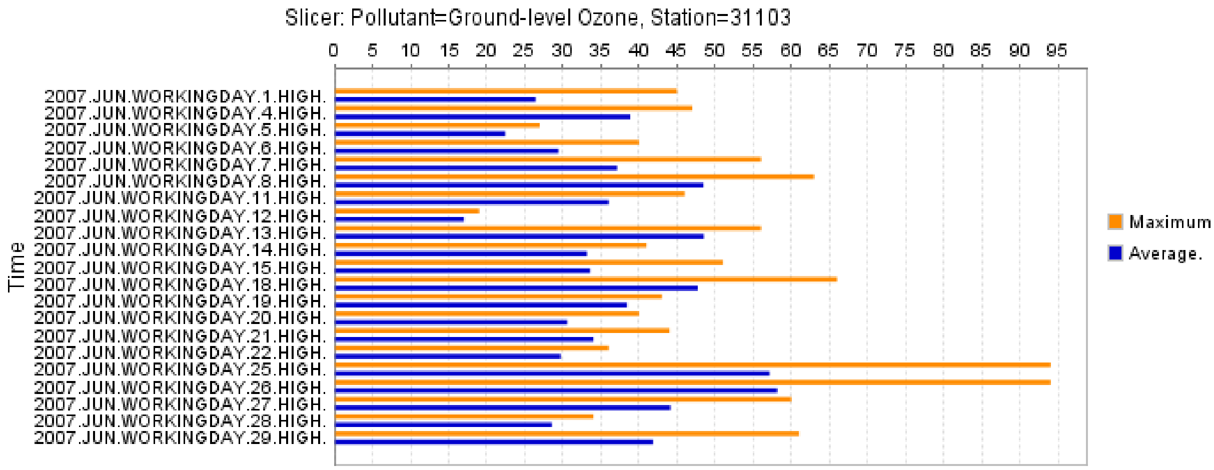

Slicer: Pollutant=Fine Particulate Matter, Station=31103
0
5
10
15
20
25
30
35
40
2007.APR.WEEKEND.
2007.AUG.WEEKEND.
2007.DEC.WEEKEND.
2007.FEB.WEEKEND.
2007.JAN.WEEKEND.
2007.JUL.WEEKEND.
2007.JUN.WEEKEND.
2007.MAR.WEEKEND.
2007.MAY.WEEKEND.
2007.NOV.WEEKEND.
2007.OCT.WEEKEND.
2007.SEP.WEEKEND.
Time
Average.
Maximum.

APPENDIX D

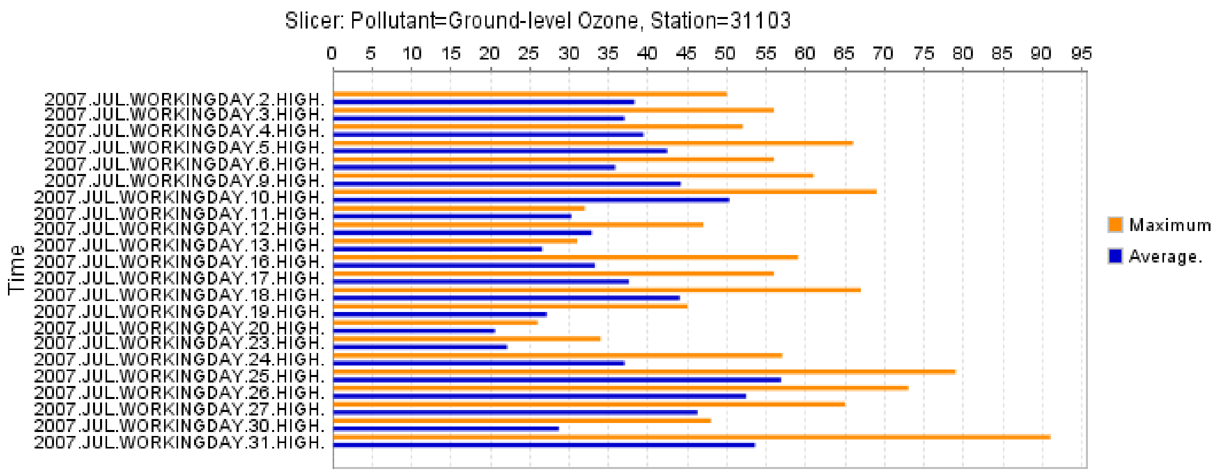
A comparison between Toronto and Ottawa, for average concentration levels of O3 and PM2.5 during the "Peak Hours" and "Non Peak Hours" for the period of April 1st to Sept 30 in the year 2007.



