USER MANUAL FOR

WHAT

Well Hydrograph Analysis Toolbox

August 4, 2015

Document written for software version 4.1.7-beta

Jean-Sébastien Gosselin[1], Christine Rivard[2], and Richard Martel[1]

[1] *Institut national de la recherche scientifique, Centre Eau Terre Environnement, 490 rue de la Couronne, Quebec City, Quebec, Canada*

[2] *Geological Survey of Canada, Quebec Division, 490 rue de la Couronne, Quebec City, Quebec, Canada*

License

Contents

List of Figures

1 Introduction

1.1 What is WHAT

WHAT (Well Hydrograph Analysis Toolbox) is a free, open source, and cross-platform interactive computer program whose main focus is the interpretation of observation well hydrographs. It is written in the Python 2.7 programming language and is currently maintained and developed by Jean-Sébastien Gosselin at INRS-ETE (`www.ete.inrs.ca`). The source code and a stand-alone executable for Windows 7 are available free of charge for download on GitHub (`www.github.com/jnsebgosselin/WHAT`).

If you encounter any problems or errors during program execution, have any questions, or have suggestions on how to improve WHAT, please contact Jean-Sébastien Gosselin at this email address: jnsebgosselin@gmail.com.

1.2 Features

Below are listed the features that are available in the current version of WHAT, as well as the ones that are going to be incorporated in future versions of the software.

Features available in current version of WHAT :

- Gapless Daily Weather Time Series Preparation :
  - Graphical interface to the online Canadian Daily Climate Database (CDCD) to search for weather station by location coordinate.
  - Automatic downloading and formatting of available data from the CDCD.
  - Estimation of missing data and automatic gapfilling of the datasets.
- Data Visualization:
  - Data exploration in a user-friendly and dynamic graphical environment.
  - Production of publication-quality graphs in vectorial formats (pdf or svg).
- Weather Data Analysis :
  - Estimation of yearly and monthly normals.
  - Estimation of the daily potential evapotranspiration.

- Water Level Analysis :
  - Calculation of the master recession curve (MRC).
  - Estimation of groundwater recharge from the MRC using a continuous water-table fluctuation (WTF) method.

Features planned for future versions of WHAT :

- Synthetic hydrograph production for the estimation of groundwater recharge and prediction of water levels.
- Assessment of the level of confinement of the aquifer at the well location based on an analysis of the barometric response function of the water level record.

1.3 Installation

WHAT can run on Windows, Linux, or OS X computer operating systems. However, a stand-alone executable of the program is currently released and tested only for the Windows 7 platform. This executable should also be compatible with Windows XP. For the Linux and OS X platforms, the software can be run directly from the source code, provided that Python 2.7 and all the required third party packages are installed on the computer (PySide, NumPy, matplotlib, xlrd, xlwt).

The stand-alone executable for Windows 7 is distributed in a Zip archive that can be downloaded freely on GitHub (`https://github.com/jnsebgosselin/WHAT/releases`). This archive contains:

- the GNU General Public License;
- a folder named "WHAT" that contains all the necessary system files for the program to run, including the file "WHAT.exe" from which the software can be started;
- a folder named "Projects" where all input and output files used or created by WHAT are stored by default. This folder includes samples of input and output files that provide a quick and convenient way to test and learn the various features of the program.

Once the content of the Zip archive has been extracted, the program can be started directly from the "WHAT.exe" executable file that is contained within the folder named "WHAT". The software can conveniently run from any location on the computer or from any storage device without the need to install the program beforehand.

1.4 Overview of the Graphical User Interface

The Graphical User Interface (GUI) of WHAT mainly consists of a menu bar, a console area, and a central view panel (Fig. 1.1). The *menu bar* is located in the top right corner of the GUI. This is where the name of the current project is displayed and where it is possible to open an existing project or create a new one. The *console* is located at the bottom of the GUI and is used to report technical information about the various tasks accomplished by the program as well as warning and error messages. The console can be collapsed vertically to save space, or can be extended to the entire window area. The *central view panel* is the main component of the GUI. It is where the various features of the software are displayed. The content of this panel is divided into four tabs: *Download Data*, *Fill Data*, *Hydrograph*, and *About*. These tabs are described in more details below and are shown in Fig. 1.2.

