

Samuel Gilbert*, Radenko Pavlovic, Hugo Landry, Paul-André Beaulieu, Michael D. Moran

1. INTRODUCTION

VAQUM (Verification of Air Quality Models) is a system developed by the Meteorological Service of Canada, which is part of Environment and Climate Change Canada, to evaluate the performance of air quality model forecasts.

2. HISTORY

Development on VAQUM began in 2008. At the time, model forecasts were evaluated using a set of scripts written in different languages, as well as FORTRAN programs. Data read in by the scripts and compiled programs had to be stored in CSV flat files. A considerable part of the program logic was dedicated to locating the proper input files and reading them, which had nothing to do with actual performance evaluation.

It quickly became obvious that such an approach was error-prone and unmaintainable. To store the data in a structured and reliable way, an SQL database was chosen. Since both observations and forecasts are located in time and space, PostgreSQL with PostGIS was the selected Relational Database Management System with geospatial support.

3. DESIGN GOALS

Since a performance evaluation system was being built from scratch, goals were established.

3.1 Automation

Before VAQUM, producing a complete set of results required a lot of human intervention. The process was time-consuming and error-prone. VAQUM was therefore designed to automate common performance evaluation tasks, such as evaluating the numerical forecasts used as guidance by forecasters to produce official public forecasts.

3.2 Reproducibility

Since performance evaluation is the basis for selecting upgrades to the air quality forecasting system, results must be reproducible. For example, it may sometimes be necessary to see

* *Corresponding author address:* Samuel Gilbert, Meteorological Service of Canada, Environment and Climate Change Canada (ECCC), Government of Canada, Dorval, QC, H9P 1J3; e-mail: samuel.gilbert@canada.ca

how a newer model version performs on a historical case that happened many years in the past. Thus a performance evaluation must be performed in such a way that it is comparable to previous ones.

3.3 High Maintainability

Routine maintenance of the system must be easy and not require an excessive amount of work. The system must be able to easily handle changes such as different observation exchange formats.

3.4 Extensibility

As time goes by, new user requirements will invariably crop up. Adding features and products must be easy and must be possible without requiring changes that might break existing products.

3.5 User Friendliness

To ensure that users can properly use the system, it must be designed in such a way that the learning curve is not too steep.

4. COMPONENTS

4.1 Database

At the very core of the system is a geospatial relational database that stores both the observations and forecasts, as well as meta-information required to support the many products.

Observations are grouped as datasets that users can select when configuring comparisons. There are datasets for observations from near-real-time air quality observation networks such as AirNow. Other datasets contain observations from quality-assured/quality-controlled sources such as the National Air Pollution Surveillance Program (NAPS) or the EPA's Air Quality System (AQS).

Forecasts are loaded into a series, which represents a sequence of model integrations

sharing the same parameters. For example, there is a series containing forecasts from the operational Regional Air Quality Deterministic Prediction System (RAQDPS). Other series contain forecasts produced with development versions of various AQ models.

An up-to-date list of observation stations and their properties is stored in a dedicated table. Since station properties, such as position, change from time to time, snapshots of the station list are kept. When creating a series, a snapshot must be selected. This ensures consistency and enables the comparison of series sharing the same snapshot, even if station information has been updated in the up-to-date list.

Verification regions are stored as geospatial vector objects, with a region being a multipolygon. Geopolitical boundaries are preloaded, but users can create arbitrary verification regions with off-the-shelf GIS applications. When configuring a comparison, users can select any number of regions for which statistics will be computed.

Other tables store computed statistics and other ancillary information such as metropolitan areas, timezones, land use, etc.

4.2 Support Utilities

Some specific tasks require processing to be done outside the database. Purpose-built support utilities fulfil those roles.

One utility reads the forecasts from RPN standard files, an in-house file format used to store model output, and extracts forecasted values interpolated at observation station locations. Those values are loaded into the database and linked to the appropriate series and model integration.

Another noteworthy support utility is an implementation of the bootstrap statistical resampling method. Since that method is computationally intensive, it has been programmed in C++ to use multiple threads. This particular implementation scales very well with multiple CPUs.

4.3 Web Application

All user interaction is done using a Web application.

The configuration module allows the creation of series, loading forecasts and configuring comparisons between compatible series. To allow only authorized users to launch computationally-

intensive tasks, access to this module is subjected to authentication and access control mechanisms.

The application also presents comparison results. This allows easy sharing of results with multiple people within the organization's network as a single hyperlink gives access to all the statistics produced for a given comparison.

Results are displayed using interactive charts. For example, a user can hover the mouse cursor over any data point in a chart to view the actual value.

Web maps enable visualization of geospatial information. Users can navigate to regions of specific interest and zoom to any scale required to grasp spatial patterns. Furthermore, clicking on a station on the maps displays station information and a time series of observed and forecasted values for the selected comparison and pollutant.

5. OPERATING MODES

5.1 Comparisons

This is the most common operating mode. It is used to guide model development by verifying that new model algorithms, parametrizations, or input datasets (such as emissions inventories) yield better results.