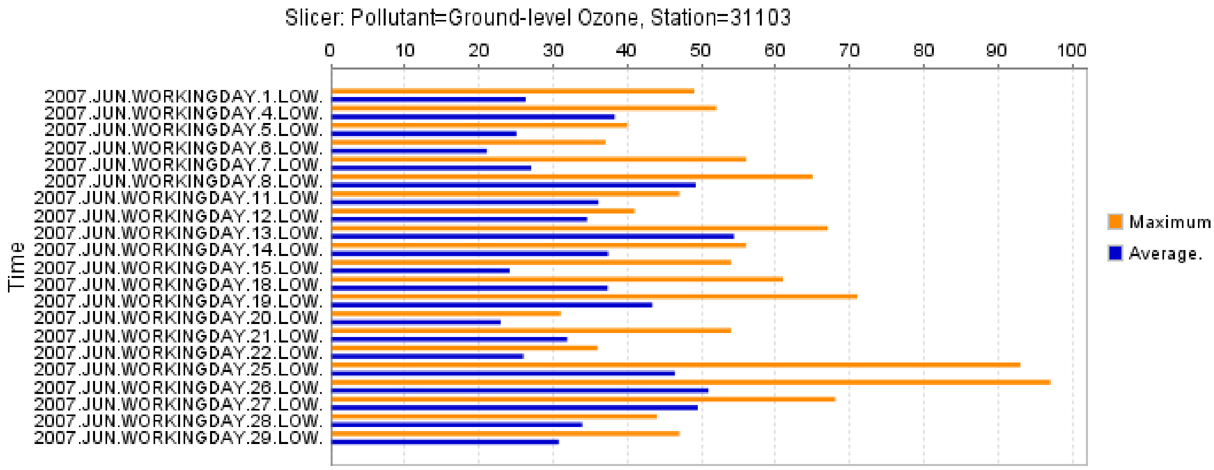


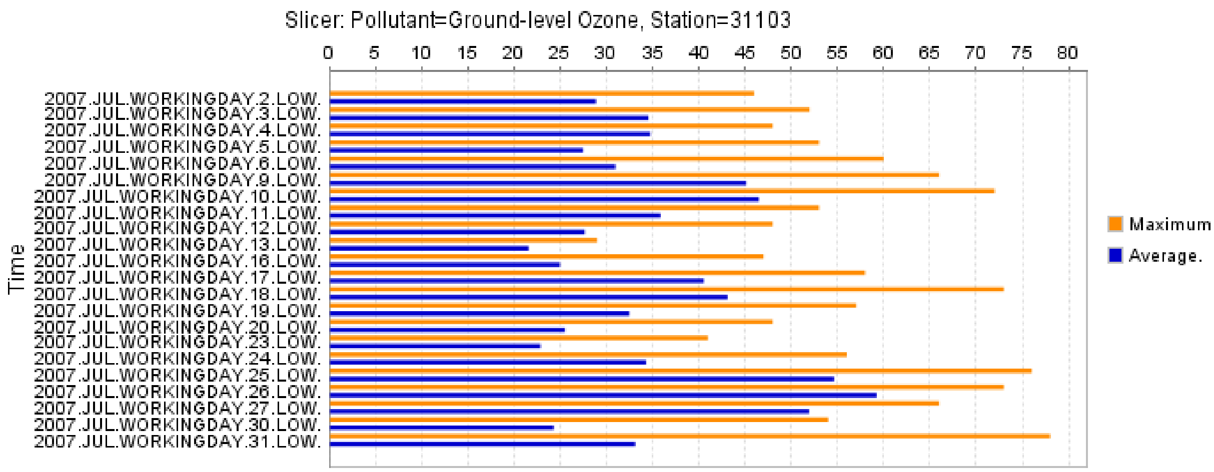


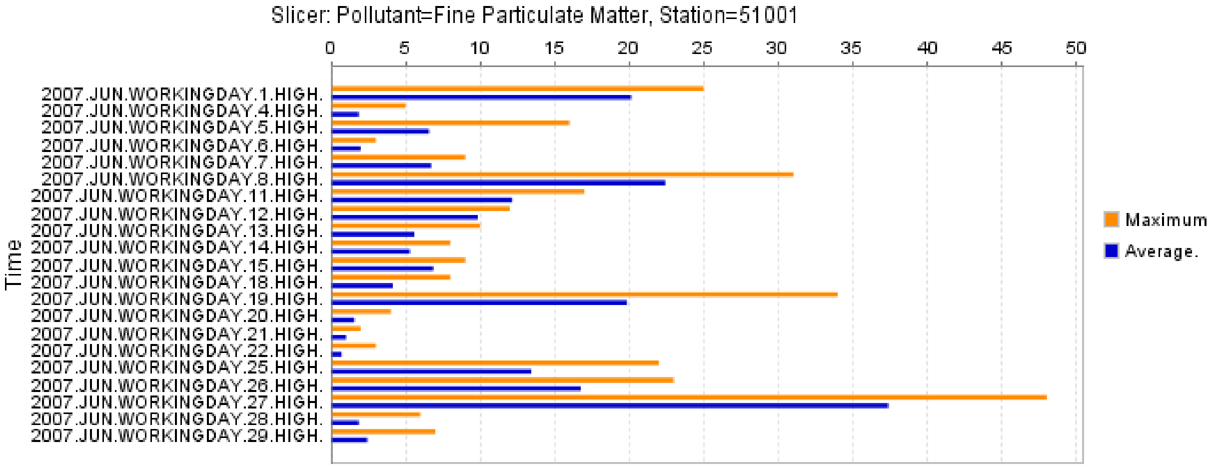
Slicer: Pollutant=Ground-level Ozone, Station=31103
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Time
2007.JUN.WORKINGDAY.1.HIGH.
2007.JUN.WORKINGDAY.4.HIGH.
2007.JUN.WORKINGDAY.5.HIGH.
2007.JUN.WORKINGDAY.6.HIGH.
2007.JUN.WORKINGDAY.7.HIGH.
2007.JUN.WORKINGDAY.8.HIGH.
2007.JUN.WORKINGDAY.11.HIGH.
2007.JUN.WORKINGDAY.12.HIGH.
2007.JUN.WORKINGDAY.13.HIGH.
2007.JUN.WORKINGDAY.14.HIGH.
2007.JUN.WORKINGDAY.15.HIGH.
2007.JUN.WORKINGDAY.18.HIGH.
2007.JUN.WORKINGDAY.19.HIGH.
2007.JUN.WORKINGDAY.20.HIGH.
2007.JUN.WORKINGDAY.21.HIGH.
2007.JUN.WORKINGDAY.22.HIGH.
2007.JUN.WORKINGDAY.25.HIGH.
2007.JUN.WORKINGDAY.26.HIGH.
2007.JUN.WORKINGDAY.27.HIGH.
2007.JUN.WORKINGDAY.28.HIGH.
2007.JUN.WORKINGDAY.29.HIGH.
Maximum.
Average.

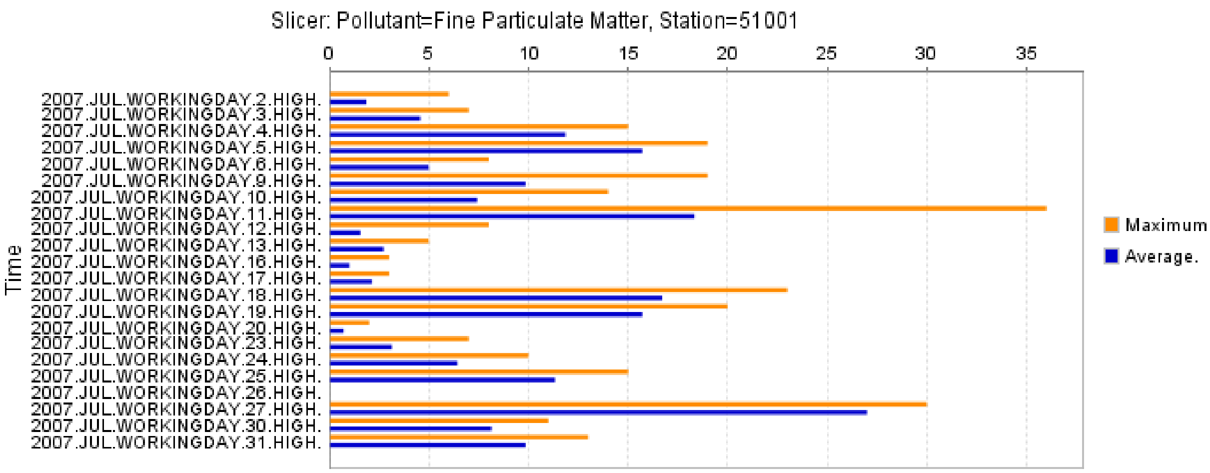


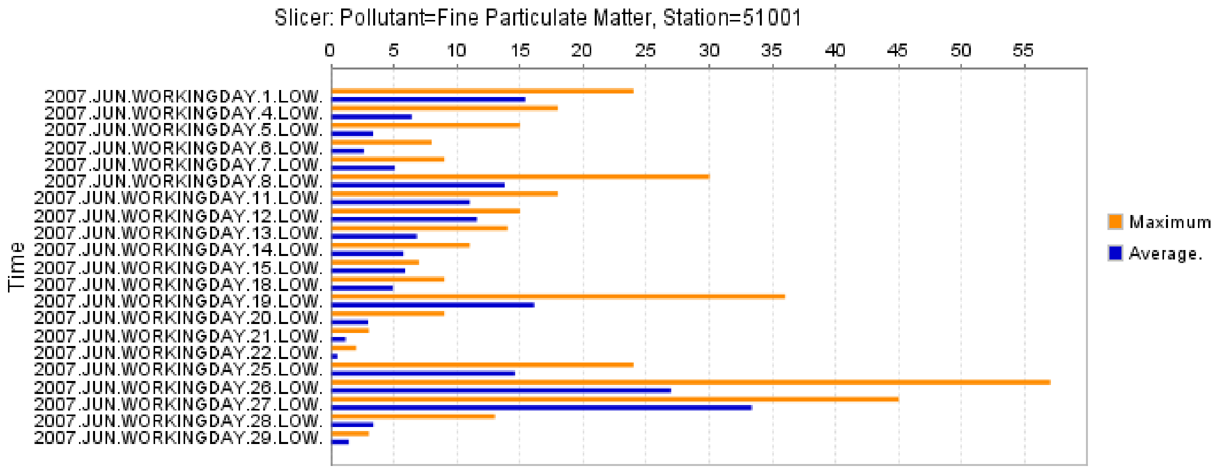
Slicer: Pollutant=Ground-level Ozone, Station=31103
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95
Time
2007.JUL.WORKINGDAY.2.HIGH.
2007.JUL.WORKINGDAY.3.HIGH.
2007.JUL.WORKINGDAY.4.HIGH.
2007.JUL.WORKINGDAY.5.HIGH.
2007.JUL.WORKINGDAY.6.HIGH.
2007.JUL.WORKINGDAY.9.HIGH.
2007.JUL.WORKINGDAY.10.HIGH.
2007.JUL.WORKINGDAY.11.HIGH.
2007.JUL.WORKINGDAY.12.HIGH.
2007.JUL.WORKINGDAY.13.HIGH.
2007.JUL.WORKINGDAY.16.HIGH.
2007.JUL.WORKINGDAY.17.HIGH.
2007.JUL.WORKINGDAY.18.HIGH.
2007.JUL.WORKINGDAY.19.HIGH.
2007.JUL.WORKINGDAY.20.HIGH.
2007.JUL.WORKINGDAY.23.HIGH.
2007.JUL.WORKINGDAY.24.HIGH.
2007.JUL.WORKINGDAY.25.HIGH.
2007.JUL.WORKINGDAY.26.HIGH.
2007.JUL.WORKINGDAY.27.HIGH.
2007.JUL.WORKINGDAY.30.HIGH.
2007.JUL.WORKINGDAY.31.HIGH.
Maximum.
Average.