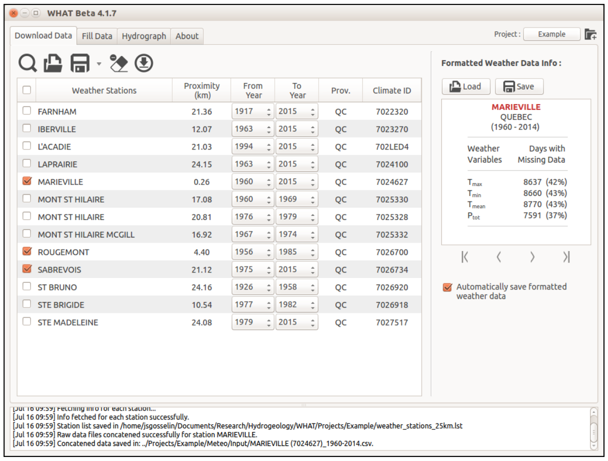
Figure 1.1: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *About* tab.

Download Data : This tab (Fig. 1.2a) provides a graphical interface to the online Canadian Daily Climate Database (CDCD), owned and operated by Environment Canada, from which it is possible to interactively search for stations by location coordinates, download the available data, and automatically organize the data in a format compatible with WHAT.

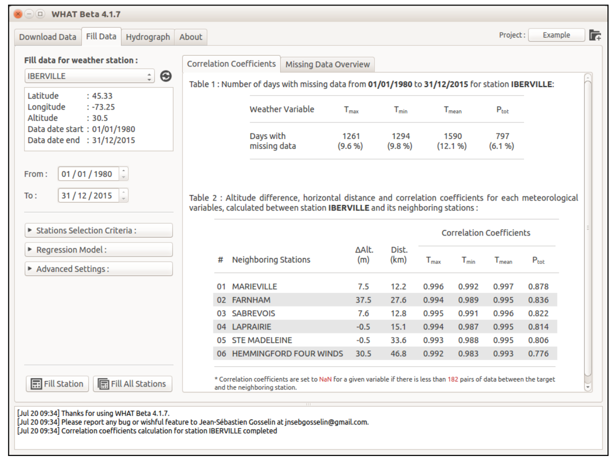
Fill Data : This tab (Fig. 1.2b) provides an automated procedure to estimate and fill the missing values in the daily weather datasets. Missing data in a dataset from a given station are estimated with data from selected neighboring weather stations using a multiple linear regression model.

Hydrograph : This tab is used for viewing and plotting both groundwater level and weather time series. For this purpose, two modes are available: the *layout* and the *computation* mode. Both modes share the same weather and water level datasets and it is possible to switch from one mode to the other at anytime. The **layout** mode (Fig. 1.2c) provides a graphical interface to interactively produce publication-quality graphs. The **computation** mode (Fig. 1.2d) consists in a dynamic graphical environment where data can be visualized, manipulated and analyzed. Various computational tools are available in this mode, including the estimation of the Master Recession Curve (MRC) and the estimation of groundwater recharge.

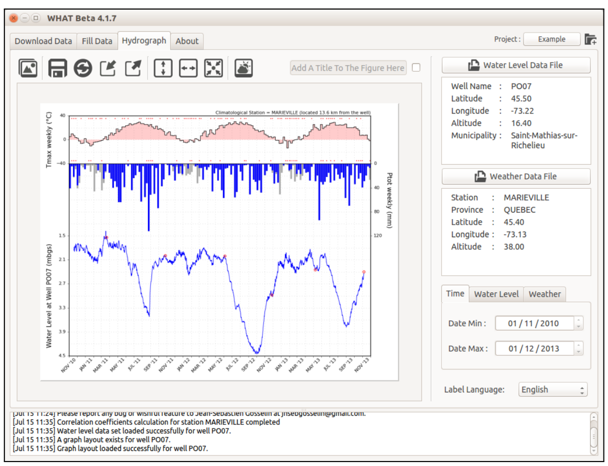
About : This tab (Fig. 1.1) displays copyright, licensing and general information about WHAT.



