ON-LINE QC AND SPATIAL ANALYSIS OF METEOROLOGICAL DATA

Radim Tolasz
RNDr. Radim TOLASZ, Ph.D.
Czech Hydrometeorological Institute
Na Šabatce 17, Praha 4 – Komořany
+420 596 900 277, tolasz@chmi.cz

Abstract
Continuous automated measurements of meteorological elements allow their online presentation to the public and to state security agencies. National Meteorological and Hydrological Services (NMHSs) are today responsible for providing quality and timely information on the meteorological situation in a comprehensible form. The basic time step of meteorological measurements in the Czech Hydrometeorological Institute is a 10-minute interval, and overview graphs of selected values in a 48-hour time window are presented at the same interval. Possible occurrence of individual missing values is interpolated by polynomial interpolation before the presentation, outages longer than 40 minutes are interpolated by the IDW interpolation method using surrounding meteorological stations. However, the length of IDW interpolation is limited to a maximum of 24 hours. IDW interpolation is defined for each station and element, and the authorized administrator of the application can modify the interpolation parameters (change the radius, add or remove surrounding stations). The CLIDATA/GIS application provides basic temperature maps (hourly, average daily, maximum daily and minimum daily), rainfall (daily sum, snow depth and total snow depth) and daily sum of global radiation. The data sources (10M data) are checked on a daily basis in the CLIDATA application using prepared outputs - adding missing and checking the interpolated values, checking defined limits and checking the element’s dependencies. Errors and outages of automated measurements arise in the event of telecommunication transmission failures or voltage fluctuations. After the check, the outputs (graphs and maps) are automatically updated.

Keywords: Data quality control, Meteorological data, Spatial analysis, Polynomial interpolation, IDW interpolation, Data presentation

AUTOMATISATION OF METEOROLOGICAL MEASUREMENT

Automatization of meteorological stations has been increasing regular data capacity. Nowadays the Czech Hydrometeorological Institute (CHMI) takes care about 350 standard automatic climatological stations and more than 100 supplementary climatological stations. These stations produce 10 minutes data that is importing to the CLIDATA database continuously. Except of synoptic (professional) and voluntary automatic meteorological and climatological stations there are a lot of data from hydrological stations and air quality control stations, stations of other companies (water managements and road managements authorities) as well as data from foreign meteorological stations. Despite reliable stations operation there are some errors during failure of whole stations or some sensors, plugging of rain gauges or during data transmission. Ten minutes data (10M) is not possible to control by methods determined for climatological data (usually twice or three times a day). Therefore we tried to find methods and algorithms for checking of regular data. The main target is to detect and repair or complete wrong or suspicious data as quickly and effectively as possible.

QUALITY CONTROL OF REGULAR 10M DATA

Methods used for regular on-line QC in CLIDATA are described together by theoretical background of these methods by Repka, et al (2017). The regular QC application is made analysis of 10M data and detected wrong, suspicious or missing data using the predefined algorithms. User has to manage the 9 groups of detected values (see Fig. 1, left part). Each corrected or complemented value is marked by standard flag. Flag means how was the value corrected (estimation, calculation, interpolation).
Examples of QC information

One part of application presents the differences between consecutive intervals. Meteorological elements are controlled according to defined limit values by comparing values of two consecutive (following) 10M intervals. If the difference is higher than limit (for example 2 hPa for air pressure, 3 °C for air temperature or 1 °C for soil temperature) the system informs the user (Fig. 2). User have to correct or delete the values, or (it is not case) may label values as OK. Values are displayed in the table and also as graph (more elements) and user can directly rewrite correct or estimated value and store it in the database with particular flag.

Another type of error detection is prepared according elements dependences (for example maximum temperature is less than minimum temperature or wind speed without wind direction) or as comparison with historical extremes for individual station. The special reason for error is on the Fig. 3 – the values are interpolated (see next chapter).
INTERPOLATION OF MISSING OR SUSPECTED VALUES

Completeness of measured data is controlled there. Complete dataset for 10M data contains 144 values per day. For selected elements (average wind speed, humidity, soil humidity, air pressure, air temperature and soil temperature) the missing values are automatically interpolated by polynom. The number of continuous interpolated values should be less than 5 (40 minutes as maximum). Longer break in time series (1440 minutes, it is 24 hours, as maximum) may be interpolated for some elements by IDW method.