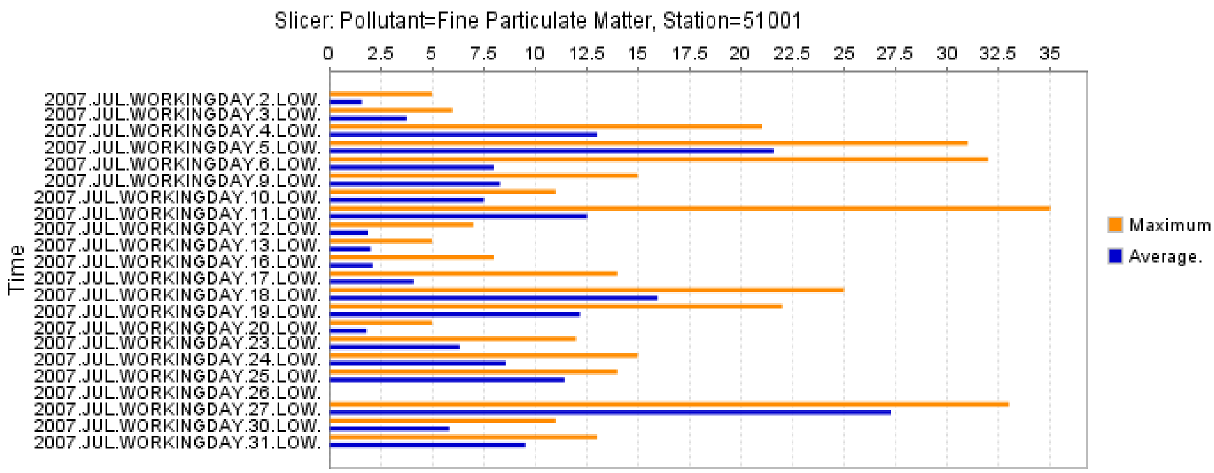

Slicer: Pollutant=Ground-level Ozone, Station=31103
0
10
20
30
40
50
60
70
80
90
100
2007.JUN.WORKINGDAY.1.LOW.
2007.JUN.WORKINGDAY.4.LOW.
2007.JUN.WORKINGDAY.5.LOW.
2007.JUN.WORKINGDAY.6.LOW.
2007.JUN.WORKINGDAY.7.LOW.
2007.JUN.WORKINGDAY.8.LOW.
2007.JUN.WORKINGDAY.11.LOW.
2007.JUN.WORKINGDAY.12.LOW.
2007.JUN.WORKINGDAY.13.LOW.
2007.JUN.WORKINGDAY.14.LOW.
2007.JUN.WORKINGDAY.15.LOW.
2007.JUN.WORKINGDAY.18.LOW.
2007.JUN.WORKINGDAY.19.LOW.
2007.JUN.WORKINGDAY.20.LOW.
2007.JUN.WORKINGDAY.21.LOW.
2007.JUN.WORKINGDAY.22.LOW.
2007.JUN.WORKINGDAY.25.LOW.
2007.JUN.WORKINGDAY.26.LOW.
2007.JUN.WORKINGDAY.27.LOW.
2007.JUN.WORKINGDAY.28.LOW.
2007.JUN.WORKINGDAY.29.LOW.
Time
Maximum.
Average.

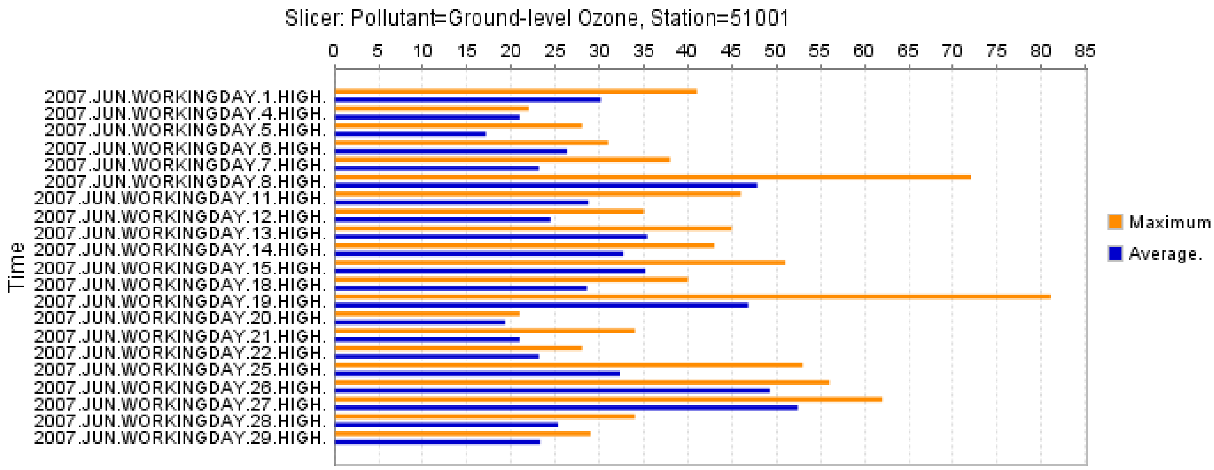

Slicer: Pollutant=Ground-level Ozone, Station=31103
0
5
10
15
20
25
30
35
40
45
50
55
60
65
70
75
80
2007.JUL.WORKINGDAY.2.LOW.
2007.JUL.WORKINGDAY.3.LOW.
2007.JUL.WORKINGDAY.4.LOW.
2007.JUL.WORKINGDAY.5.LOW.
2007.JUL.WORKINGDAY.6.LOW.
2007.JUL.WORKINGDAY.9.LOW.
2007.JUL.WORKINGDAY.10.LOW.
2007.JUL.WORKINGDAY.11.LOW.
2007.JUL.WORKINGDAY.12.LOW.
2007.JUL.WORKINGDAY.13.LOW.
2007.JUL.WORKINGDAY.16.LOW.
2007.JUL.WORKINGDAY.17.LOW.
2007.JUL.WORKINGDAY.18.LOW.
2007.JUL.WORKINGDAY.19.LOW.
2007.JUL.WORKINGDAY.20.LOW.
2007.JUL.WORKINGDAY.23.LOW.
2007.JUL.WORKINGDAY.24.LOW.
2007.JUL.WORKINGDAY.25.LOW.
2007.JUL.WORKINGDAY.26.LOW.
2007.JUL.WORKINGDAY.27.LOW.
2007.JUL.WORKINGDAY.30.LOW.
2007.JUL.WORKINGDAY.31.LOW.
Time
Maximum.
Average.