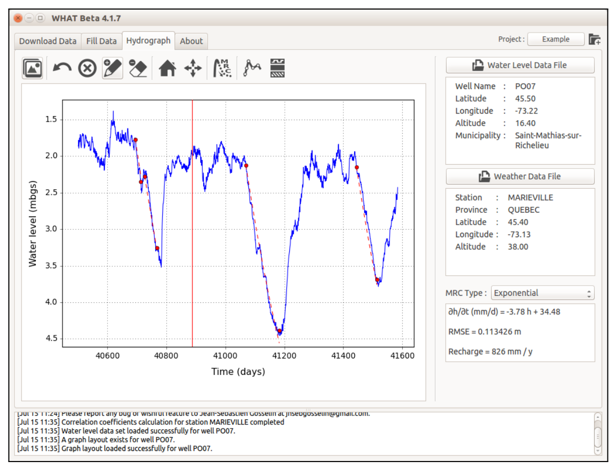
(a) *Download Data* tab.



(b) *Fill Data* tab.



(c) *Hydrograph* tab in *Layout* mode.



(d) *Hydrograph* tab in *Computation* mode.

Figure 1.2: Screenshots of WHAT v4.1.7-beta captured in Ubuntu Linux 15.04. showing : (a) the *Download Data* tab, (b) the *Fill Data* tab (c) the *Hydrograph* tab in *Layout* mode, and (d) the *Hydrograph* tab in *Computation* mode.

2 Data Management by Projects

2.1 Introduction

The data are managed in WHAT by project, that is to say, all input and output files relative to a particular project are stored in a common folder named “project folder”. This file management system allows to easily backup or copy the data related to a particular project since all the files are saved at the same location.

The first time WHAT is started, the software will automatically open the project “Example”, which includes all the files necessary to easily and quickly test the different features of the software. The title of the current project is shown in the menu bar located in the upper right corner of the WHAT window. Only one project at a time can be open per instance of WHAT.

2.2 Create a New Project

The creation of a new project can be started by clicking on the *New Project* button located in the menu bar (Fig. 1.1). This will open a new dialog window (Fig. 2.1) where information about the project can be entered such as its title, author, and location coordinates. Clicking on the button *Save* will create a new project folder named after the project’s title. Moreover, information related to the project are saved in a file with an extension “.what”. It is possible to change the directory where the project is saved by clicking the folder icon located next to the *Save in Folder* directory path.

For example, information related to the project *My New Project* by John Doe, in Fig. 2.1, would be saved in the file named “My_New_Project.what”, in the folder named “My New Project”, located in the directory “*...\WHAT\Projects*”.

2.3 Open a Project

It is possible to open an existing project by clicking on the button *Project*, which displays the name of the current project, in the upper right corner of the WHAT window. This will open a new dialog window, from which an existing project file (*.what) can be selected and opened. The current project name will then change for the name of the project that was just selected.

The path to the project folder is stored in a relative format. This means that if the location of the project folder is changed relative the executable of the software (“WHAT.exe”), WHAT will need to be redirected to the new location of the project by repeating the procedure described in the paragraph above.

