Atmospheric Environmental Information - An overview with Canadian examples

A.C. McMillan, D. MacIver and W. B. Sukloff
Atmospheric Environment Service
4905 Dufferin St., North York,
Ontario, Canada, M3H 5T4.
e-mail: Ann.McMillan@ec.gc.ca

Abstract

The study of meteorology has been revolutionized by modern computing. In this paper, an overview of the handling of meteorological, climatological and air quality data is provided. Some suggestions are given as to a framework for designing systems to handle atmospheric environmental data. Several examples of Canadian systems that are in use now are presented, some of which illustrate application of the framework.

Keywords

Meteorology, air quality data, data management, quality control, Research Data Management and Quality Control (RDMQ) System, CAPMoN, NAPs, NAtChem.

Environmental Software Systems Vol.2 R. Denzer, D.A. Swayne & G. Schimak (Eds.)

1 INTRODUCTION

The collection and management of environmental data has traditionally been done by the scientist collecting the data, who has often had no training in informatics, and has little time, energy or resources to invest in designing a robust, efficient system for data handling.

In this paper, an update on meteorological informatics in Canada will be provided. Then, the development of a system for handling air quality data will be presented to illustrate a process by which a framework for information can be created.

2 METEOROLOGICAL-CLIMATOLOGICAL DATA

It has long been recognized that the weather is a global phenomenon. Hence, some of the earliest international data exchange activities took place in the context of the World Meteorlogical Organization (WMO). The Canadian Weather Service, dating back almost to Confederation, has just celebrated its 125 birthday. Progress in weather prediction has been closely tied to informatics progress.

There are five major networks for atmospheric observations in Canada (Canadian *ad hoc* GCOS Committee, 1995):

- daily climatological observing network;
- hourly observing network both manned and automatic including synoptic observations;
- supplementary programs conducted at hourly observing and daily climatological stations;
- upper air observing network;
- marine programs.

The Canadian Meteorological Center (CMC), located in Dorval, near Montreal, obtains observations from around the world and assembles them into global and regional forecasts using numerical weather prediction techniques. Currently, this largest super computer center in Canada, uses NEC technology. Recent summaries of the computing facility and communications approach can be found in Dansereau, 1996 and Gagne, 1996. Part of the challenge of weather prediction is the need to have information products ready in real time.

Canada also has a program for observing and archiving climate data, for which real time means days, months or years rather than hours. For example, the collection of climate data is done by a variety of partners, including volunteers, who agree to follow the Atmospheric Environment Service (AES) Guidelines for Co-operative Climatological Autostations, 1992. Acceptable data loggers, sensors, calibration and quality assurance standards are provided in this report. The data may be submitted to AES for archiving on a variety of electronic media and disseminated to a multitude of user groups in the private and public sectors.

The high quality of the national climate data base and the subsequent development of different application models has helped protect the health and safety of Canadians. Building codes for wind and snow loadings; drought risk and irrigation models; climate change impact studies; biodiversity stress and prediction models are some of the ways that climate data has been used for socio-economic benefits. The development of Geographic Information Systems (GIS) further advanced climate research and interpretation by layering together different thematic surfaces. Climate process models that describe the general circulation of the atmosphere have used climate data bases for spatial verification.

The climate observing network relies on the systematic observations of trained volunteers. Taken twice daily each day, volunteer data is shared for environmental research, prediction and scenario development. Weather and climate are common topics of conversation at the kitchen table, over the backyard fence, or through the newspaper and radio. Record snowfall, low temperatures, road weather conditions, tornado warnings are all discussed. Amateur radio operators across Canada act as "eyes and ears" of Environment Canada during severe weather. Advances in electronic climate station technology provide a new opportunity to expand into remote areas, such as the climate network supporting the Forest Biodiversity Network in Canada.

3 DATA MANAGEMENT AND QUALITY CONTROL OF AIR QUALITY DATA

As new environmental measurement networks have been implemented, the AES has recognized that a consolidated approach is needed for managing and quality controlling the data. The Research Data Management and Quality Control (RDMQ) system is a software tool created for this purpose. RDMQ was designed to manage and quality control (QC) data from diverse environmental measurements taken in the laboratory and the field. The term "focus" is used in RDMQ to identify separate groups of data. Current focuses include:

- Acid precipitation data from the Canadian Air and Precipitation Monitoring Network (CAPMoN).
- Tropospheric Ozone data from CAPMoN.
- Toxic chemical concentrations in air and precipitation measured in the Integrated Atmospheric Deposition Network (IADN), a network of sites around the Great Lakes.