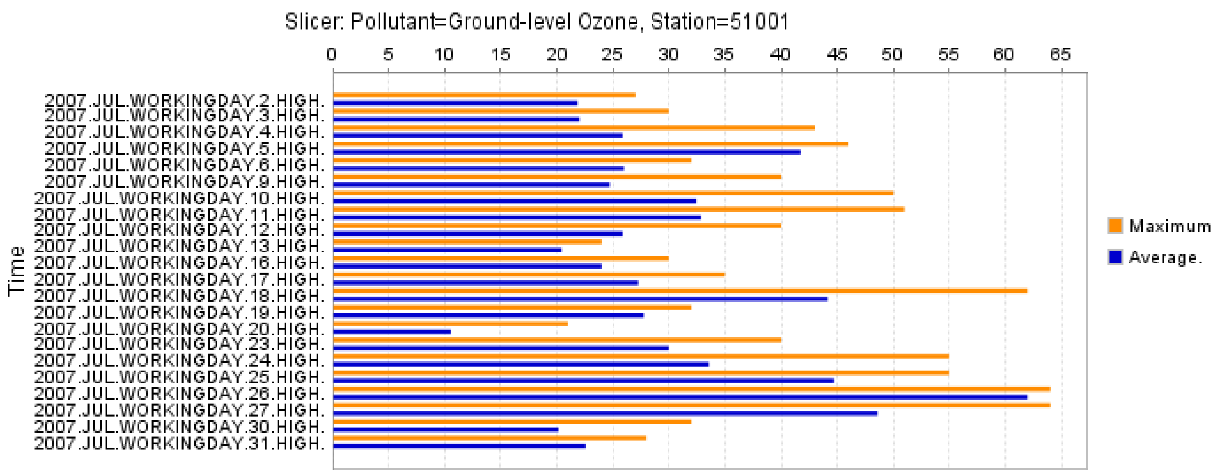

Slicer: Pollutant=Fine Particulate Matter, Station=51001
0
5
10
15
20
25
30
35
40
45
50
Time
2007.JUN.WORKINGDAY.1.HIGH.
2007.JUN.WORKINGDAY.4.HIGH.
2007.JUN.WORKINGDAY.5.HIGH.
2007.JUN.WORKINGDAY.6.HIGH.
2007.JUN.WORKINGDAY.7.HIGH.
2007.JUN.WORKINGDAY.8.HIGH.
2007.JUN.WORKINGDAY.11.HIGH.
2007.JUN.WORKINGDAY.12.HIGH.
2007.JUN.WORKINGDAY.13.HIGH.
2007.JUN.WORKINGDAY.14.HIGH.
2007.JUN.WORKINGDAY.15.HIGH.
2007.JUN.WORKINGDAY.18.HIGH.
2007.JUN.WORKINGDAY.19.HIGH.
2007.JUN.WORKINGDAY.20.HIGH.
2007.JUN.WORKINGDAY.21.HIGH.
2007.JUN.WORKINGDAY.22.HIGH.
2007.JUN.WORKINGDAY.25.HIGH.
2007.JUN.WORKINGDAY.26.HIGH.
2007.JUN.WORKINGDAY.27.HIGH.
2007.JUN.WORKINGDAY.28.HIGH.
2007.JUN.WORKINGDAY.29.HIGH.
Maximum.
Average.

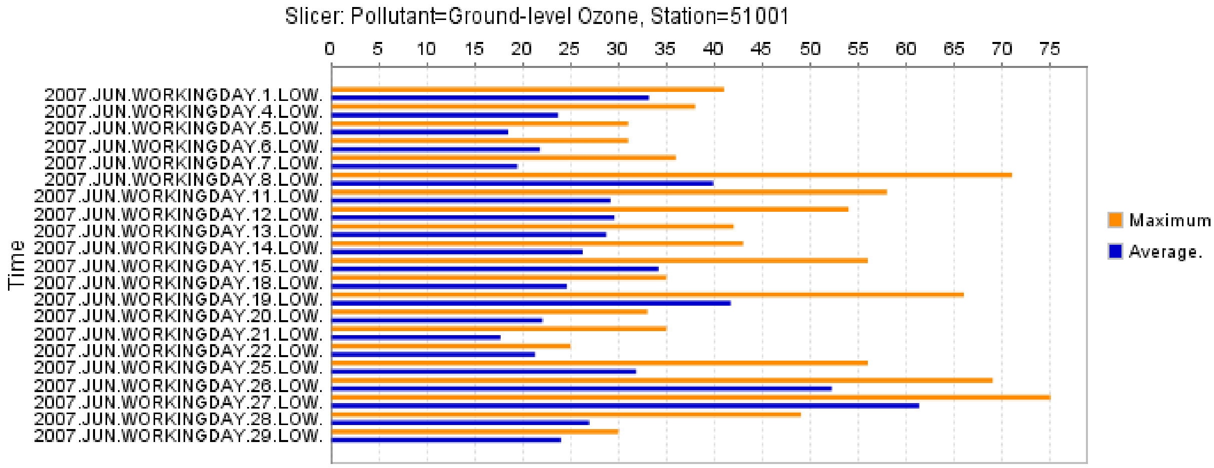

Slicer: Pollutant=Fine Particulate Matter, Station=51001
0
5
10
15
20
25
30
35
Time
2007.JUL.WORKINGDAY.2.HIGH.
2007.JUL.WORKINGDAY.3.HIGH.
2007.JUL.WORKINGDAY.4.HIGH.
2007.JUL.WORKINGDAY.5.HIGH.
2007.JUL.WORKINGDAY.6.HIGH.
2007.JUL.WORKINGDAY.9.HIGH.
2007.JUL.WORKINGDAY.10.HIGH.
2007.JUL.WORKINGDAY.11.HIGH.
2007.JUL.WORKINGDAY.12.HIGH.
2007.JUL.WORKINGDAY.13.HIGH.
2007.JUL.WORKINGDAY.16.HIGH.
2007.JUL.WORKINGDAY.17.HIGH.
2007.JUL.WORKINGDAY.18.HIGH.
2007.JUL.WORKINGDAY.19.HIGH.
2007.JUL.WORKINGDAY.20.HIGH.
2007.JUL.WORKINGDAY.23.HIGH.
2007.JUL.WORKINGDAY.24.HIGH.
2007.JUL.WORKINGDAY.25.HIGH.
2007.JUL.WORKINGDAY.26.HIGH.
2007.JUL.WORKINGDAY.27.HIGH.
2007.JUL.WORKINGDAY.30.HIGH.
2007.JUL.WORKINGDAY.31.HIGH.
Maximum.
Average.