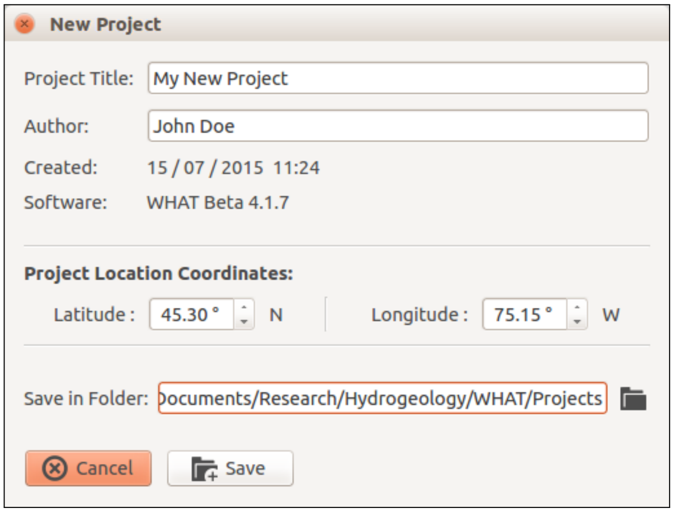


Figure 2.1: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *New Project* dialog window.

2.4 Project Folder Structure Overview

In addition to the project folder and the ".what" file that are created when saving a new project, WHAT automatically generates various files and sub-folders that are required for its execution. This file organization is briefly described below and an example is presented in Fig. 2.2. The project folder contains two sub-folders named "Meteo" and "Waterlvl" and a few other files.

Meteo : The sub-folder *Meteo* contains three sub-folders named respectively Raw, Input and Output. The **Raw** folder is where the weather data downloaded from the CDCD are saved, for each year separately, as csv (coma-separated values) files. All the files related to a same station are saved within a common folder, named after the name of the station and its climate ID. For example, in Fig. 2.2, the raw data file "eng-daily-01011980-12311980.csv", which contains weather data from the station "Marieville" for the year 1980, is saved within a folder named "MARIEVILLE (7024627)", where the number in parentheses is the climate ID of the station.

The folder **Input** contains the formatted weather data files produced from the raw data files. These are tsv (tab-separated values) files that are named after the name of the station, its climate ID, and the first and last year of the data record.

The folder **Output** is where the gapless weather time-series are saved in tsv files with the extension ".out". The files with the extension ".log" are tsv files that contain detailed information about the missing daily weather values that were estimated to fill the gaps in the weather datasets. The files with the extension ".err" contains a time-series of estimated weather values that were produced with a cross-validation re-sampling technique. These estimated values can be used to evaluate the accuracy of the method.

Waterlvl : The sub-folder "Waterlvl" is the preferred location where the water level datasets related to a same project should be stored. These files can be either in a Microsoft Excel spreadsheet file format (xls) or in a tab-separated values text format (tsv).

Other Files : The files with an "lst" extension correspond to the lists of weather stations from the Canadian Daily Climate Database (CDCD) that were saved by the user from the *Download Data* tab. The file "graph layout.lst" is a resource file, in which are stored the layout of the well hydrographs that were produced from the *Hydrograph* tab. The file "weather_datasets_summary.log" is a tsv file that contains a summary of all the weather data files included in the "Input" folder. The file "waterlvl_manual_measurements.xls" contains all the manual water-level measurements from field visits.

