Presentation of an Algorithm for the Automated Estimation and Gap-Filling of Missing Data in Daily Weather Records

J.S. Gosselin[a,*], R. Martel[a], C. Rivard[b]

[a]*Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 rue de la Couronne, Quebec City, Quebec, Canada*
[b]*Geological Survey of Canada, Quebec Division, 490 rue de la Couronne, Quebec City, Quebec, Canada*

Abstract

Daily weather data are useful in several areas of Earth sciences, including hydrology, hydrogeology and agronomy. However, weather datasets are often incomplete. The estimation of missing data can be a complex and tedious task. This is particularly the case for daily precipitation because of their high spatial and temporal variability. A user friendly, menu-driven, and interactive computer program for rapid and automatic completion of daily climatological series has been developed. Missing data for a given weather station are estimated using a multiple linear regression model, generated using data from nearby stations. For daily precipitation, it is possible to activate an option that forces the algorithm to preserve the probability distribution of data. This is an advantage over conventional approaches that tend to overestimate the number of wet days and underestimate the high intensity precipitation events. The software also allows downloading and automatic formatting of raw data available on the Environment Canada website. The software is demonstrated for two weather station located in Monteregie Est region, southern Quebec. Cross-validation was used to check the method and to define the optimal parameters to minimize the error in estimating missing daily precipitation.

Keywords: `heat transport, recharge assessment, uncertainty analysis, subsurface temperature time series`

1. Introduction

Daily weather data are useful in several fields of Earth sciences, including hydrology, hydrogeology and agronomy. However, weather datasets are, most of the time, incomplete. This can represent a major hindrance in various applications, such as for the use of hydrological or hydrogeological simulators (e.g. SWAT, HELP, SHAW), which heavily depend on these data. Filling the gaps in weather datasets can quickly become a tedious task as the length of the data records and the number of stations increase. Moreover, it can also be quite complex when aspects such as time-efficiency of the method and accuracy of the estimated missing values are taken into account. This is particularly

*Corresponding authors
Email address: `jnsebgosselin@gmail.com` (J.S. Gosselin)

true for the estimation of missing daily precipitation data because of their high spatial and temporal variability (Simolo et al., 2010). Although various methods to estimate missing daily weather data are documented in technical papers (i.e. DeGaetano et al., 1995; Simolo et al., 2010), few published tools to perform this task efficiently and automatically are available.

This paper addresses this issue by presenting an open source algorithm, written in the Python programming language, that was developed to fill the gaps in daily weather datasets with an automated, robust, and efficient method. The missing weather data in the records of a given weather station (hereafter called the target station) are estimated through a Multiple Linear Regression (MLR) model using synchronous measurements from neighboring stations. The algorithm also includes an option to asses the validity of the method and the uncertainty of the estimated missing values through a cross-validation resampling technique.

In addition, a Graphical User Interface (GUI) for the algorithm has been developed and is included in the WHAT software (Gosselin, 2015). The algorithm and the WHAT software are both available for free at this web address: `https://github.com/jnsebgosselin/WHAT`. An example of an application of the algorithm, using the GUI provided in WHAT, is also presented at the end of this paper for the Montérégie Est study area (southern Quebec, Canada).

2. Description of the Algorithm

The algorithm described in this paper is based on the classical MLR (Multiple Linear Regression) method presented in Eischeid et al. (2000). The MLR method is a robust and well known spatial interpolation technique that can indirectly account for local effects, such as topography, land cover, land use and surface water. While creating serially complete daily datasets of air temperature and total precipitation for the western U.S., Eischeid et al. (2000) found that the MLR method consistently outperformed the other classical methods tested (normal ratio, inverse distance, optimal interpolation, and single best estimator). The same result was also found by Xia et al. (1999) for a study in Bavaria, Germany. Moreover, in a study conducted in Iran for different climate conditions (dry to extra humid conditions), Kashani and Dinpashoh (2011) found that the estimation obtained with the MLR method compared well with those obtained with more recent methods, more specifically the artificial neural network (reference) and the genetic programming (references) techniques.