Each focus has its own data customized dictionary and set of QC checks. RDMQ presents a common set of tools in which the data dictionary and QC checks are applied. The guiding principles in RDMQ are as follows:

- Every datum has an associated flag describing its quality.
- Every record is identified by a focus name, station ID, record type, and the date and time the sample was measured.
- Quality control checks are self-documenting.
- Quality control checks are automated, requiring scientists to evaluate only the data anomalies.

Prior to the development of RDMQ, QC was performed manually in an ad hoc manner. Experience showed that it was not sufficient to apply QC procedures "by eye". Data visualization techniques such as time series plots improved the process, but data anomalies continued to go unnoticed. Another problem occurred when personnel changed, resulting in different interpretations of the visualization plots and data tables. We concluded that the only way to perform QC in a consistent and complete manner was through the use of a computer program. In developing the QC computer program, we found that a mental script is often followed by the scientist. This script provided the basis for the QC coding. In fact, we have not encountered a manual QC check that could not be programmed. Visualization techniques are still used in RDMQ, but they are only used as a final quality assurance check. Any insight into the data discovered by using visualization techniques are coded and added to the automated checks performed by RDMQ.

Creation of a new focus requires scientists to define the measurement variables and QC checks. Scientists using RDMQ can start from QC checks created in other focuses. RDMQ is inherently a collaborative tool resulting in a reduction in the time required to implement a new focus. The user-friendly graphical user interface (GUI) provides a consistent presentation of the data across measurement systems. This dramatically shortens the learning curve for implementing new focuses. RDMQ allows scientists to record the causes and suggested corrective action for each of the QC checks. This "expert system" feature is easily accessible when evaluating QC checks, allowing scientists the option of handing over the task of QC to technicians. AES has four contractors who are familiar with RDMQ. The result is that scientists implementing new focuses do not have to pay for time spent learning the data management and QC tool.

3.1 IMPLEMENTING NEW FOCUSES IN RDMQ.

The heart of RDMQ lies in its QC checks. These allow scientists to concentrate their efforts on data anomalies, reducing the time required to QC the data. In order to produce a good set of QC checks, scientists must have sufficient information about the circumstances in which the measurements were taken including comments originating from the field and from the laboratory. Codification of comments provides this information in a way that is usable by a computer program. Rare occurrences can still be documented using free-format textual fields. RDMQ also provides a mechanism for recording metadata such as major station events, station location and siting information. We are moving towards using the Internet and the Intranet for documentation. The RDMQ manual is now on a web site with URL HTTP://airquality.tor.ec.gc.ca/rdmq. RDMQ can produce HTML tables documenting a focus' data dictionary, QC checks, and station event logs.

RDMQ is written entirely in the computer software package "SAS", from Cary, North Carolina. SAS can produce all of its reports as HTML web pages and has a complete suite of statistical and visualization tools that can work directly on the RDMQ datasets. SAS and RDMQ run identically on both Unix RISC processors and personal computers, allowing scientists to select the most appropriate platform based on data volume, hardware availability, cost, and support infrastructure.

3.2 NAtChem

Once data have been quality controlled using the RDMQ system, the data are transferred to the National Atmospheric Chemistry Data Base System (NAtChem), an integrated database and analysis facility for diverse environmental data. The current data contributors to NAtChem are shown in Table 1. Our experience in setting up NAtChem for acid precipitation measurements revealed that a large percentage of time is spent on translating data into a standardized structure. Because much of this work is already performed in RDMQ, data originating from RDMQ can be rapidly loaded into NAtChem. Using RDMQ forces this work to be done at the front end, reducing the time required to load data into an integrated database.

Table 1: Contributors to NAtChem (Ro, et al, 1995)

CAPMoN	Canadian Air and Precipitation Monitoring Network (Environment Canada)
APIOS-C	Acidic Precipitation in Ontario Study - Cumulative Network (Ontario Ministry of the Environment and Energy)
APIOS-D	Acidic Precipitation in Ontario Study - Daily Network (Ontario Ministry of the Environment and Energy)
GLPN	Great Lakes Precipitation Network (Environment Canada)
REPQ	Reseau d'enchantillonnage des precipitations du Quebec (Ministere de l'Environnement det de la Faune du Quebec)
NBPMN	New Brunswick Precipitation Monitoring Network (New Brunswick Department of Municipal Affairs and the Environment)
NSPSN	Nova Scotia Precipitation Study Network (Nova Scotia Department of the Environment)
NEPMoN	Newfoundland Acid Precipitation Monitoring Network (Newfoundland Department of Environment)
MNPC	Manitoba Network for Precipitation Collection (Manitoba Environment)
PQMPA	Precipitation Quality Monitoring Program in Alberta (Alberta Environment)
BCPCSN	British Columbia Precipitation Chemistry Sampling Network (British Columbia Ministry of the Environment, Lands and Parks)
NADP/NTN	National Atmospheric Deposition Program (NRSP-3) / National Trends Network (United States NADP/NTN Coordination Office)
CASTNET	Clean Air Status and Trends Network (United States Environmental Protection Agency)

4 ADAPTATION AND MITIGATION DATA

Much environmental work has the objective of understanding environmental processes or managing the environment for some purpose. In order to do this successfully it is often necessary to relate environmental changes to effects on biodiversity and human health. There are special challenges involved in handling observations of biodiversity and other ecosystem indicators, however, increasingly the communities taking these observations are recognizing the need for greater quantification, standardization, systematization and inter-linking of data bases.