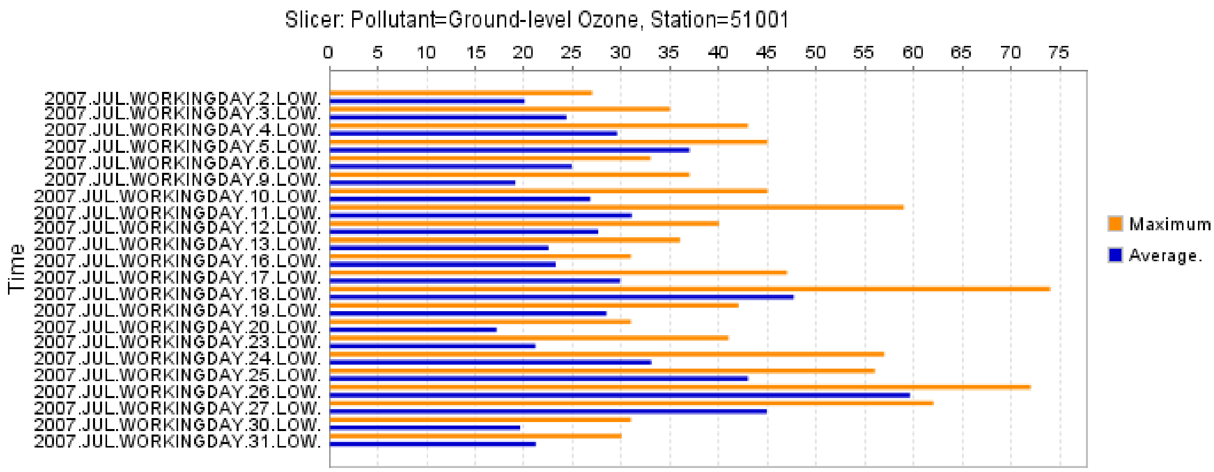

Slicer: Pollutant=Fine Particulate Matter, Station=51001
0 5 10 15 20 25 30 35 40 45 50 55
Time
2007.JUN.WORKINGDAY.1.LOW.
2007.JUN.WORKINGDAY.4.LOW.
2007.JUN.WORKINGDAY.5.LOW.
2007.JUN.WORKINGDAY.6.LOW.
2007.JUN.WORKINGDAY.7.LOW.
2007.JUN.WORKINGDAY.8.LOW.
2007.JUN.WORKINGDAY.11.LOW.
2007.JUN.WORKINGDAY.12.LOW.
2007.JUN.WORKINGDAY.13.LOW.
2007.JUN.WORKINGDAY.14.LOW.
2007.JUN.WORKINGDAY.15.LOW.
2007.JUN.WORKINGDAY.18.LOW.
2007.JUN.WORKINGDAY.19.LOW.
2007.JUN.WORKINGDAY.20.LOW.
2007.JUN.WORKINGDAY.21.LOW.
2007.JUN.WORKINGDAY.22.LOW.
2007.JUN.WORKINGDAY.25.LOW.
2007.JUN.WORKINGDAY.26.LOW.
2007.JUN.WORKINGDAY.27.LOW.
2007.JUN.WORKINGDAY.28.LOW.
2007.JUN.WORKINGDAY.29.LOW.
Maximum.
Average.


Slicer: Pollutant=Fine Particulate Matter, Station=51001
0 2.5 5 7.5 10 12.5 15 17.5 20 22.5 25 27.5 30 32.5 35
Time
2007.JUL.WORKINGDAY.2.LOW.
2007.JUL.WORKINGDAY.3.LOW.
2007.JUL.WORKINGDAY.4.LOW.
2007.JUL.WORKINGDAY.5.LOW.
2007.JUL.WORKINGDAY.6.LOW.
2007.JUL.WORKINGDAY.9.LOW.
2007.JUL.WORKINGDAY.10.LOW.
2007.JUL.WORKINGDAY.11.LOW.
2007.JUL.WORKINGDAY.12.LOW.
2007.JUL.WORKINGDAY.13.LOW.
2007.JUL.WORKINGDAY.16.LOW.
2007.JUL.WORKINGDAY.17.LOW.
2007.JUL.WORKINGDAY.18.LOW.
2007.JUL.WORKINGDAY.19.LOW.
2007.JUL.WORKINGDAY.20.LOW.
2007.JUL.WORKINGDAY.23.LOW.
2007.JUL.WORKINGDAY.24.LOW.
2007.JUL.WORKINGDAY.25.LOW.
2007.JUL.WORKINGDAY.26.LOW.
2007.JUL.WORKINGDAY.27.LOW.
2007.JUL.WORKINGDAY.30.LOW.
2007.JUL.WORKINGDAY.31.LOW.
Maximum.
Average.


Slicer: Pollutant=Ground-level Ozone, Station=51001
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
Time
2007.JUN.WORKINGDAY.1.HIGH.
2007.JUN.WORKINGDAY.4.HIGH.
2007.JUN.WORKINGDAY.5.HIGH.
2007.JUN.WORKINGDAY.6.HIGH.
2007.JUN.WORKINGDAY.7.HIGH.
2007.JUN.WORKINGDAY.8.HIGH.
2007.JUN.WORKINGDAY.11.HIGH.
2007.JUN.WORKINGDAY.12.HIGH.
2007.JUN.WORKINGDAY.13.HIGH.
2007.JUN.WORKINGDAY.14.HIGH.
2007.JUN.WORKINGDAY.15.HIGH.
2007.JUN.WORKINGDAY.18.HIGH.
2007.JUN.WORKINGDAY.19.HIGH.
2007.JUN.WORKINGDAY.20.HIGH.
2007.JUN.WORKINGDAY.21.HIGH.
2007.JUN.WORKINGDAY.22.HIGH.
2007.JUN.WORKINGDAY.25.HIGH.
2007.JUN.WORKINGDAY.26.HIGH.
2007.JUN.WORKINGDAY.27.HIGH.
2007.JUN.WORKINGDAY.28.HIGH.
2007.JUN.WORKINGDAY.29.HIGH.
Maximum.
Average.

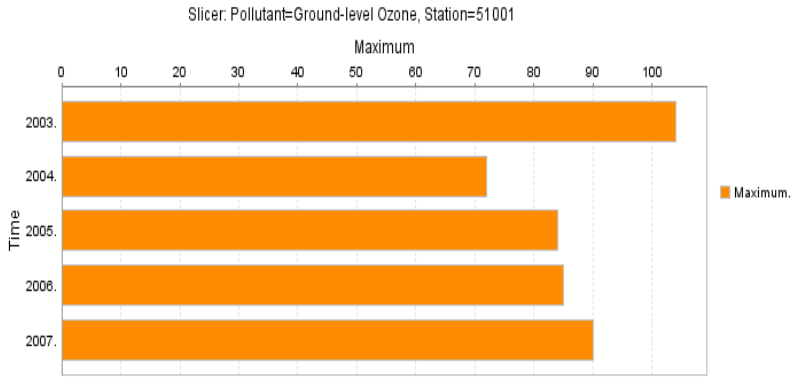

Slicer: Pollutant=Ground-level Ozone, Station=51001
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Time
2007.JUL.WORKINGDAY.2.HIGH.
2007.JUL.WORKINGDAY.3.HIGH.
2007.JUL.WORKINGDAY.4.HIGH.
2007.JUL.WORKINGDAY.5.HIGH.
2007.JUL.WORKINGDAY.6.HIGH.
2007.JUL.WORKINGDAY.9.HIGH.
2007.JUL.WORKINGDAY.10.HIGH.
2007.JUL.WORKINGDAY.11.HIGH.
2007.JUL.WORKINGDAY.12.HIGH.
2007.JUL.WORKINGDAY.13.HIGH.
2007.JUL.WORKINGDAY.16.HIGH.
2007.JUL.WORKINGDAY.17.HIGH.
2007.JUL.WORKINGDAY.18.HIGH.
2007.JUL.WORKINGDAY.19.HIGH.
2007.JUL.WORKINGDAY.20.HIGH.
2007.JUL.WORKINGDAY.23.HIGH.
2007.JUL.WORKINGDAY.24.HIGH.
2007.JUL.WORKINGDAY.25.HIGH.
2007.JUL.WORKINGDAY.26.HIGH.
2007.JUL.WORKINGDAY.27.HIGH.
2007.JUL.WORKINGDAY.30.HIGH.
2007.JUL.WORKINGDAY.31.HIGH.
Maximum.
Average.