A flowchart of our gap-filling algorithm is shown in fig. 1. The algorithm consists of two nested loops: the external 'Loop A' iterates over the weather variables of the dataset (min, max, and mean air temperature and total precipitation), while the inner 'Loop B' iterates over the missing values in each weather data series. Each missing value is estimated independently with a two-step procedure in 'Loop B'. The first step consists in the selection of the neighboring stations. The second step consists in building a MLR model, estimating the missing value, and filling the corresponding gap in the data series. The gap-filling algorithm of fig. 1 is described in more details in the following sections of this paper.

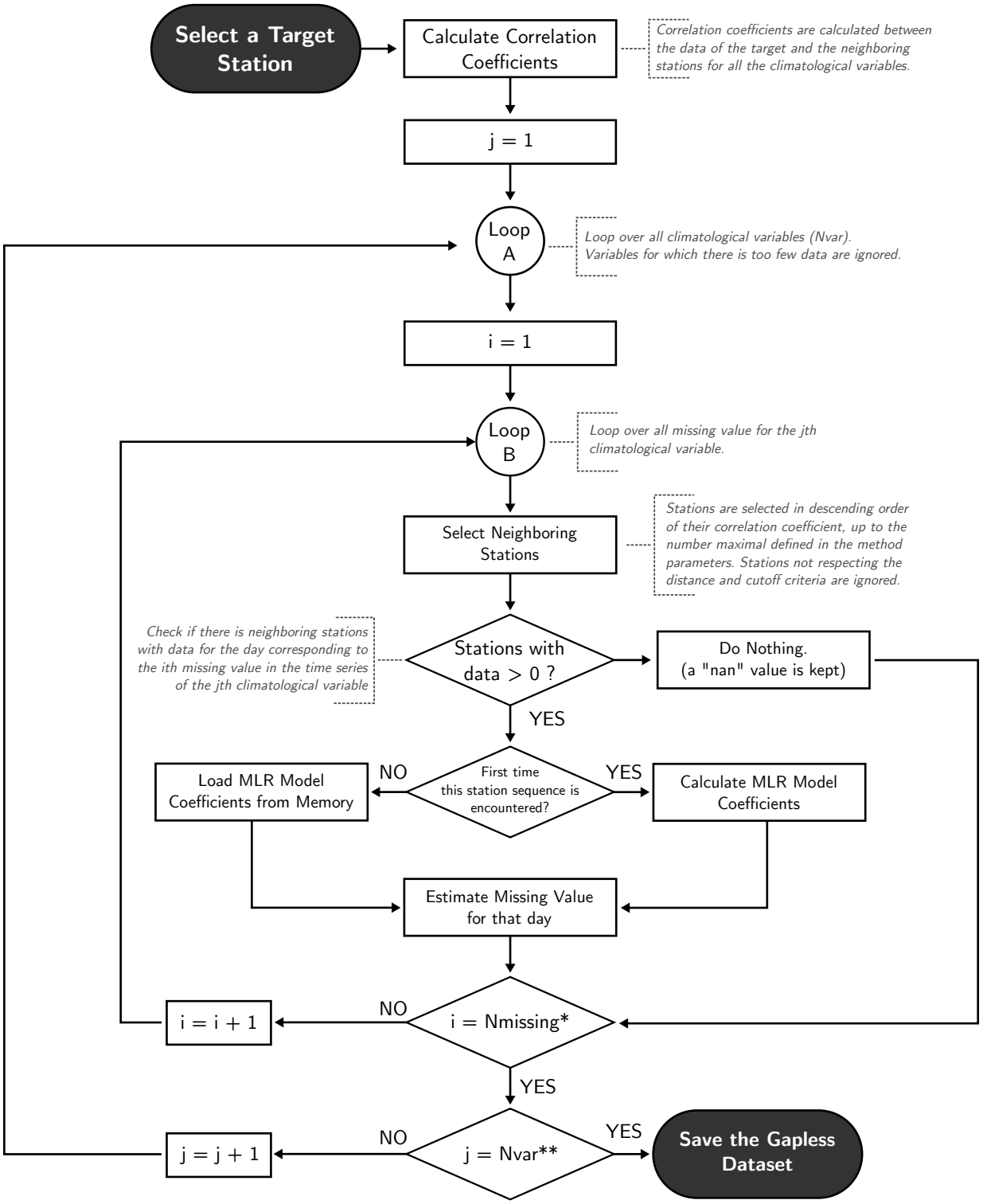
Select a Target Station
Calculate Correlation Coefficients
Correlation coefficients are calculated between the data of the target and the neighboring stations for all the climatological variables.
j = 1
Loop A
Loop over all climatological variables (Nvar). Variables for which there is too few data are ignored.
i = 1
Loop B
Loop over all missing value for the jth climatological variable.
Select Neighboring Stations
Stations are selected in descending order of their correlation coefficient, up to the number maximal defined in the method parameters. Stations not respecting the distance and cutoff criteria are ignored.
Check if there is neighboring stations with data for the day corresponding to the ith missing value in the time series of the jth climatological variable
Stations with data > 0 ?
Do Nothing. (a "nan" value is kept)
YES
NO
First time this station sequence is encountered?
YES
Load MLR Model Coefficients from Memory
Calculate MLR Model Coefficients
Estimate Missing Value for that day
NO
i = Nmissing*
i = i + 1
YES
NO
j = Nvar**
YES
j = j + 1
Save the Gapless Dataset