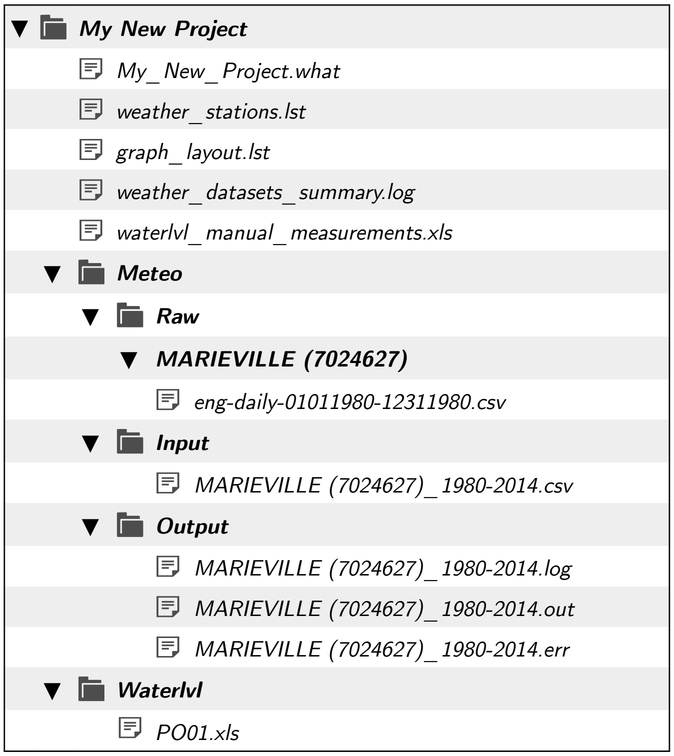


Figure 2.2: File organization of the project folders.

3 Creating Gapless Daily Weather Datasets

3.1 Downloading and formatting data from the CDCD

3.1.1 Introduction

The Canadian Daily Climate Database (CDCD) contains daily data for air temperature and precipitation dating back to 1840 to the present for about 8450 stations distributed across Canada. Data can be downloaded manually on the Government of Canada website (`www.climate.weather.gc.ca`) for each year individually and saved in a csv file. This process involves a lot of repetitive manipulations and is a time consuming task. Moreover, the re-organization of the individual data files, saved for each year separately, in a more convenient format can also represent a tedious task when done manually. Alternately, it is possible to order a DVD containing the entire database for a small fee. This option has the disadvantage of only providing an image in time as data cannot be updated.

WHAT alleviates this process by providing a graphical interface to the online CDCD that allows to query for stations interactively using location coordinates, download the available data for the selected weather stations, and automatically organize the data in a format compatible with WHAT. These features are available in the *Download Data* tab shown in Fig. 3.1. This tab consists of a toolbar located at the top of the interface, an area where are displayed the current list of weather stations for which data can be downloaded, and a side-panel to the right where can be manage the formatting of the weather data files that were downloaded for each year individually.

3.1.2 Searching for weather stations

Before any weather data can be downloaded from the online CDCD, a list of stations must first be provided to WHAT. This can be done by selecting an already existing list of stations (files with a "lst" extension) by clicking on the opened document icon located in the toolbar.

Alternatively, it is possible to search for weather stations in the online CDCD by clicking on the magnifying glass icon in the toolbar. This will open a new dialog window (see Fig. 3.2) where it is possible to search for weather stations located within a given radius around a set of location coordinates (latitude and longitude) in decimal degrees. It is possible to further narrow down the search by including only stations with data available within a given period and/or with data available for a minimum number of years.

When all the parameters have been specified, the search for weather stations in the online CDCD can be initiated by clicking on the *Search Stations* button. The resulting stations are automatically displayed in a table, along with information regarding their proximity to the location coordinates used for the search, the years for which data are available, their province, and their climate ID. Selecting stations and clicking on the *Add Stations* button will add them to the current list of weather stations displayed in the *Download Data* tab.

It is possible to remove any weather station from the current list by selecting them and clicking on the toolbar eraser icon. The station list can be saved by clicking on the toolbar floppy disk icon. Alternatively, it is also possible to generate a list of weather stations by creating manually a lst file without using the graphical interface. This is done by retrieving station information from