Slicer: Pollutant=Ground-level Ozone, Station=51001
Time
Maximum.
Average.

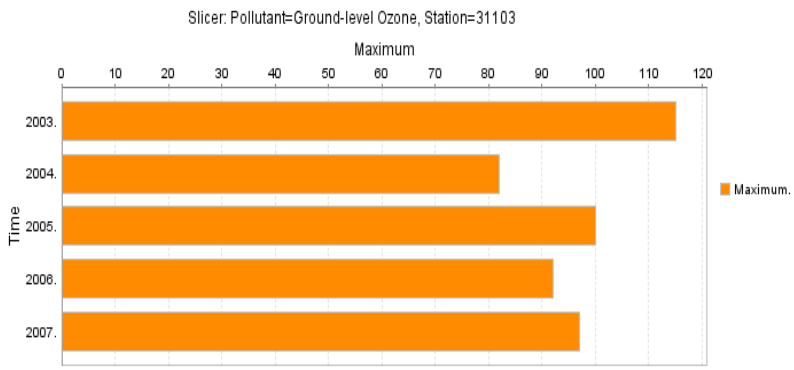

Slicer: Pollutant=Ground-level Ozone, Station=51001
Time
Maximum.
Average.

APPENDIX E

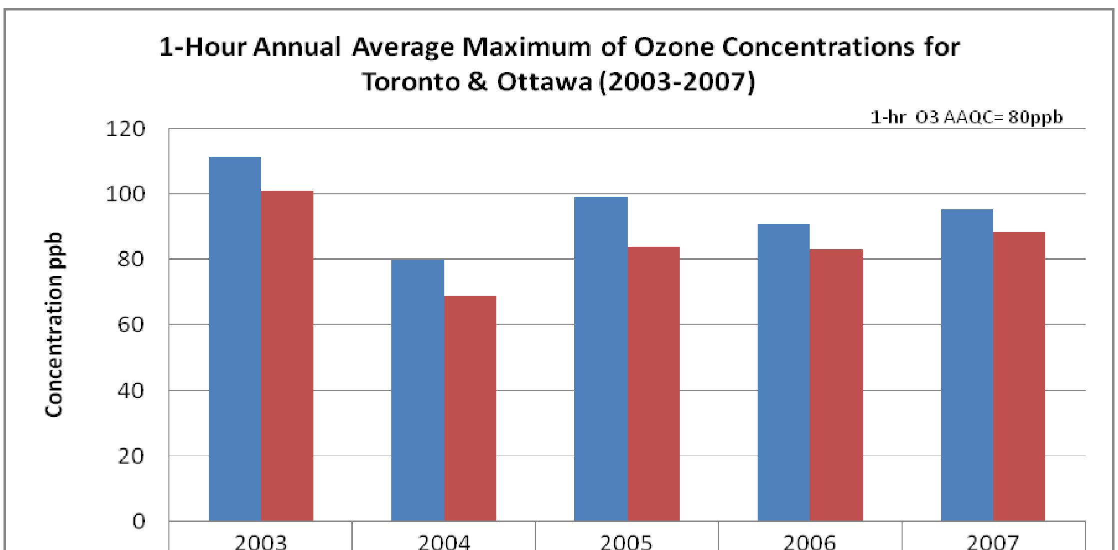
A 5 year (2003-2007) trend analysis for O3 and PM2.5 for the city of Toronto and Ottawa



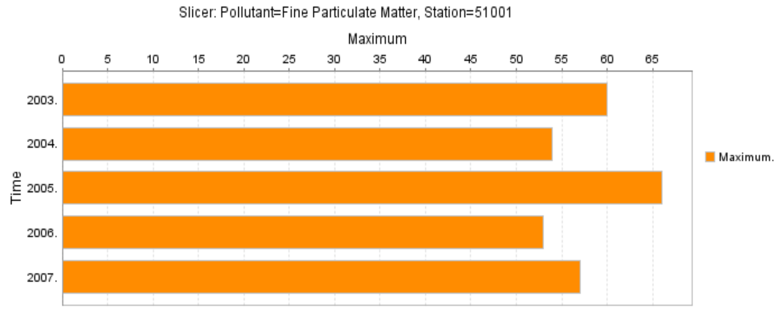
Time	Measures
Year	Maximum
2003	104
2004	72
2005	84
2006	85
2007	90



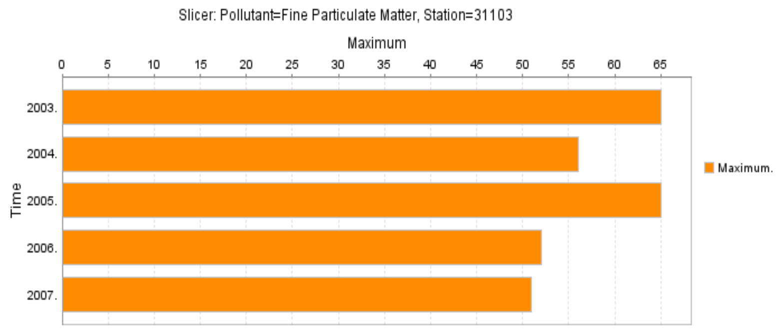
Time	Measures
Year	Maximum
2003	115
2004	82
2005	100
2006	92
2007	97



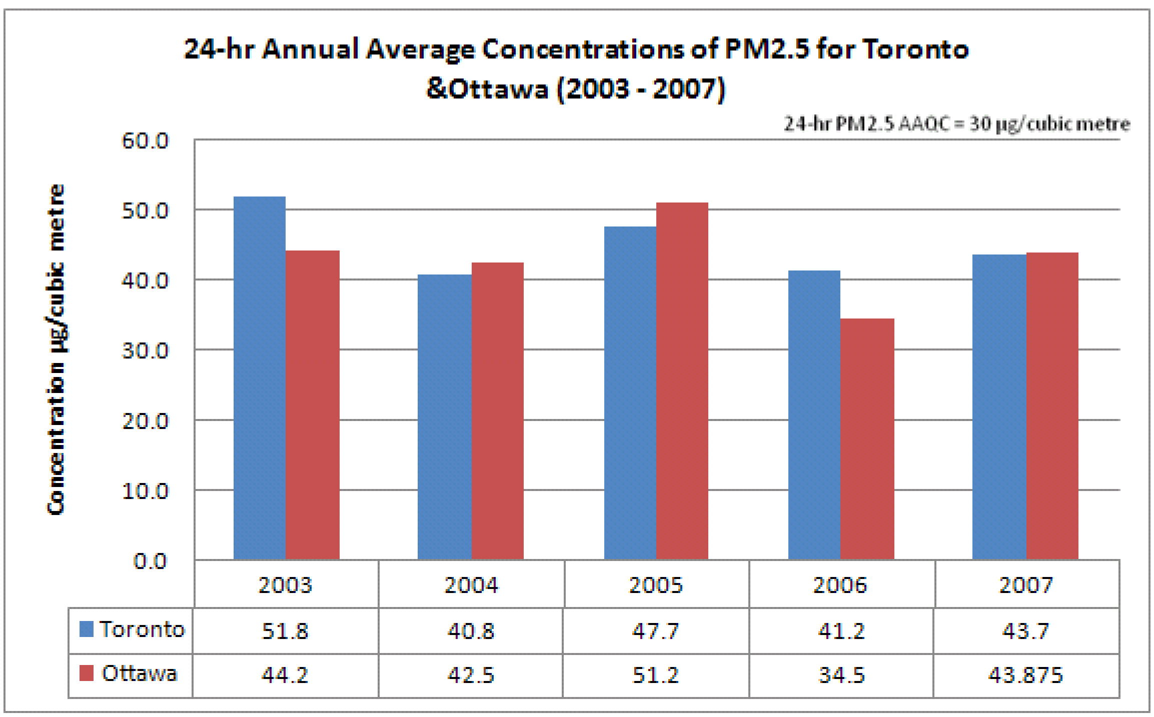
	2003	2004	2005	2006	2007
Toronto	111.5	80	99.5	91	95.5
Ottawa	101	69	84	83.5	88.5



