9.9 A NEW AUTOMATED AND INTERACTIVE SYSTEM FOR SOLAR RADIATION DATA PROCESSING AND QUALITY CONTROL

Tara H. Sopoco*, L. Dale Boudreau and Mark Armstrong
Meteorological Service of Canada, Downsview, Ontario, Canada

1. INTRODUCTION

Quality assurance (QA), quality control (QC) and data processing are experiencing a transformation at the Meteorological Service of Canada (MSC). Past practices are no longer effective in meeting demand for quick access to accurate data. Surface solar radiation data is in high demand by the climate monitoring and modeling communities, remote sensing, agricultural and forestry interests and by the engineering sectors in Canada. The system of processing and delivering data is being modernized to meet demand as well as to respond to new network realities.

The MSC has developed a new tool that enables the processing, quality assurance and quality control of solar radiation data from the Canadian solar radiation network. This replaces the previous system, introduced in 1988, which was a landmark program at the time. The old system was based on data being sent in on cassette every two weeks, requiring a large amount of human data handling along the data path from source through to QA/QC and the archive. This manually intensive process, along with budget constraints and reduced human resources for the QA process, have resulted in a backlog of unprocessed data. For the past several years quality control of solar radiation data has mainly been applied for targeted stations and on a request basis. The new system is more efficient, more automatic and takes advantage of current research and technology.

A new data management framework has been accepted for the future of MSC. This framework emphasizes the importance of a single official value for each MSC data element. Also important are the maintenance of data history for all changes, real-time automatic QA/QC, and data storage in a long-term archive where data are reliable and easily accessible in near real-time.

*Corresponding author address: Tara H. Sopoco, National Archives and Data Management Branch, MSC, 4905 Dufferin St., Downsview, ON, CANADA M3H 5T4; email: Tara.Sopoco@ec.gc.ca.

At the same time as the development of the new processing system and the implementation of a new data management framework, a network overhaul is taking place. The network is shifting from just over 40 stations with various surface radiation fields including global irradiance, diffuse irradiance, reflected irradiance, net irradiance and incoming infrared irradiance, to a network of 50 stations focusing on global radiation plus 3-4 CORE stations that will collect global, diffuse, reflected and direct irradiance as well as incoming and outgoing infrared irradiance. The 50 global radiation stations will be in the Reference Climate Station (RCS) network. This is a network of automatic stations with a long period of record and will be observing other elements important for climate monitoring. The CORE stations are manned research sites that collect detailed measurements of atmospheric composition and radiation.

The changes are now being made to the network and should be in place by the end of 2005. The data transfer is automatic; on an hourly basis at RCS stations and on a daily basis at CORE stations. The new processing system takes advantage of the more frequent transmission of data. Data with only minor problems can be processed and available in the archive, with the appropriate flags, within a day of receipt.

2. SYSTEM OVERVIEW

The new system is composed of three main parts: the Domain Layer, the Storage Layer and the Presentation Layer, as seen in Figure 1. The user can interact with the system through the Presentation Layer, using the manual QC graphical user interface (GUI) or the web display. The Domain Layer contains the Java code, built using object oriented analysis and design concepts including common design patterns (Controller, Facade, Singleton). It is in control of the automatic ingest (Ingester), processing (Decoder, Postprocessor, QA'er, Auto QC'er) and database connections (Database Broker). In the Storage Layer, the Oracle 9i database holds the data and the metadata (e.g. station and instrument details, QA/QC flags, error messages) in a cluster environment using Oracle's real application cluster software (RAC). Formerly, the data were held in individual files and directories corresponding to each two-week magnetic tape or ftp transmission.

