3.2 ASSIMILATION OF OBSERVATION DATA INTO SCRIBE

C. Landry, M. Ouellet, R. Parent and R. Verret
Development Branch
Canadian Meteorological Center

1. Introduction

Automated weather forecasts for all regions of Canada have been available for quite some time based on SCRIBE (Verret et al, 1993), a system developed at the Canadian Meteorological Centre (CMC). This system has until now been restricted to using guidance from numerical weather prediction (NWP) models and their derived statistical outputs, namely Perfect Prog (PP) (Klein et, 1959) or updateable Model Output statistics (UMOS) (Wilson and Vallée, 2002). In practice, this means that the generated forecasts are solely based on model data, without any explicit observation data. Therefore, the SCRIBE product generator is totally unaware of recent weather events, and this limitation is particularly acute for weather products that are generated long after the model run. This “blind” effect would generally result in forecasts that are not as up-to-date in their first 18 to 24 hours, were it not for the adjustments made by the operational forecasters. One of the key impetus for the work presented in this paper was to minimize these necessary manual adjustments.

It has been demonstrated in several experiments that most of the modification work that forecasters perform on the SCRIBE guidance is to merge model forecasts with current observations. A system is under development that will merge the SCRIBE concepts with the latest local observations, using a statistical approach. This will eliminate most of the errors in the initial period of the forecasts and is expected to reduce by 50% the time spent by forecasters to quality control the initial parts of the SCRIBE forecasts. The added value to the SCRIBE forecast in the initial period is expected to be of the order of 5-10%. The design of the system must be open enough to ingest radar, satellite and lightning data at a later stage.

2. The SCRIBE System

SCRIBE is an Expert System capable of generating automatically or interactively any type of weather products for a region or a specific locality (see Verret et al, 1997 for more details). The data that feed the system come from a set of matrices which are generated after the 00:00Z and 12:00Z model runs. These matrices contain different types of weather elements such as NWP output, statistical guidance form PP model and UMOS model and other analyses model and climatology data. The time resolution is 3 hours. Once ready these regular matrices are sent to each regional SCRIBE system. Upon arrival, the data are processed by the Concept Generator and are synthesized and downsized to a set of well defined weather elements *concepts*. These concepts, which can be displayed on a graphical interface, are then updated by the forecaster to reflect the latest obaservarions and understanding of the weather situation. Verification as shown that value is added by the forecaster to the initial SCRIBE set of weather concepts but only for the first 24 hours. Beyond this period (day 2-3) little or no value is added to the forecast. From these concepts, the Product Generator can produce either automatically or manually any type of weather products, which can then be sent to the client. Figure 1 shows the main steps in the Scribe data processing.

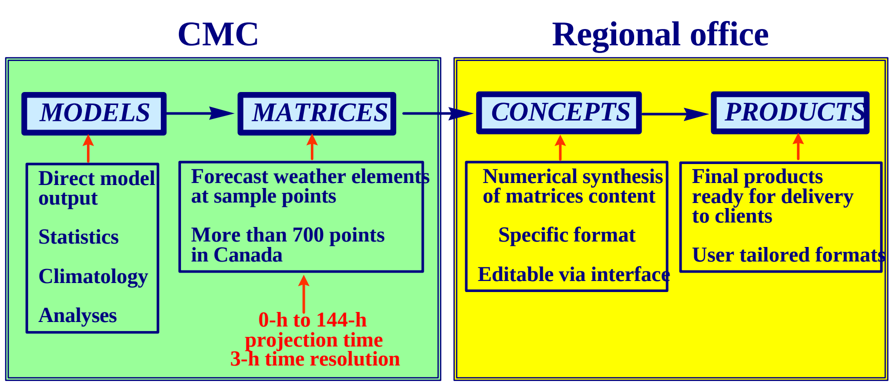
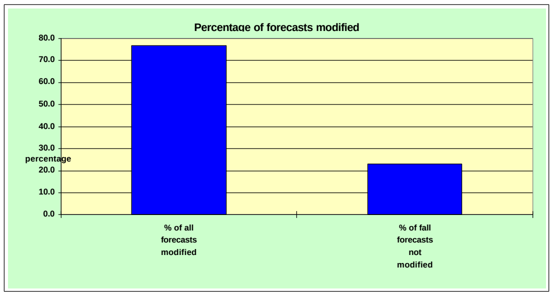


Fig. 1. Basic Scribe System information flow with the different data format.