Time	Measures
Year	**Maximum**
2003	60
2004	54
2005	66
2006	53
2007	57

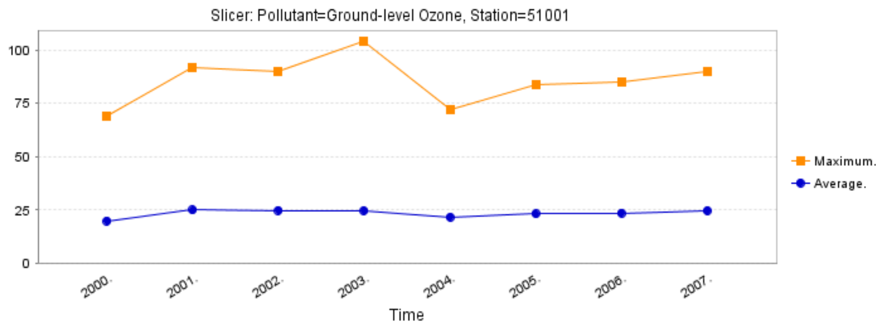


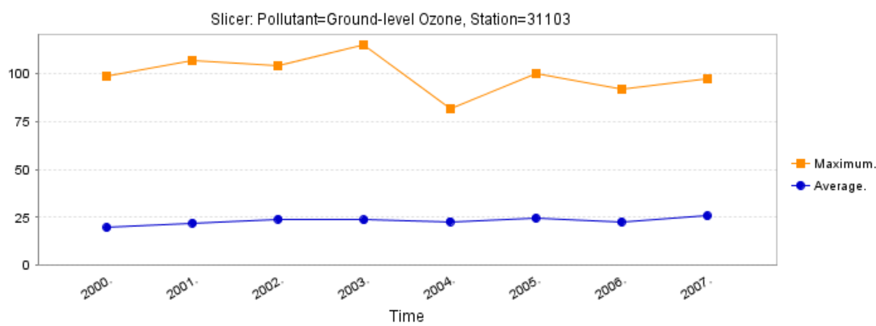
Time	Measures
Year	**Maximum**
2003	65
2004	56
2005	65
2006	52
2007	51

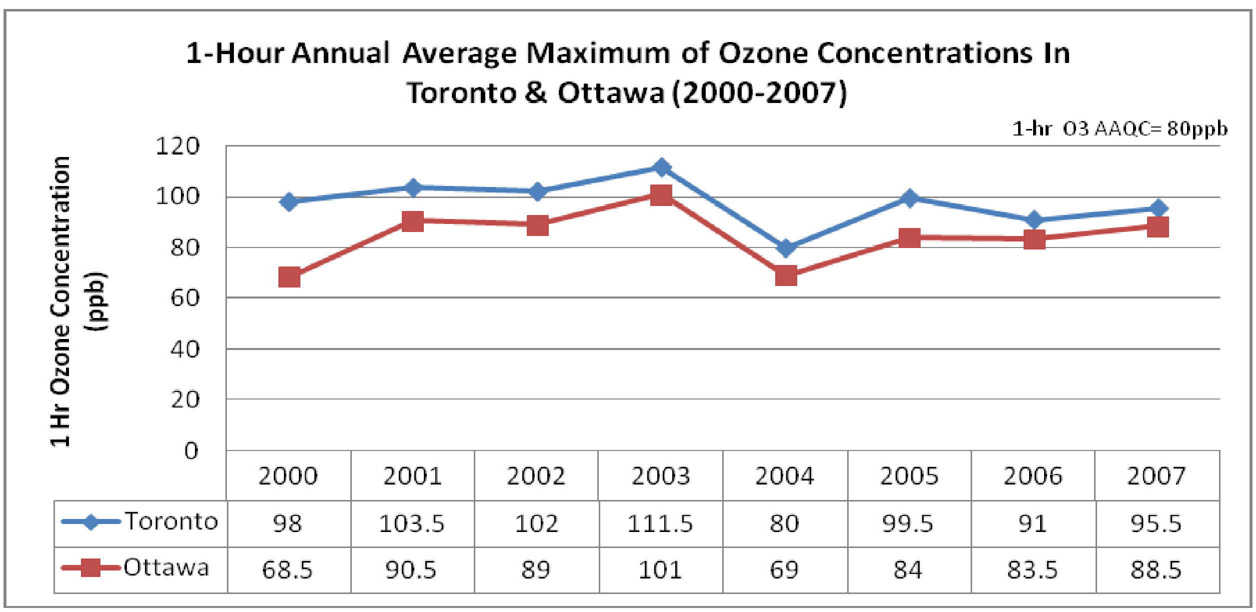


APPENDIX F

Trend analysis of ground level ozone (2000-2007) concentrations for the city of Toronto and Ottawa







	2000	2001	2002	2003	2004	2005	2006	2007
Toronto	98	103.5	102	111.5	80	99.5	91	95.5
Ottawa	68.5	90.5	89	101	69	84	83.5	88.5

APPENDIX G:

The table below presents the Station ID, City, Address, Latitude, Longitude, and Height of instruments for air intake purposes.

Table 2: Station ID, City, Address, Latitude, Longitude, and Height of instruments