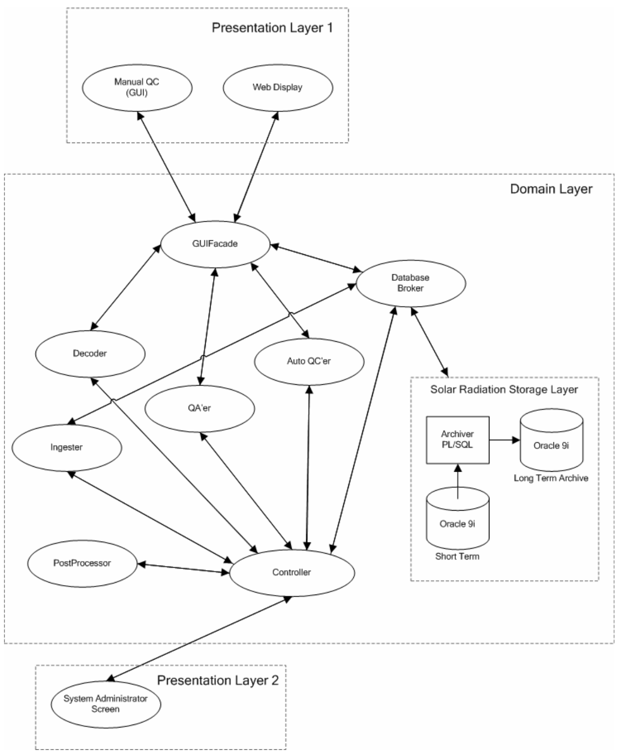

Figure 1. Conceptual system model: view of the high level components and layers. Presentation layer comprised of graphical user interfaces, domain layer with the business logic and the storage layer for the database.

With the new design of the processing system and automatic data transmission, data that are successfully processed are uploaded to the archive and available much more quickly than in the past. Only serious problems with data integrity, such as a missing station ID, stop processing. If automatic QA uncovers erroneous data, they are flagged as such and may be made missing or corrected by automatic QC. The data move into the archive with the appropriate flags, removing the necessity for manual QC before the data are archived. Note that for the purposes of this paper, QA is defined as checking and flagging data, QC is defined as changing data (e.g. correcting or deleting errors).

However, the new system has maintained the capabilities for manual quality control and has advanced features for visual display and efficient targeting of QC. Thumbnail graphs give a quick view of the data. Error reports direct the technician to problems and in the interactive GUI, features have been designed for easy access, efficient information transfer and fast data manipulation.

2.1 AUTOMATIC COMPONENTS

Several new automatic components of the system have been incorporated; from the delivery of data and provision of metadata, to the archiving of data immediately after processing. The efficiency gains and avoidance of delays in the process mean huge time savings and more readily accessible data. There will also be data quality improvements, since problems can be detected in near real-time and station or sensor failures can be corrected on a more timely basis.

The network is currently being modernized from one that used magnetic tapes and paper printouts mailed to the QC office biweekly, to a system with data arriving in hourly bulletins via modem or satellite for most stations and daily ftp transmissions for the remaining. This avoids an average delay in processing of two weeks. After the data arrives and accumulates for a day, processing begins automatically. Previously, significant manual data handling was required and problems with cassette tape sound quality or mail delays were common.

From this point, the decode of the data files occurs, separating the data into elemental form. This is followed by the conversion from millivolts to W/m^2, including temperature compensation for those instruments that require it. In this module, there are also QA checks and corrections for temperature before it is used in the temperature compensation.

For longwave radiation in particular, the accuracy of temperature data is important. In the previous QC system, temperature was only checked against physical extremes. Now, values are compared to monthly normals and against other temperature measurements at the same station or nearby stations. They are also checked for large variations in short time periods (step test). Where necessary, the data can often be estimated using another sensor at the same station or linear interpolation for short periods. If the temperature data is missing or erroneous for 24 hours, the technician receives an email alert so that the problem can be rectified.

After the Converter, offsets and corrections for background noise, power fluctuations and instrument drift are applied to the appropriate channels in the PostProcessor. At this stage there is also a check for incorrect time readings: if night-time data exceeds normal background noise limits for an extended period of time, processing must stop and the technician is informed. In the backlogged data, incorrect time readings were not uncommon, although this problem will nearly disappear in the new network configuration. This is the last stage at which data processing can stall. Data that runs successfully through the PostProcessor will continue to the archive.