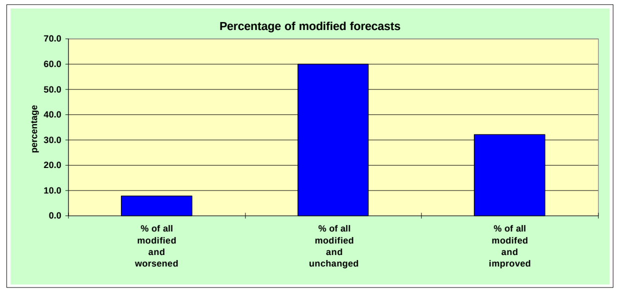


Fig. 2. Verification results for 480 forecasts using the Scribe Observation prototype system. The graphic on the left shows the distribution between the modified and unmodified forecasts. The graphic on the right shows whether the forecasts were improved, worsened or unchanged by the modifications.

3. First Prototype

As a first attempt, a simple method was developed to assimilate the latest observational data into the SCRIBE product generator (Petrucci, et al, 1999). This assimilation subsystem was extracting relevant weather surface observations at each regional observation station. Then, the SCRIBE weather matrices that contain the numerical forecast data was updated with the observation data. Using simple algorithms, the numerical content of the matrices was replaced with the appropriate observed values, when and where available. Further processing of the matrices up to the forecast outputs then proceeded in the regular fashion. An ensemble of procedures was established to update key weather elements, as well as relaxation methods to merge observational data with forecast data beyond the current time. Figure 2 shows subjective verification results over 480 public weather forecasts. 78% of the forecast were changed by the modified SCRIBE system and out of these 32% were improved, 60% were unchanged and 8% showed some deterioration.

Although these preliminary results were encouraging, a deeper analysis of the behavior of this first prototype raised many problems and led us to another approach were observed data should be used to modify weather elements at a later stage in the SCRIBE data processing. Moreover, the use of short-term objective forecasting model (weather element, radar data, and lightning data) is essential to link observed data to the model data.

4. Architecture of the New System

The aim of the present project is to provide observed and nowcasting data to the SCRIBE system on a continuous basis. At any time, the pure model data can be replaced interactively by the latest observations and short range forecast weather elements for the next 9 hours. The time resolution of the observed and nowcasting data is increased to 1 hour. To fulfill this task, three dynamic databases are prepared. The first one will contain the observed weather elements, the second will contain the short-term forecast weather elements and the third one will synthesize and merge the observed and forecast data into a consistant set of continuous weather elements.

Observations Database

Observed data are extracted on an hourly basis and are used in different ways such as bias correction, tendency recognition, and short term persistence. This will provide a direct and positive impact on the beginning of forecast time period. The database will contain data form different sources: METAR, SPECI, Synoptic observations, radar data and lightning data. Up to 300 stations are used to feed the database. A Quality Control Module will be used to remove or replace erroneous data that could deteriorate the forecast. Automatic stations will be processed with specific rules to take into account their related type of data. Another module will remove redundant SPECI that do not add information to the knowledge of the weather situation.

Radar data will be extracted from the dynamic CMC North-American Mosaic Radar Database inside a 15 km radius around the station where radar coverage is available. A similar strategy will be applied to the observed lightning strikes available from the Canadian Lighting Detection Network.

Forecasted Weather Elements Database

A database of forecast weather elements is prepared every hour. It contains the short-term weather element forecasts that provide a link between model and observed data. These data come from different sources. Three-hourly data from the CMC operational regional model are interpolated to a one-hour time resolution.