Station_ID	Longitude	Latitude	City	Address	Air Intake Height (Meters)
12008	- 83 02 37.21	42 18 56.8	Windsor	467 University Ave	8
12016	- 83 04 23.25	42 17 34.43	Windsor	College/South St	4
13001	- 82 12 29.92	42 24 13.33	Chatham	435 Grand Ave W	15
13021	- 82 13 05	42 14 58	Merlin	Middle Rd/MOE water Pump Stn	3
14064	- 81 24 18.32	42 58 56.23	Sarnia	Front St/CN tracks	3
15020	- 81 44 34.4	43 19 59.1	Grand Bend	Water Treatment Plant	10
15025	- 81 12 23.12	43 00 24.19	London	900 Highbury Ave	4
16015	- 81 09 46.36	42 40 19.5	Port Stanley	43665 Dexter Line	5
18007	- 81 32 58.96	44 18 52.1	Tiverton	Concession Rd 2	4
21005	- 80 17 33.52	43 08 18.97	Brantford	324 Grand River Ave	5
22071	- 80 15 50	42 51 08	Simcoe	HWY 3 Blue Line Rd	4
26060	- 80 30 13.7	43 26 37.8	Kitchener	West Ave/Homewood	5
27067	- 79 14 5.13	43 09.36 16	St Catharines	Argyle Cres	4
28028	- 80 15 51.0	43 33 05.8	Guelph	Exhbition St	4
29000	- 79 51 42.04	43 15 27.99	Hamilton	Elgin/Kelly	4
29114	- 79 51 44.1	43 13 45.31	Hamilton	Vickers Rd/E.18 St	3
29118	- 79 54 27.85	43 15 26.81	Hamilton	Main St W/HWY 403	3
31103	- 79 23 17.24	43 39 46.72	Toronto	Bay/Wellesley St	10
33003	- 79 16 26.61	43 44 52.52	Toronto	Kennedy/Lawrence	4
34020	- 79 25 03.80	43 46 53.76	Toronto	Hendon/Yonge St	5
35125	- 79 35 18	43 38 38	Toronto	125 Resources Rd	8
44008	- 79 48 09.52	43 18 54.37	Burlington	HWY2 NorthShore Blvd E	5
44017	- 79 49 08.2	43 29 12.9	Oakville	8th line GlenashtonDr/Halton Reservoir	12
45025	- 78 51 03	43 53 31	Oshawa	2200 Simcoe St N./Durham College	7
46089	- 79 46 51.27	43 41 55.47	Brampton	525 Main St N/Peel Manor	5
46109	- 79 39 09	43 36 57	Mississauga	Frank McKechnie Ctr	10
47045	- 79 42 8.32	44 22 56.47	Barrie	83 Perry st	5
48002	- 79 15 58	43 57 52	Stouffville	HWY47/ E. of HWY 48	5
48006	- 79 28 59.65	44 02 39.49	Newmarket	Eagle St/ Mc Caffrey Rd	5
49005	- 80 02 17.37	45 20 16.34	Parry Sound	7 Bay St	5
49010	- 78 55 58.64	45 13 27.41	Dorset	HWY 117/Paint Lake Rd	3
51001	- 75 40 33.6	45 26 03.59	Ottawa	Rideau/Wurtemburg St	4
51002	- 75 42 51.05	45 22 57.08	Ottawa	960 Carling Ave	5
51010	- 77 26 28.27	45 59 48.18	Petawawa	Petawawa Research forest facility	6
52022	- 76 31 41.94	44 12 58.54	Kingston	752 King St W	13
54012	- 77 23 43.8	44 09 01.9	Belleville	2 Sidney St/Water Treatment Plant	10
56010	- 75 11 23.80	44 53 59.09	Morrisburg	County Rd 2/Morrisburg Water tower	5
56051	- 74 44 6.84	45 01 4.67	Cornwall	Bedford/Third St	4
59006	- 78 20 46.35	44 18 06.88	Peterborough	10 Hospital Dr	10
63203	- 89 17 24.56	48 22 45.75	Thunder Bay	421 James St N	15
71078	- 84 18 35.67	46 31 59.53	Sault Ste Marie	Sault College	8
75010	- 79 26 57.43	46 19 23.52	North Bay	Chippawa St/Dept National Defence	4
77219	- 80 57 46.62	- 46 28 32.48	Sudbury	Ramsey Lake Rd	3

REFERENCES:

BOOKS:

Giovinazzo, W., 2000. *Object-Oriented Data Warehouse Design: Building a Star Schema.* Prentice Hall.

Godish, T. 2004. Air Quality. 4th ed. Chemical Rubber Company Press, Taylor & Francis Group.

Jarke, M., Lenzerini.M, Vassiliou.Y, Vassiliadis.P, 2003. *Fundamentals of Data warehousing*, 2nd ed.New York, Springer-Verlag Berlin Heidelberg

Kimball, R, Reeves.L, Ross.M, Thornthwaite.W, 1998. *The Data warehouse Lifecycle Toolkit,* 1st ed Wiley

Moussiopoulos, N.ed., 2003. *Air Quality in Cities.* 1st ed.New York, Springer-Verlag Berlin Heidelberg

Qin.Y, Tonnesen.G.S, Wang.Z, 2004. *One hour and eight hour average zone in the California South Coast air quality management district: trends in peak values and sensitivity to precursors.* Atmospheric Environment 38 (2004) 2197-2207.

Rob, P. & Coronel, C., 2009 *Database Systems: Design Implementation and Management.* 8th ed. Boston Massachusetts: Thompson course Technology

Schwartz, J & Hayward, S. 2007. *Air Quality in America: A Dose of Reality on Air Pollution Levels, Trends and Health.* American Enterprise Institute for Public Policy Research

Singh, H.,1999 *Interactive Data Warehousing*, Prentice Hall

Shen.L., 2005. *Internet GIS for Air Quality Information Service for Dalian China*. Masters University of Waterloo, Ontario, Canada

World Health Organization, 1999.*Monitoring Ambient Air Quality for Health Impact Assessment,* European Series No 85, World Health Organization Regional Office for Europe, World Health Organization Press

World Health Organization, 2005 *Air Quality Guidelines Global Update* Particulate Matter, ozone, nitrogen dioxide and sulfur dioxide. Europe, World Health Organization Press

INTERNET RESOURCES:

Apache Tomcat, 2008, Version 6.0.16 [Online] Available at http://tomcat.apache.org/download-60.cgi [Accessed at 27 January 2009]

Environment Canada, 2004. *Criteria Air Contaminants and Related Pollutants*[Online] Available at http://www.ec.gc.ca/cleanair-airpur/Pollutants-WSBCC0B44A-1_En.htm] (Updated 18 August 2006) [Accessed 29 March 2009]

Ever green Foundation, 2000.Schools need nature. [Online] Available at : http://www.evergreen.ca/en/programs/schools/index.sn [Accessed 4 February 2010]

Geodatabase, Data Storage and Management for GIS,1995, [Online] Available at http://www.esri.com/software/arcgis/geodatabase/index.html [Accessed at 15 June 2009]

GeoKettle, Open Source Geospatial business Intelligence, 2010, [Online] Available at http://www.spatialytics.org/projects/geokettle/ [Accessed at 5 January 2010]

Health Canada, *Environmental Contaminants*, 2008 [Online] (Updated 28 August 2009) Available at http://www.hc-sc.gc.ca/ewh-semt/contaminants/index-eng.php [Accessed at 3 September 2009]

Heart & Stroke Foundation's Annual Report on Canadian's Health, 2008. [Online] Available at http://www.heartandstroke.com/site/apps/nlnet/content2.aspx?c=ikIQLcMWJtE&b=4955951&ct=4974269 [Accessed 17 March 2009]

My SQL, 2009, About MySQL, [Online] Available at http://www.mysql.com/about/ [Accessed at 27 January 2009]

Ontario Ministry of Environment, 2006, Clarkson Airshed Study, A scientific approach to improving air quality, [Online] Available at http://www.ene.gov.on.ca/envision/techdocs/6031e.pdf [Accessed 15 January 2010]

Ontario Ministry of environment, 2010, Historical Air Quality Pollutant Data [Online] Available at http://www.airqualityontario.com/history/ [Accessed 15 January 2009]

Statistics Canada, 2009, Population growth is higher in most provinces, territories, [Online], (Updated at 22 September 2009) Available at http://www12.statcan.ca/census-recensement/2006/as-sa/97-550/p4-eng.cfm [Accessed at 3October 2009]

The world fact book, 2007, *Canada* [Online] Available at https://www.cia.gov/library/publications/the-world-factbook/geos/ca.html [Accessed 27 February 2009]

Weather Network, 2009, Weather and your home, The climate of Ontario [Online] Available at http://www.theweathernetwork.com/index.php?product=wxhome&placecode=caon0728&pagecontent=on [Accessed at 1February 2009]

Ministry of Environment, Air Dispersion Modelling Guideline for Ontario , 2009, [Online] Available at http://www.ene.gov.on.ca/en/air/ministry/index.php#em [Accessed at 28 March 2009]

Ontario Medical Association, 2001. *Ontario's Air: Years of Stagnation.* [Online] Available at http://www1.oma.org/phealth/report/airq001.pdf [Accessed 6 February 2009]

Master Thesis in Geographical Information Science (LUMA-GIS)

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland. (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research. (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing. (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia. (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010)

8: *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010)

9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010)