Figure 1

2.1. Correlation Coefficients Calculations

Correlation coefficients are calculated between the available data of the target station and those of the neighboring stations for each weather variable individually. The coefficients are calculated using all the available data. However, neighboring stations that have less than 182 days (half a year) of synchronous data with the target station or that have a correlation coefficient below a value of 0.35, for a given weather variable, are not used to fill the gaps in the data for that weather variable. The 0.35 threshold is based on the value used by Eischeid et al. (2000).

Moreover, it is possible to discard completely from the gap-filling procedure neighboring stations that are located further of the target station than specified thresholds, either in the horizontal or the vertical direction. The default values are set to 100 km and 350 m for the horizontal and vertical distance respectively, based on the values found in the literature Tronci et al. (1986); Xia et al. (1999); Simolo et al. (2010).

2.2. Selection of the Neighboring Stations

As stated by Eischeid et al. (2000), the selection of neighboring stations is critically important for the accurate estimation of missing weather data. Problems arise though because the set of neighboring stations with available data can vary from one day to the other. Therefore, the process of selecting the neighboring stations for the generation of a MLR model is repeated for each missing value in the dataset of the target station.

The neighboring stations with available data are selected in descending order of their correlation coefficient, up to a maximal number of stations that can be specified as a parameter in the algorithm. The default value for the maximal number of neighboring stations used for the generation of the MLR models is 4. Tests run by Eischeid et al. (2000) showed that using more then 4 neighboring stations did not significantly improve, and may even have degraded, the accuracy of the estimates. If for a given day with a missing value, no neighboring stations have a measured value, no calculation is done and a 'NaN' value is kept in the dataset.

2.3. Generation of the Multiple Linear Regression Model

Each time a MLR model is generated for a given sequence of neighboring stations, the result is stored into memory. Therefore, after a set of neighboring stations have been selected for a given day where a data is missing (section 2.2), the program checks first if this sequence of stations has already been encountered before for the current weather variable. If so, the stored MLR parameters will be used directly to estimate the missing data for the current day. Otherwise, a new MLR model will be generated from the newly encountered set of neighboring stations to estimate the missing data. Since a MLR model is generated only once for a given sequence of neighboring stations, the algorithm becomes faster with time as the number of MLR models stored into memory increases.