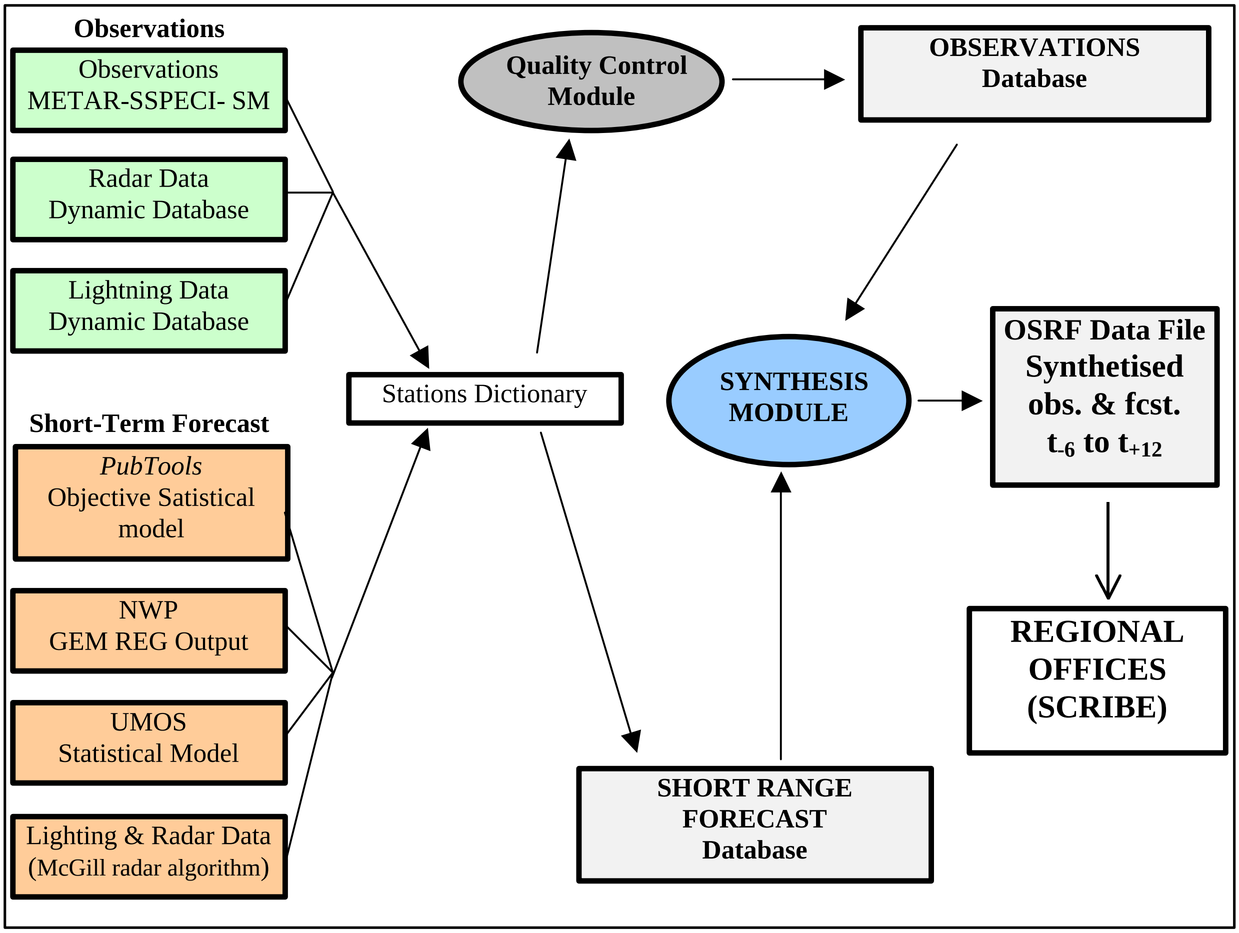


Fig. 3. Architecture of the system: Data processing at CMC to produce the Observation and Short Range Forecast data file.