Two approaches, the coarse filter and fine filter approaches have proved successful. At the coarse filter level, numerous observation networks and historical data bases exist in most agencies. Few of these data bases can be connected in a scientifically sound manner. The atmosphere is common to all biodiversity point observations and may provide the universality or "glue" to meaningfully interpret and predict environmental change over time and space. At the fine filter level or site level, integrated monitoring systems are designed from the on-set with greater attention to understanding interrelated processes and sharing of the observation schedules, elements and units of measurement. In this latter case, atmospheric measurements are designed to be integrated with comparable biotic, aquatic and terrestrial observations.

5 CONCLUSIONS

This paper has been presented in a rather top down way, which is exactly the reverse of the usual way in which environmental informatics develops. This has been done to make the suggestion to the community that a broader vision of systems being developed for specific applications would make eventual linking and integrating of local systems much easier by encouraging local practitioners to deal with issues such as QC on a local scale, but within certain guidelines to ensure that the products meet compatibility standards in order to be integrable.

Environmental data is collected for a variety of purposes, and these determine what is collected, where, how frequently and with what degree of scientific rigor. Science may require detailed measurements, continuous in time, to resolve small spatial areas, in order to study a particular process. Policy may require long-term trends built up over a regional area to assess, for example, the success of an air quality or climate change management strategy. In using environmental information for purposes other than what it was initially collected for, it is important to consider the appropriate scales of the problem under consideration and to try to match data to that requirement.

The AES experience with dealing with atmospheric environmental data has led to some lessons learned:

- partnerships are very important;
- atmospheric environmental informatics, has trends that go out of style, investing in classics (such as QC and decent documentation) gives the best value over time;
- using top end technology doesn't always result in robust, reliable systems, choose the technology to fit the requirements of the application, not the other way around;
- while environmental systems generally develop from the specific to the general, there are enough examples of large general systems around to guide thinking more broadly at the design stage, leading to a more flexible system.

6 REFERENCES

Atmospheric Environment Service (1987) Documentation for the Digital Archive of Canadian Climatological Data (surface) Identified by Element.

Canadian *ad hoc* GCOS Task Group (1995) The Case for Canadian Contributions to the Global Climate Observing System (GCOS), May 5, 1995.

Climate Information Branch (1992) AES Guidelines for Co-operative Climatological Autostations. Guide 92-1. Version 2, October,1992.

Dansereau, M. (1996) CMC Infrastructure. *The Canadian Meteorological Centre Review,* **62**, Volume 3, number 3.

Gagne, A. (1996) Wide Area Network. *The Canadian Meteorological Centre Review,* **63**, Volume 3, number 3.

Hogue, R. (1996) Characteristics of the data assimilation system, forecast models and operational runs. *The Canadian Meteorological Centre Review*, **54-61**, Volume 3, number 3.

Ro, C., R. Vet, D. Ord, A. Holloway, S. Yum (1995) Acid Precipitation in Eastern North America. *National Atmospheric Chemistry Data Base (NAtChem) - 1992 Annual Report.* June 1, 1995.

Sukloff, W.B., Allan, S. (1995) The Research Data Management and Quality Control User Manual. *RDMQ User Manual, HTTP://airquality.tor.ec.gc.ca/rdmq*

7 BIOGRAPHIES

Ann McMillan leads the Science Assessment and Policy Integration Branch (APAC) of AES. This group's main products are science assessments of major atmospheric issues such as "acidifying emissions", "smog" and "stratospheric ozone". She has degrees in Math, Computer Science, and Engineering. Unlike most Weather Services, AES includes an active research group which not only provides research improving the weather forecast, but also provides in air quality issues.

Don MacIver is APAC's Issue Manager for Biodiversity and Issue Integration. He has diverse experience with agencies seeking to manage natural resources. He has written extensively about biodiversity and its link to the atmosphere, especially from a climate change perspective. Data requirements for performing integrative studies is an area of his present interest.

William (Bill) Sukloff is a Computer Systems Analyst in the Air Quality Research Branch (AQRB) of AES. He specializes in software development for environmental information systems.