QA checks are the next level of processing. These have been greatly enhanced and expanded since the program, rather than the QC technician, is expected to catch the majority of data problems. Physical limit checks are based on the physically possible values on a yearly or monthly basis and detect the most serious errors. Model limits will detect more subtle errors. A clear air curve model is used for global radiation and adapted for diffuse, reflected (with snow cover/albedo also taken into account) and net radiation. For incoming longwave, the model limits are calculated using air temperature.

Across quantities checks are applied between the different radiation fields. Reflected radiation must be a fraction of global radiation, based on albedo. Diffuse radiation and net radiation must not exceed global radiation. Longwave and diffuse irradiance have step tests applied, with the limits derived from the analysis of large data sets. Trend checks have been developed for reflected with global irradiance and net with global irradiance. In these checks, the rise and fall pattern of the data is compared to another element which should respond in a similar pattern, and although small variations are permitted, continued differing patterns in the data signify a problem with one of the sensors.

The system has been designed so that the limits used in the QA checks are not hard-coded, but stored in the database and are therefore able to be adjusted and tuned. As well, the code is modular so that adding or removing QA checks in the future will be a smooth process.

Each QA check sets a flag in the database to signify whether the check was passed, the severity of the error and if the value was above or below the limit. These flags are used to create text messages indicating errors and warnings to the QC technician and also direct the application of automatic QC.

The Auto QC module is designed to correct systematic or simple problems of short duration so that the QC technician can concentrate on the more subtle errors or serious issues of longer duration. Currently, it is restricted to linear interpolation for time periods of up to ten minutes. On a clear day, for global radiation, the clear air model may be used to correct a section of data. Flags are set indicating if the value has been estimated by automatic QC.

After each major step in the processing, the data set is saved to the database, to record a history of the changes made. At the end of automatic QC, the data is also formatted and prepared for the public archive. Both hourly and minutely values are stored with flags indicating the quality of data and estimations. With the minutely data, there will also be QC indicators that will give more information about the problems with data and the solutions that were taken. Although all data with a “Q” (QC’d) status is considered to have successfully passed through an approved MSC QA/QC system, the added information will allow the discriminating data user to select the best data set for their purpose.

This whole process from ingest to auto QC takes a matter of minutes. Therefore, data can be in the national archive within a day of being recorded. However, if there are significant problems with the data that cannot be corrected through automatic QC, the data will not be given “Q” status. The technician will handle these problems through the interactive GUI.

2.2. PRESENTATION LAYER

At this point, no computerized algorithm can replace the full benefit of a skilled human technician for spotting subtle problems and determining the cause or for deciding if suspect data is valid or erroneous. The interactive GUI and web display have been built to effectively use the time that a technician can spend on solar radiation data.

The interactive QC application was created in Macromedia Flash so that it can be delivered through the web, and takes advantage of Flash’s low-bandwidth technology. Because of this, the application can be run from any computer with an Internet connection and the Flash plug-in. It was built to deliver real-time interactive manipulation of graphed data points and direct data editing capabilities. The user manipulates the data in real-time on their desktop, independent of server-side support. Flash is extremely fast, which is important when dealing with large data sets, and database connectivity is integrated using ColdFusion MX.

Figure 2 shows the Week Summary view. This page will open when a station and date are chosen. It gives the QC technician a wide range of information. They can compare the values from each sensor for the week. They can also compare the curve from different sensors. Temperature data is displayed when available. The colored bar above each day indicates the snow conditions (white – snow, green – no snow, grey – unknown). The curve for each day is color-coded using the error flags in the database. The worst error for the day determines the color, with red indicating an erroneous value, yellow indicating suspicious values and green indicating an error-free condition.