A short-range statistical model based on Multiple Discriminant Analysis (MDA) provides probabilities of occurrence of weather elements such as cloud cover, visibility, and precipitation (occurrence, types, convective/stratiform) for the short term. Observed Radar data are projected in the short term using extrapolation techniques developed by McGill University. Similar techniques are used to forecast the lighting clusters in the short term.

Synthesized Weather Element Database

The observed and forecast weather elements available in the databases are analyzed and synthetized into a consistent sequence of hourly weather events starting from 6 hours (t_{-6}) before the system cutoff time to 12 hours after (t_{+12}). This is done by an ensemble of rules. The format output is a file containing a set of continuous weather elements at one-hour time resolution. This Observed and Short Range Forecast (OSRF) data file will be sent to regional offices were the SCRIBE system will use it to update the regular SCRIBE concepts. Figure 3 shows the data flow leading to the generation of the OSRF data file and Figure 4 shows an example of the output.

Concepts generator and Data Merger

As for the regular Scribe matrices, the OSRF data file will be converted in the Scribe *Concepts* format. An adapted concept generator, specially designed to manipulate the hourly weather element data, will transform the OSRF files into concepts. The forecaster will be able to update the SCRIBE concepts interactively by merging the OSRF concepts with the regular concepts, display the result on the SCRIBE interface and generate the product text. Figure 5 illustrates this process.

```
STN: CYYC    |SKY|CIG|        PRECIPITATION        |POP|   ACC   | TEMP  |    VENT   |  VIS   |
date      hr |/10|   |pcpn1|pop|pcpn2|pop|pcpn3|pop|   | l  | s  | T | Td| dd| vv|gst|(NM) |Tp|
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 0600| 10|005| -R  |100|     |   |     |   |100| 0.8|    |  4|  4| 33| 18|   | 4.0 |PP|
20020930 0700| 10|006| -R  |100|     |   |     |   |100| 0.8|    |  3|  3| 20| 24|   | 6.0 |PP|
20020930 0800| 10|005| -R  |   |     |   |     |   |100| 0.8|    |  2|  2|360| 18|   | 7.0 |  |
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 0900| 10|005| -R  |100|     |   |     |   |100| 0.8|    |  2|  2|350| 17|   | 6.0 |PP|
20020930 1000| 10|004| -R  | 50| -S  | 50|     |   |100| 0.4| 0.4|  1|  1|339| 14|   | 6.0 |PP|
20020930 1100| 10|004| -S  |100|     |   |     |   |100|    | 0.8|  1|  1|331| 16|   | 5.0 |PP|
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 1200| 10|005| -S  |100|     |   |     |   |100|    | 0.8|  1|  1|340| 21|   | 4.0 |PP|
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 1300|  9|004| -S  |100|     |   |     |   |100|    | 1.3|  1|  1|343| 20|   | 3.0 |PP|
20020930 1400|  9|005| -S  |100|     |   |     |   | 98|    | 1.3|  1|  0|349| 20|   | 3.0 |PP|
20020930 1500|  9|007| -S  | 99|     |   |     |   | 95|    | 1.3|  1|  0|355| 20|   | 3.0 |PP|
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 1600|  9|007| -S  | 95|     |   |     |   | 85|    | 1.0|  2|  1|355| 22|   | 9.0 |  |
20020930 1700|  9|008| -S  | 83|     |   |     |   | 76|    | 1.0|  1| -1|356| 24|   | 3.0 |  |
20020930 1800|  9|009| -S  | 78|  -R | 22|     |   | 68| 0.2| 0.8|  1| -2|356| 26|   | 9.0 |  |
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 1900|  9|010| -SW | 73|  -RW| 27|     |   | 53|    | 0.1|  1| -3|353| 27|   | 9.0 |  |
20020930 2000|  9|010| -SW | 93|     | 07|     |   | 58|    | 0.1|  2| -2|351| 25|   | 9.0 |  |
20020930 2100|  9|010| -RW | 62|  -SW| 38|     |   | 37|    | 0.1|  2| -2|360| 20|   | 9.0 |  |
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----+--+
20020930 2200|  9|150| -RW | 66|     | 34|     |   | 50|    |    |  2| -2|360| 20|   | 9.0 |  |
20020930 2300|  9|400|     |   |     |   |     |   | 21|    |    |  2| -2|360| 20|   | 9.0 |  |
20021001 0000|  9|500|     |   |     |   |     |   | 17|    |    |  2| -2|360| 20|   | 9.0 |  |
-------------+---+---+-----+---+-----+---+-----+---+---+----+----+---+---+---+---+---+-----|--|
```