The MLR models can be generated using either an Ordinary Least Square (OLS) or a Least Absolute Deviations (LAD) criteria. Since daily precipitation series are generally characterized by long-tailed and positively skewed distributions, the LAD method is more appropriate than the OLS method since it is more robust to outliers (Menke, 1989; Eischeid et al., 2000). The downside in using the LAD method is an increase in computation time by about a

factor 10 compared to the OLS method. The resolution of the MLR models with the LAD method is achieved using an iterative reweighted least-squares method (reference).

2.4. Estimating Missing Daily Values

The value of the weather data for the target station is estimated from the synchronous measurements of the neighboring stations using the MLR model, such as:

$$Y_t = a_0 + \sum_{k=1}^{N} a_k \cdot X_k(t_i) \tag{1}$$

where $Y(t)$ is the value estimated for the target station at time t, $X_k(t)$ is the synchronous measured data for the k^{th} neighboring stations, a_k are the regression coefficients of the MLR model, and N is the total number of neighboring stations that were used for the regression. The intercept term, a_0, is estimated for the air temperature, but is set to zero for precipitation.

For precipitation, it is possible to have a negative regression coefficient for the less correlated neighboring stations used to generate the MLR model. If so, the MLR model will sometimes yield small negative values for the precipitation. To correct that, negative values estimated for daily precipitation are always set to zero.

2.5. Uncertainty of the estimated values

The accuracy of each MLR model generated throughout the gap-filling procedure is estimated by computing the root mean of squared residuals of the regression. It is also possible to evaluate the accuracy of the whole method (instead of each MLR model individually) for the entire dataset of given weather station with a Leave One Out (LOO) cross-validation procedure. The procedure consists in estimating a value for each day and each weather variable for which a measured data is available in the dataset (in addition to the days with missing data). In other words, the loops A and B in the flowchart of fig. 1 iterates over all the days and all the weather variables of the dataset instead of only iterating over days with a missing data. Before estimating a value for a given day, the corresponding measured data is temporarily discarded from the dataset of the target station to avoid self-influence of this observation on the generation of the MLR model. The accuracy of the method is then estimated by computing the Root-Mean-Square Error (RMSE) between the estimated weather data and the respective non-missing observations in the original dataset of the target station.

Since a new MLR model must be generated for each day independently, estimation of the accuracy of the method with the cross-validation procedure significantly increase the computation time of the gap-filling procedure. This is especially true if the MLR model is generated with the least absolute deviation regression method. For this reason, the cross-validation procedure is by default not activated in the algorithm.

3. Operation of the Algorithm

It is possible to use the algorithm with the Graphical User Interface (GUI) that is included in the free and open source software WHAT (Well Hydrograph Analysis Toolbox). A detailed description on the use of the algorithm with WHAT is provided in the user guide of the software (Gosselin, 2015).

Alternately, the gap-filling algorithm can be run directly from the command line in a Python interpreter version 3.4 or 2.7 or later. The external libraries *NumPy*, *Matplotlib*, *xlrd*, *PySide* and *Statsmodels* are required for the program to run. A minimal working example of an application is documented at the end of the python file. Some data samples to run the example are also provided with the algorithm.

The present section of this paper covers the format of the input data that is required for running the algorithm and the various outputs that are generated after a gap-less weather dataset has been successfully produced with the algorithm. Additional information about the input and output of the gap-filling algorithm is also provided in the user guide of the WHAT software.

3.1. Input Data

It is possible to use weather data from any sources with the gap-filling algorithm, as long as the data are saved in tab-separated values file with the '.csv' extension. Also, the labels in the first column of the file must be faithfully observed, since the algorithm is reading these to know where to retrieve the station information and the weather data within the file. It is recommended to use a copy of one of the sample files that are provided with the algorithm and fill-in directly the station information and the weather data. A "NaN" value must be entered where data are missing. The daily data must also be in chronological order, but do not need to be continuous over time. That is, missing blocks of data (e.g., several days, months or years) can be completely omitted from the time-series.