Interpolation by polynom

If the length of the break is up to four 10M values, system calculates the missing value of average wind speed, humidity, soil humidity, air pressure, air temperature and soil temperature by polynom. The application uses the Ferguson’s 3rd degree interpolation polynomial (Salomon 2006) for interpolation of missing values (T1 to maximum T4). Starting point (P1) is the last value in the time series and the end point (P2) is the first value after the interruption. For the definition of the polynomial the tangents (t1 and t2) are defined by the last two points (P1 and P0) before the interruption and by the first two points (P2 and P3) after interruption (Fig. 4).

Figure 3. Example of regular QC for TMA and T (maximum temperature less than temperature)
The interpolated values are flagged by special flag for interpolated values (Fig. 5).

Interpolation by IDW method

If the length of the break is longer than 4 and less than 1440 missing values system uses the inverse distance weighting (IDW) statistic for interpolate average wind speed, humidity, air pressure, air temperature and soil temperature. We use the IDW interpolation for one hour precipitation amount and one hour sun shine duration for the break of maximum 144 (24 hours) missing values also. Administrator may define the parameters of IDW for each station and each element in particular (Fig. 6).
The interpolated values are flagged by special flag for interpolated values (Fig. 7). The application prefers the polynomial interpolation of selected meteorological elements (average wind speed, humidity, soil humidity, air pressure, air temperature and soil temperature) to IDW interpolation (the short break is always interpolated by polynom). For one hour precipitation amount and one hour sun shine duration application uses only IDW interpolation.

**SPATIAL ANALYSIS OF METEOROLOGICAL DATA**

The values of the selected meteorological elements are automatically presented in the map section of the Czech Hydrometeorological Institute in the 1 hour, 1 day and 1 month time scales (Fig. 8). This is only possible because the values are subject of the automatic checking and labelling of missing and suspicious values, described above. The application CLIDATA uses the MLWR method for spatial analysis of temperature and snow cover and the IDW method for precipitation and solar radiation.
Figure 8. Hourly, daily and monthly maps of selected meteorological elements (www.chmi.cz)

Description of MLWR method

The application uses the digital terrain model of the Czech Republic (grid 100 x 100 m). In the first step, the most suitable stations are selected for each point based on the combination of terrain and the smallest distance (Němec, Stříž 2011). Each station is assigned a weight according to the formula:

\[ W = w_h \cdot W_h + w_v \cdot W_v + w_{a1} \cdot W_{a1} + w_{a2} \cdot W_{a2} + w_{c1} \cdot W_{c1} + w_{c2} \cdot W_{c2} \], where

\( W_h \) is weight of the horizontal distance between the grid point and the station,
\( W_v \) is weight of the vertical distance between the grid point and the station,
\( W_{a1}, W_{a2} \) are weights of slope orientation and slope between the grid point and the station,
\( W_{c1}, W_{c2} \) are the terrain convexity weights between the grid point and the station and
\( w_h, w_v, w_{a1}, w_{a2}, w_{c1}, w_{c2} \) are optional coefficients, the sum of which is equal to one.

The values of the individual coefficients depend on the meteorological element and the specific meteorological situation. The analysis uses for each point \( n \) the highest weight stations (maximum number is 10 stations) from which the element value is calculated by the smallest square method. The method may calculate the correction according to the differences in the values measured at the stations and the calculated regressions.

**CONCLUSION**

AWS data is stored in databases at 1 hour or more often at 10 minutes intervals. The AWS in CHMI imports about 28,000 values each 10 minutes. It is not possible to rely on manual user QC of this data. Therefore, applications are developed to check, label, and optionally update and repair data automatically. The paper shows several examples of such methods used in the CHMI in the CLIDATA application. This control system is further developed and edited. Our data in on-line mode can be presented to the public nowadays. CLIDATA prepares maps (and other graphical outputs) of selected meteorological elements for the CHMI portal. The public has the opportunity to continuously monitor the quality of our data and, if necessary, to draw attention to untreated discrepancies and errors.

**REFERENCES**


**BIOGRAPHY**

Since 1986 I am working with Czech Hydrometeorological Institute (CHMI) as climatologist, in the period of 2003-2011 as Deputy Director. I am WMO expert for Climate Database Management System (CDMS) and for climate data. I am a coordinator for developing and using of Czech Climate Database Application CLIDATA which is used in the CHMI since the year 2000. In cooperation with World Meteorological Organisation (WMO) the CLIDATA replaced the CLICOM in more than 30 meteorological services around the world (Estonia, Latvia, Lithuania, Montenegro, Serbia, Georgia, Tanzania, Ethiopia, Ghana, Namibia, Nigeria, Dominican, Trinidad and Tobago, etc.). Since 2011 I am GEO Principal and since 2014 IPCC Focal Point for Czech Republic. I am author/co-author of some scientific articles and publications and since 2012 chief editor of Czech Meteorological Bulletin.