Fig. 4. Example of Observations & Short Range Forecast (OSRF) data file for Calgary CYYC. Data file starts with 7 hours of synthesized observations, including the current observation, followed by 12 hours of forecasted data.

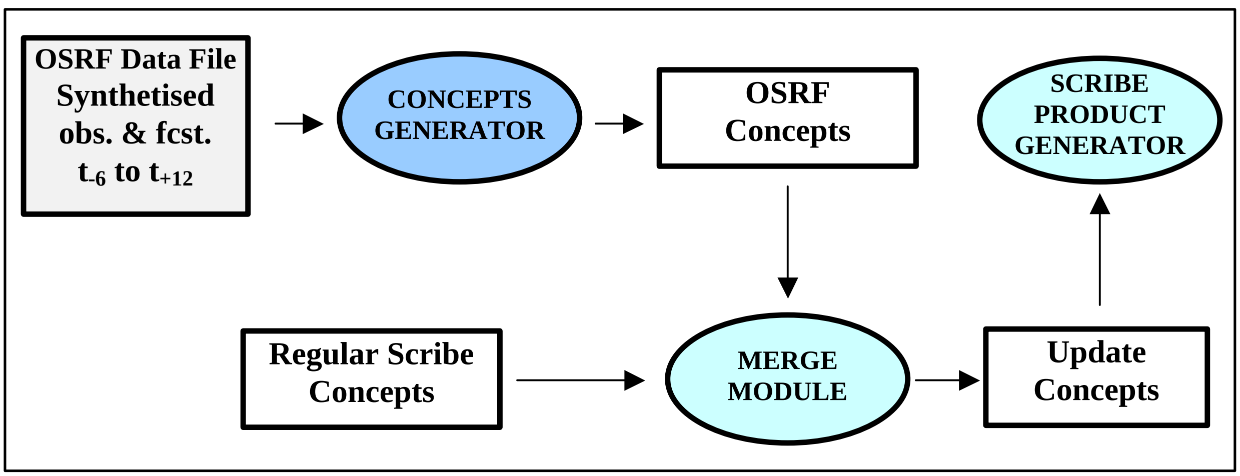


Fig. 5. Architecture of the system: Data processing in Regional Scribe system to update regular set of concepts.

5. Conclusion

The assimilation system is currently under development. A first prototype is expected to be ready in March 2003 for evaluation and testing. The validation and the integration of the nowcasting modeling technique, which are currently under development, will also be ready at that time.

6. Reference

Klein, W. H., B. M. Lewis and I. Enger, 1959: Objective prediction of five-day mean temperature during winter. J. Meteor. 16, 672-682.

Petrucci, F., Rochon, M. and Verret, R. , 1999: *Assimilation of observation data into Scribe.* Preprints, Sixth Workshop on Operational Meteorology. Halifax, Nova Scotia, Canada, 29 November. - 3 December, 1999, pp 77-81

Verret, R., D. Vigneux, J. Marcoux, F. Petrucci, C. Landry, L. Pelletier and G. Hardy, 1997 : *SCRIBE 3.0 A Product Generator*. Preprints, 13th International Conference on Interactive Information and Processing Systems for Meteorology, Oceanography and Hydrology. AMS. Long Beach, California, February 2-7, 1997, pp 392-395.

Wilson, L. J. and M. Vallée, 2002: The Canadian Updateable Model Output Statistics (UMOS) System: design and development tests. Weather and Forecasting, 17, 206-222.