5.2 Continuous Performance Monitoring

This mode allows fully automated performance evaluation. For example, after each integration of the RAQDPS, forecasts are loaded into VAQUM. As soon as observations covering the entire forecasted period become available, computation of statistics is performed. Results are presented as times series with points for each integration.

5.3 In-Depth Research

VAQUM allows people with sufficient knowledge of the SQL language to quickly explore observation and forecast datasets. With direct access to the database, one can quickly drill down in the data to identify specific events or prototype new statistical evaluation tools.

Since VAQUM is based on standards established by the Open Geospatial Consortium, it is also very easy to export data to allow further processing with other tools.

6. COMPARISON WORKFLOW

Automated processes at MSC decode near-real-time observations received in various formats from observation providers and insert the decoded values into the database.

Through the Web application, a user can create a new series to hold the forecasts from a development version of the model. The application prompts the user to select a station snapshot and enter other details, such as a comparison name and a short description.

The user can then launch forecast extraction by providing a path where the model output files are stored as well as the model level corresponding to the surface. Because of the potential volume of data being processed, this operation can be quite long. It is therefore executed asynchronously and a notification sent by e-mail informs the user of completion. Forecasts are extracted at the station locations specified in the snapshot associated to the series.

A comparison can be created by selecting two series, a base or reference case and an experimental or test case, which share the same station snapshot. The period is configured by selecting a range of integrations to compare. A default set of regions, observation datasets, and pollutants is automatically configured for the comparison. The user can then launch the statistical computation with a single mouse click.

VAQUM pairs up corresponding forecast hours from all matching integrations from both series. It then finds corresponding observations for each forecast validity date and time. Once the creation of the data triplets (observation, base forecast, test forecast) is complete, statistical computation can begin in earnest.

Geospatial predicates are used to find the intersection of observation stations and verification regions selected for the comparison. This allows isolation of included triplets to produce per region statistics. The bootstrap utility is then used to generate distributions of statistics which can, in turn, be employed by statistical tests to determine if differences between the statistics produced for each series are significant at a given confidence level.

When all of the computing is complete, an e-mail containing a link to the Web application with the comparison's results is sent to the user.

7. PRODUCTS

Through its design, VAQUM avoids pre-

rendering charts and maps. All products are always up to date since they are generated "on the fly" with data contained in the database. Here are a few examples :

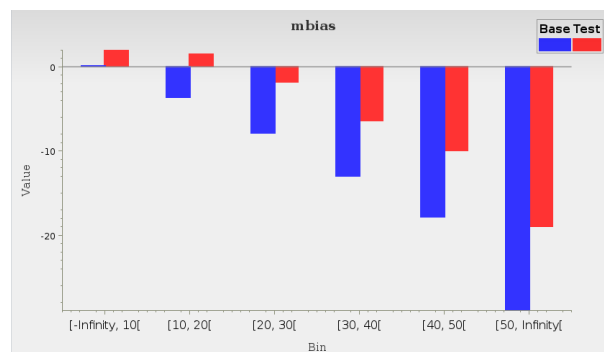


Figure 1: Per-observed-value-bin statistics. Not shown in this figure is the pollutant and region since they are user-selected in the Web application.

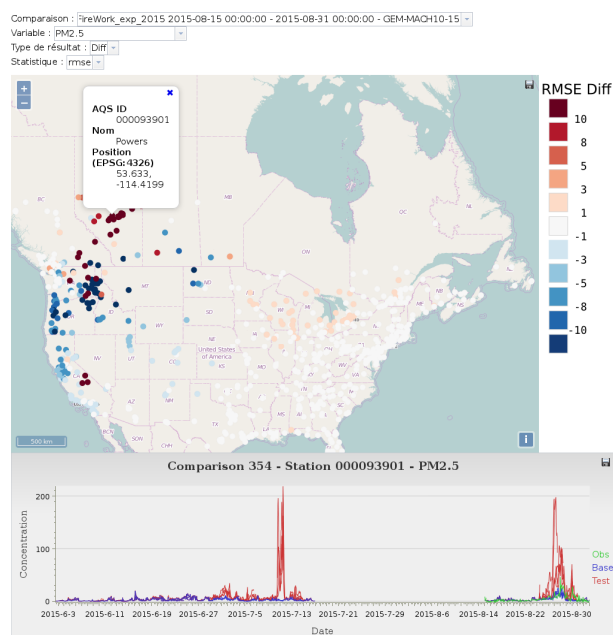


Figure 2: Screen capture of the Web application showing a map view of results. The selected statistic is plotted on the map for each station. Clicking on a station shows a pop-up containing its information and also displays a time series of observations and forecasts spanning the selected comparison

8. CONCLUSION

VAQUM has proven itself to be a valuable tool to guide AQ model improvements. It is also a fully geo-enabled platform that allows the quick and efficient development of new performance evaluation products.