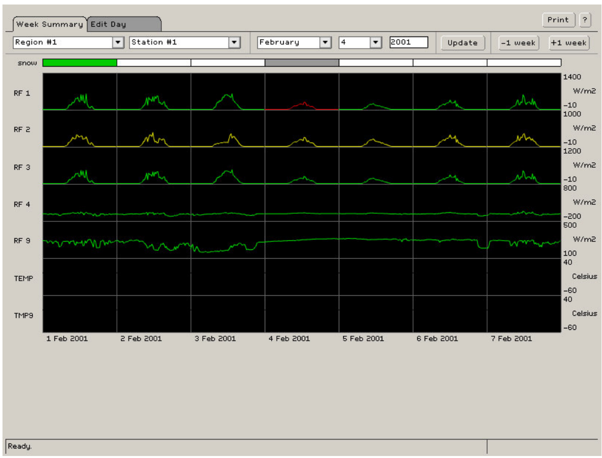

Figure 2: Week Summary in the interactive GUI. Snow conditions are indicated by the bar along the top of each day. The color of the line indicates the severity of errors in the data.

By clicking on a thumbnail, the user is taken to the Edit Day tab, where the data for one day is shown in a full-screen view (see Figure 3). The time is shown along the x-axis and value in W/m^2 on the y-axis. At the top are several options to allow the user to transfer data with the domain layer: Save is used to save a temporary version of the data; QA Check sends the data through the QA module; Finalize is used when the technician has completed work on the day; Raw Data allows the download of a raw data file for editing, before it is processed.

In this view the user has many options on each of the tabs in the lower half of the screen. The Data Options tab has a list box allowing the user to choose multiple channels to display as well as one to choose processing stages. This feature allows the technician to see the transformation of the data after each step in the processing, so that comparisons can be made, e.g. to determine if the corrections made by automatic QC were appropriate. This view also has a small display listing the errors in the display. By clicking on the text for an error or warning, a red cross shows as a marker on the screen to indicate the affected area. The View Data button allows the technician to bring up a spreadsheet-style view of the data where direct data editing can be done.

On the Display Options tab there are various features that can be displayed on the graph. In Figure 3, sunrise and sunset are displayed by the vertical orange line. As well, the mouse position is shown in the upper right-hand corner of the plot. When this option is active, a white cross slides along the curve, indicating the position of the mouse, and the value at that minute. There are also options to display cloud cover amounts, shadows, cleaning/maintenance times, and to plot hourly averages or QA limits.

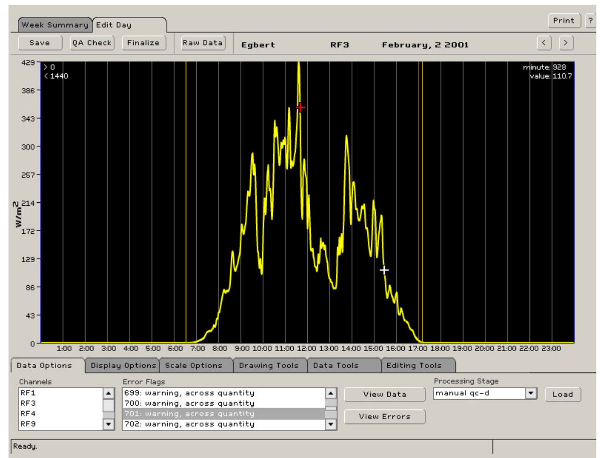

Figure 3. Edit Day view. In this view, the technician has the use of many display options and tools for drawing and editing data, as well as clearing errors.

The Scale Options tab has several preset scale options as well as a custom zoom tool. Drawing Tools has a basic segment tool and a more sophisticated Bezier tool. Data Tools include a feature that allows you to clip part of another curve or model and drop it in to replace a section of the primary curve. It also has tools for the technician to manually complete corrections for the diffuse radiation shade band and for heating caused by shortwave radiation in the pyrgeometer for incoming longwave. Editing tools have various features for directly changing the data, e.g. the time shift feature, which makes it possible to correct for erroneous time stamps in the data.