All the input weather data files must be saved in one single location that must be specified to the gap-filling algorithm. The algorithm will automatically scan this location for valid weather data files and will store the data in memory for analysis.

3.2. Parameters

The gap-filling algorithm is written as a Python class object, with the method parameters defined as class attributes. When using the gap-filling algorithm directly in a Python interpreter (without the GUI), the method parameters are specified by directly defining the value of their corresponding class attribute. An example is given at the end of the python file and each parameters is documented in the help section of the algorithm class, within the code. Additional information is also provided in the user guide of the WHAT software. A list of the different method parameters for the current version of the algorithm is presented in table 1.

3.3. Output

All the outputs that are produced after a gap-less weather dataset has been produced successfully are saved in a sub-folder named after the name of the target station, in a directory that must be specified to the gap-filling algorithm.

The gap-less weather datasets are saved in a tab-separated values file with a '.out' extension. Detailed information about the estimated values that were used to fill the gaps in the data series are also saved in an accompanying '.log' file. An histogram showing the yearly and monthly weather normals, calculated from the gap-less data series previously generated with the algorithm, is also produced and saved in a pdf format.

The results from the cross-validation procedure, if the option is enabled in the algorithm, are saved in a '.err' file. A figure comparing the probability density function of the original daily precipitation series to the estimated is also produced and saved in a pdf format. Scatter plots comparing the estimated and measured weather data are also produced for each variable of the dataset and saved in a pdf format. A list of the different output files that are produced with the current version of the algorithm is presented in table 2.

Table 1: List of method parameters for version 1.0 of the gapfilling algorithm.

Parameter name	Default value	Description
Nbr_Sta_max	4	Set the maximal number of neighboring stations that is used for the generation of the MLR models to estimate the missing daily weather data.
limitDist	100 km	Neighboring stations that are farther away from the target station than the specified value are completely excluded from the gap-filling procedure.
limitAlt	350 m	Neighboring stations with an absolute elevation difference with the target station that is higher than the specified value are completely excluded from the gap-filling procedure.
regression_mode	LAD	Define the optimization criteria that is used for the regression in the generation of the MLR model as described in section 2.3. The two options available are *OLS* (Ordinary Least Squares) or *LAD* (Least Absolute Deviations).
full_error_analysis	False	When set to *True*, the accuracy of the method, for the dataset of the target station, will be estimated with the cross-validation procedure described in section 2.5.
add_ETP	False	When set to *True*, daily potential evapotranspiration will be estimated from the daily temperature data series and will be saved in the '.out' file, along with the gapless data series produced with the gapfill algorithm.

Table 2: List of outouts for version 1.0 of the gapfilling algorithm. The name of the file are given for the weather station BROME.

File name	File type	Description
BROME (7020840)_1980-2009.out	data	
BROME (7020840)_1980-2009.err	data	
BROME (7020840)_1980-2009.log	data	
weather_datasets_summary.log	data	
weather_normals.pdf	figure	
precip_PDF.pdf	figure	
Max Temp (deg C).pdf	figure	
Max Temp (deg C).pdf	figure	
Mean Temp (deg C).pdf	figure	
Min Temp (deg C).pdf	figure	

4. Application: Monteregie Est Case Study

4.1. Materials and Method

4.1.1. Study Area

The algorithm was tested using data from 32 land-based Canadian weather stations in and around the Monteregie Est region, located in southern Quebec, Canada. This region covers a total area of 9032 km^2, from the St. Lawrence River at its northern limit to the border of the United States (states of New York and Vermont) at its southern limit (see Figure X). It is characterized by strongly variable topography and land cover conditions. The climate of this region is characterized by significant seasonal differences in temperature, resulting in warm summers and cold winters. Total precipitation, as rain or snow, are distributed rather evenly throughout the year.

