|  |  |  |
| --- | --- | --- |
| WORLD METEOROLOGICAL ORGANIZATIONCOMMISSION FOR BASIC SYSTEMSOPAG on DPFS**MEETING OF THE CBS (DPFS) TASK TEAM ON SURFACE VERIFICATION**Geneva, switzerland20-21 oCTOBER 2014 |  | DPFS/TT-SV/Doc. 6(16.X.2014)\_\_\_\_\_\_\_Agenda item : 6ENGLISH ONLY |

**Data quality control in the EMET verification system at the Canadian Meteorological Centre**

*(Submitted by Tom Robinson)*

##### Summary and purpose of document

This document provides background information on the observational quality control system under development for the CMC EMET verification application

##### Action Proposed

The meeting is invited to take note of the work being done on quality control of surface observations at the CMC, for input to the discussion on surface verification standards.

Observational quality control systems which involve comparison of the observations with model-based fields are problematic insofar as the model is then verified against an observation data-set that is itself influenced by the model being verified. To avoid this problem in the development of the EMET verification system, a model-independent quality control system is under development for the application.

This quality control system for observations ingested by the EMET database is made up of the following points:

* Gross error check
* Statistical thresholds calculated from the history of observations available in the DB
* Identification of abnormal variations in the observed values
* Comparison with neighbouring stations in space and time

The **gross error check**, which consists of checking for meteorologically absurd values, is done prior to ingestion into the database and hence these data are not even considered in the verification system.

The **statistical thresholds** are calculated for all stations in the database within a geographical box of 10 degrees longitude by 2 degrees latitude, on a month by month basis and in 3-hour slices, as well as 1000 m slices in the vertical, in order to reasonably ensure geographical, seasonal and diurnal consistency. The thresholds can be applied with a certain degree of flexibility, with the goal of eliminating bad observations while keeping the rare but nonetheless real ones and they can be updated periodically as needed or as a sufficient amount of new data is ingested.

The thresholds are calculated from statistics of the observations contained in the database and then the observations are compared with this threshold. An observation which passes the threshold is considered “suspect”. The thresholds are calculated with as much precision as possible, taking into account the time of day, in 3-hour increments, the month and geographical position. They are then applied with a certain amount of flexibility, with the goal of eliminating bad observations while keeping the rare but nonetheless real ones. The thresholds are updated periodically.

There is also the possibility of creating exceptions. If a station is systematically penalized because its data does not resemble that of any other stations within the box, a separate box for it alone may be created.

**Abnormal variations** are determined based on the coherence within the time-series of the observed element.

A score is given to the observations based on the statistical thresholds and the abnormal variations. When an observation is ingested into the database, it begins with a score of 100, which is then reduced based on the comparison with the statistical thresholds and its time-series coherence. Once the score reaches a certain value it is considered suspect. At this point, in order to differentiate between a gross error and a rare but real meteorological phenomenon, it undergoes a **comparison with neighbouring stations** within 150 km and within an interval of plus or minus 3 hours. If it fails this comparison, it is then rejected. If there are no other stations within the imposed spatial and temporal limits, then the observation may be accepted or rejected depending on the score from the first two checks.

Note that the neighbour check also takes altitude into account. The difference in altitude between stations is multiplied by a factor of 100 (i.e. a 10m difference = 1 km horizontal distance). In this way, stations which are close in horizontal distance but whose altitudes differ too much are not considered as neighbouring.

To date the quality control has only been used for temperature observations. Work has begun on wind observations, but there are a number of complicating factors in comparison with the temperatures.

Other ideas considered for the quality control system include checking for successive nul variations (check for instrument problems), inter-variable checks (e.g. report of snow with temperature of +10C) and frequency of reports, but to date these have proven to be too complex to incorporate.

**Results**:

The quality control with respect to temperature does not make that much of a difference when verifying against synoptic observations. As figure 1 indicates, the scores do not in general change much at all with or without the QC. This is perhaps not overly surprising as the SYNOP data is of relatively high quality and as the numbers of the observations on the graphs indicate, only about 0.05% of the observations are eliminated.

It is expected that with other variables such as precipitation and possibly winds, a higher rate of rejection is likely. A much more sophisticated quality control in the Canadian Precipitation Analysis (CaPA) rejects significant numbers of observations, particularly in winter, due to known biases of certain widely used precipitation measuring instruments. Due to the shear number of rejects with the CaPA QC, not only the values of the scores but the conclusions to be drawn in comparing two models changes significantly.



 

Figure 1: Mean forecast error of the CMC Global and Regional Deterministic Prediction System surface temperature forecasts scored against the North America SYNOP observation network over the period December 1, 2013 to January 31, 2014. Top: without quality control on the observations; Bottom: with quality control

 

 

Figure 2: Heidke Skill Score of the CMC Global and Regional Deterministic Prediction System 6-hour precipitation accumulation forecasts. Scores are calculated using the observations from the Canadian Precipitation Analysis (CaPA). Top: without quality control; Bottom: with the CaPA quality control. The confidence intervals are calculated from bootstrap on the difference between the two models.

An example of scoring against METAR observations below concerns the verification of Mean Sea-level Pressure forecasts. A series of METAR observations below shows an obviously erroneous value.

|  |  |
| --- | --- |
|  23:00:00  |  1.3  |
|  00:00:00  |  1.25  |
|  01:00:00  |  1.31  |
|  02:00:00  |  1.54  |
|  **03:00:00**  |  **-21.75**  |
|  04:00:00  |  1.56  |
|  05:00:00  |  1.46  |
|  06:00:00  |  1.55  |
|  07:00:00  |  1.48  |
|  08:00:00  |  1.48  |

With no quality control applied, the verification shows an anomaly in the time series of standard deviation of the forecast error, which is then cleaned up when the QC is applied (figure 3).





Figure 3: Standard deviation of the forecast error of Mean Sea-level Pressure forecasts from two versions of the CMC Regional Deterministic Prediction System. Scores are calculated against METAR observations over the south-eastern United States. Top: no quality control on the observations; Bottom: Quality control applied on the observations.