Figure 4 is an example of how a technician could deal with an error in the data. The view is zoomed in to a section of data. Guides (vertical blue lines) were dragged in to indicate the area the technician would like to edit. The positioning of the guides is made accurate by having them snap to the nearest minute and by displaying the position of the left and right guides in the top left-hand corner of the plot. Here, the segment tool is used to draw a straight line.

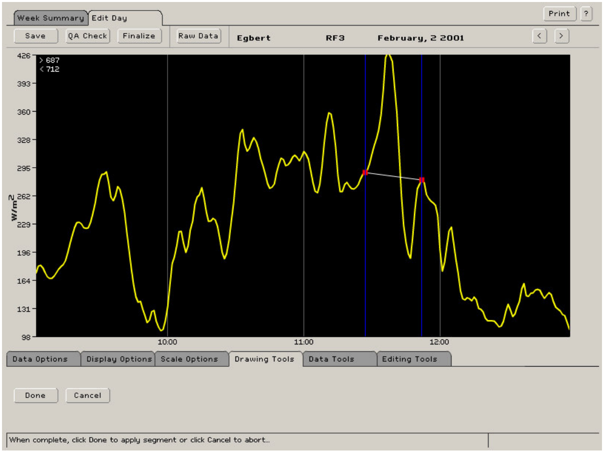

Figure 4. The Drawing Tools Tab. The segment tool and guides (vertical blue lines) are being used to correct a problem indicated by the red cross.

When the technician has made some changes and would like to check if the data now passes the QA checks, a Check button can be clicked to rerun the QA for that particular day. After all the errors from the Data Options tab have been either cleared, corrected or the erroneous data has been removed, Finalize saves the final version of the data and prepares the data for the public archive. This new tool makes the quality control of solar radiation data more efficient, allowing the technician to concentrate on those areas that need attention and providing them with all the necessary information to make the correct decisions when dealing with the data.

There is one other tool that has been designed to help the QC technician and those responsible for the instrumentation to keep abreast of problems. This tool is referred to as the Web Display. For the QC technician, the Web Display can be used to view a long-term picture of a station by displaying thumbnail plots for a month of data. This can help detect problems that develop over time, or recurring problems, such as a shadow that passes over the instrument each day. It can also display all data for a specific station on a particular day that the technician decides to focus on. This tool also has metadata screens where information about the instruments installed at the site, the datalogger program, and the limits used in QA checks can be modified.

This tool is also useful at the regional level. The display can show all the data for the previous day at all stations in a particular region, allowing for performance monitoring by staff responsible for the stations and instruments. The metadata screens are also valuable at the regional level. If a problem is detected in the graphs, the station and sensor information can be checked to ensure that the calibrations have been entered correctly. This feature also encourages better communication and information flow between the QC technician and the region responsible for the station maintenance.

3. SUMMARY

The new processing and quality control system for the MSC solar radiation network is an amalgamation of the latest research on automatic quality assurance and control methods, the latest Internet application technology, and enhancements to features in the original QC system built in 1988. It emphasizes the automatic processing and near real-time data access, but allows for detailed and accurate manual quality control where necessary.

Automatic data processing allows most current solar radiation data to be available within a day. Backlogged data can be dealt with efficiently by loading it into the system for immediate archiving and flagging for subsequent interactive QC. Manual QC is efficiently targeted, while the freedom for the technician to work with any data is maintained. This is a ground-breaking new system, with intensive data manipulation possible over the web. It was designed with maximum flexibility in its modular design and will be a useful tool for quality control to deal with the new network and data management realities.

4. ACKNOWLEDGEMENTS

Our thanks are extended to the team that has helped in the development of the interactive tools: Branden J. Hall of Joshua Davis Studios, Dave Yang of Quantumwave Interactive Inc. and MSC Ontario Region.