4.1.2. Weather Dataset

A total of 32 weather stations were selected from the Canadian Daily Climate Database (CDCD) based on the availability and continuity of the measured weather data between 1980 and 2014. Table 3 presents the list of these selected stations with their corresponding climate ID, location coordinates (latitude and longitude), altitude, time periods for which data were available, yearly averages and percentage of days with missing data. Most of the information presented in table 3 are generated automatically when loading data into the gap-filling routine and saved in a file named 'weather_datasets_summary.log' within the previously defined output folder. The geographical disposition of the weather stations is also presented in the map of fig. 2.

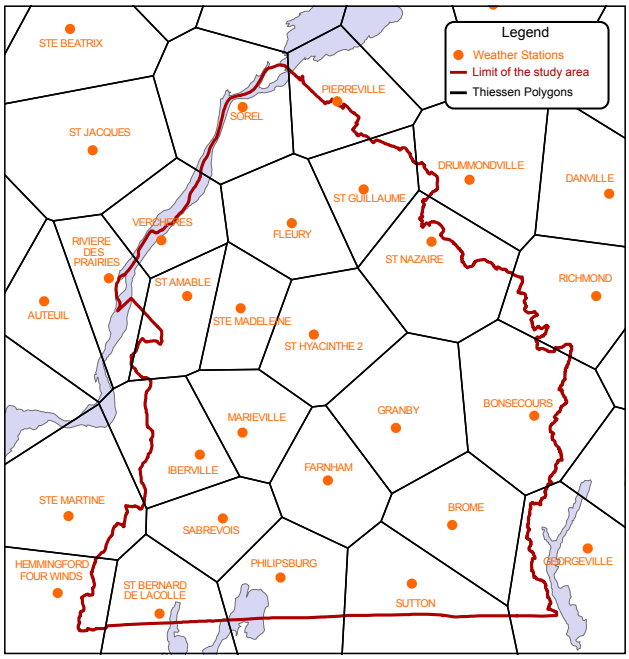

Figure 2: Locations of the weather stations in the Monteregie Est area.

The data were downloaded and formatted to the format described in section 3.1 using the software WHAT (Gosselin et al., 2015). WHAT provides a graphical interface to the online CDCD that allows to search for stations interactively using location coordinates, download the available data for the selected weather stations, and automatically organize the data in a format that is compatible with the gap-filling algorithm presented in this paper.

Total annual precipitation for all of these stations, the mean annual precipitation is 1100 mm/y. The highest total precipitation are observed at the Brome station (~1280 mm/y), while the lowest at the Sorel station (~960 mm/year). Mean annual air temperature in the study area is 5.9 °C, ranging from 4.3 to 6.7 °C while average monthly temperatures fluctuate between −12 to 21 °C. The minimum monthly temperatures are observed in January −17.1 to −13.6 °C) while the maximum monthly temperatures are observed in July (24 to 26.7 °C). The highest temperatures are observed at the Philipsburg et Saint-Bernard stations and the lowest at Bonsecours station.

4.1.3. Gap-Filling the Data

Missing data were estimated for the weather station located inside the Monteregie Est regions. The method was tested with the parameter value that are set by default in the algorithm, as shown in table 1.

The missing weather data and fill the gaps in the weather dataset of the station located within the study area. Stations bordering the limits but outside the study area were used to fill the data of the stations within the area, but their data were not filled. Data from these stations were use only to improve the spatial distribution of the neighboring stations for the weather station in the study area located near the limits.

The default value for the method parameters were kept in the algorithm. That is, the maximum number of neigh-

Table 3: List of selected weather stations in the study area and related information about missing data for the 1980-2014 period

#	Station name	Lat. °N	Lon. °W	Alt. m	% of days with missing data				Yearly Averages			
					T_{max} %	T_{min} %	T_{mean} %	P_{tot} %	T_{max} °C	T_{min} °C	T_{mean} °C	P_{tot} mm
1	Auteuil	45.65	73.73	53.0	20.8	19.1	22.8	18.4	11.3	1.4	6.4	989.7
2	Bonsecours	45.40	72.27	297.2	4.7	5.1	6.3	3.2	10.0	−0.8	4.6	1226.2
3	Brome	45.18	72.57	205.7	2.6	2.3	3.0	2.3	11.0	−0.4	5.3	1296.7
4	Bromptonville	45.48	71.95	130.0	3.9	4.2	6.0	1.8	11.1	−0.1	5.5	1137.9
5	Danville	45.82	71.98	190.0	30.8	31.1	33.1	30.2	10.6	0.4	5.5	1074.5
6	Drummondville	45.88	72.48	82.3	2.5	2.4	3.4	1.8	11.0	1.5	6.2	1122.1
7	Farnham	45.30	72.90	68.0	5.2	4.7	6.1	3.7				
8	Fleury	45.80	73.00	30.5	1.1	1.2	1.6	1.1				
9	Georgeville	45.13	72.23	266.7	26.5	26.2	27.6	5.7				
10	Granby	45.38	72.72	175.0	1.3	1.3	2.1	0.5				
11	Hemmingford	45.07	73.72	61.0	6.0	6.0	6.8	4.6				
12	Iberville	45.33	73.25	30.5	7.3	7.5	9.7	4.2				
13	Magog	45.27	72.12	274.0	4.2	4.2	5.3	4.6				
14	Marieville	45.40	73.13	38.0	10.2	10.4	11.2	9.8				
15	Nicolet	46.20	72.62	30.4	4.1	4.2	5.2	3.6				
16	Philipsburg	45.03	73.08	53.3	4.9	5.1	6.9	3.2				
17	Pierreville	46.08	72.83	15.2	6.3	5.4	6.9	4.9				
18	Richmond	45.63	72.13	123.1	3.3	3.3	3.8	4.0				
19	Riviere des Prairies	45.70	73.50	9.0	2.4	4.0	4.8	1.4				
20	Sabrevois	45.22	73.20	38.1	25.5	26.0	27.1	5.2				
21	Sorel	46.03	73.12	14.6	5.7	5.9	6.2	4.5				
22	St. Amable	45.67	73.30	41.1	8.6	10.4	11.8	7.7				
23	St. Bernard	45.08	73.38	49.3	9.6	9.7	10.5	9.0				
24	St. Guillaume	45.88	72.77	43.9	4.3	4.6	5.7	3.0				
25	St.Hyacinthe 2	45.57	72.92	33.0	6.7	6.9	7.5	6.6				
26	St. Jacques	45.95	73.58	69.0	11.7	11.9	14.0	10.8				
27	St. Janvier	45.73	73.88	61.0	40.5	40.7	41.6	21.9				
28	St. Nazaire	45.73	72.62	68.6	3.8	3.8	5.4	2.7				
29	Ste. Madeleine	45.62	73.13	30.0	5.8	6.4	7.0	5.1				
30	Ste. Martine	45.22	73.85	38.1	6.9	6.6	7.8	6.4				
31	Sutton	45.07	72.68	243.8	0.4	0.6	0.7	0.5				
32	Vercheres	45.77	73.37	21.0	3.5	3.5	4.7	2.2				

boring station was set to 4, the horizontal and vertical distance thresholds were kept at 100 and 350 values respectively. The data were filled using the OLR and LRM method to see if it makes any real difference. Also, during the entire process, the option for the cross-validation was activated.

The accuracy of the method was also assessed with different value imposed on the maximum number of neighboring stations to see the impact on the accuracy of the method.

4.2. Results

Tests have shown that inclusion of more than four stations does not significantly improve the interpolation and may in fact degrade the estimate.

The quality of the estimates is strongly affected by seasonality. Stations at higher elevations are difficult to estimate accurately, in large part because of the topographical diversity of the surrounding stations leading to degradation of spatial coherence among stations.

The tendency for all of the methods to have a negative bias is indicative of the nature of precipitation distributions to be positively skewed (interpolated values will tend to cluster about the median error rather than the mean).

According to Xia et al. (1999), the two most important factors in climatology are the inter-correlations in the station network, and the seasonal variations in the relations between the stations.

4.3. Discussion

However, weighing and regression-based techniques, including the MLR method, all tend to overestimate the number of rainy days, while heavy precipitation events are systematically underestimated. Therefore, the rainfall probability distribution is usually not preserved with these techniques). However, Simolo et al. (2010) have proposed a two-step procedure to modify the MLR method to address these issues.

An alternative approach would have been to calculate the correlation coefficient with a subset of data from the target series centered around the missing value, as it was done in Simolo et al. (2010) for instance. This approach allows for a better representation of the seasonal variations in the relationships between the stations. The downsides include a more complex algorithm to implement and a reduction of the method robustness and efficiency.

5. Conclusion

References

DeGaetano, A.T., Eggleston, K.L., Knapp, W.W., 1995. A Method to Estimate Missing Daily Maximum and Minimum Temperature Observations. Journal of Applied Meteorology 34, 371–380. URL: http://journals.ametsoc.org/doi/abs/10.1175/1520-0450-34.2.371, doi:10.1175/1520-0450-34.2.371. d270.

Eischeid, J.K., Pasteris, P.A., Diaz, H.F., Plantico, M.S., Lott, N.J., 2000. Creating a serially complete, national daily time series of temperature and precipitation for the Western United States. Journal of Applied Meteorology 39, 1580–1591. doi:1520-0450(2000)039<1580:CASCND>2.0.CO;2. d181.

Gosselin, J.S., 2015. WHAT (Well Hydrograph Analysis Toolbox). URL: https://github.com/jnsebgosselin/WHAT.

Gosselin, J.S., Rivard, C., Martel, R., 2015. User Manual for WHAT. Document written for software version 4.1.7-beta. Technical Report. INRS-ETE, Quebec City, Qc, Can. URL: https://github.com/jnsebgosselin/WHAT/raw/master/WHATMANUAL/WHATMANUAL.pdf.

Kashani, M.H., Dinpashoh, Y., 2011. Evaluation of efficiency of different estimation methods for missing climatological data. Stochastic Environmental Research and Risk Assessment 26, 59–71. URL: http://link.springer.com/article/10.1007/s00477-011-0536-y, doi:10.1007/s00477-011-0536-y. d265.

Menke, W., 1989. Geophysical Data Analysis: Discrete Inverse Theory. 1 edition ed., Academic Press, San Diego.

Simolo, C., Brunetti, M., Maugeri, M., Nanni, T., 2010. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. International Journal of Climatology 30, 1564–1576. doi:10.1002/joc.1992. d184.

Tardivo, G., Berti, A., 2012. A Dynamic Method for Gap Filling in Daily Temperature Datasets. Journal of Applied Meteorology and Climatology 51, 1079–1086. URL: http://journals.ametsoc.org/doi/abs/10.1175/JAMC-D-11-0117.1, doi:10.1175/JAMC-D-11-0117.1.

Tronci, N., Molteni, F., Bozzini, M., 1986. A comparison of local approximation methods for the analysis of meteorological data. Archives for Meteorology, Geophysics, and Bioclimatology, Series B 36, 189–211. URL: http://link.springer.com/article/10.1007/BF02278328, doi:10.1007/BF02278328. d218.

Xia, Y., Fabian, P., Stohl, A., Winterhalter, M., 1999. Forest climatology: estimation of missing values for Bavaria, Germany. Agricultural and Forest Meteorology 96, 131–144. doi:S0168-1923(99)00056-8. d182.