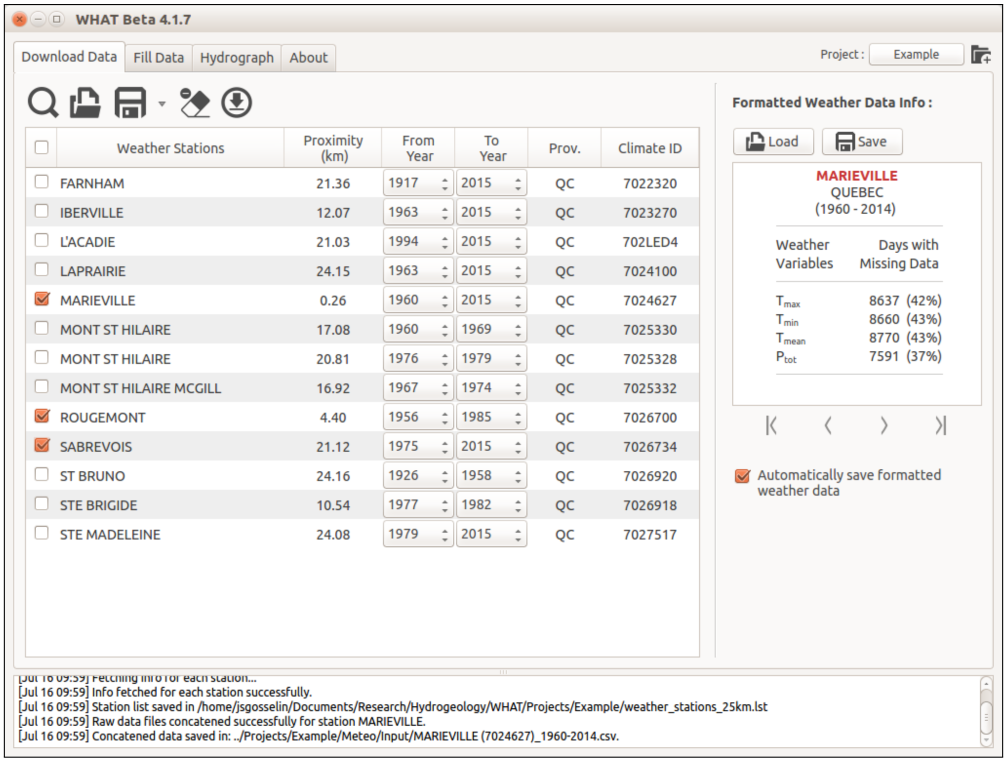


Figure 3.1: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Download Data* tab.

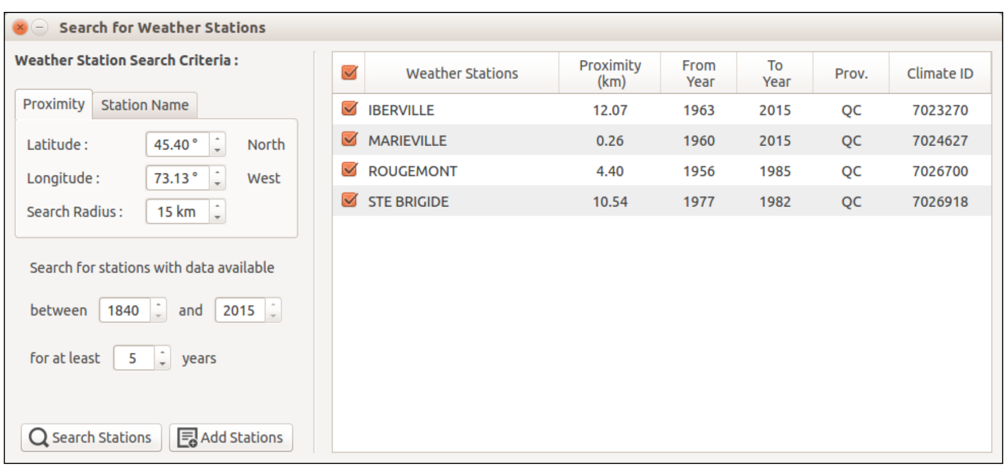


Figure 3.2: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *graphical interface* to the online CDCD database.

their unique url directly on the government of Canada website (www.climate.weather.gc.ca) and saving the information in a tsv (tabular-separated values) text file with a "lst" extension. A detailed example is presented in Appendix A

3.1.3 Downloading the weather data

Daily weather data can be downloaded from the online CDCD by selecting the desired stations from the list displayed in the *Download Data* tab and clicking on the toolbar icon with the encircled downward arrow. Data will be downloaded for the years specified for each selected station and the results will be saved automatically as a csv (comma-separated values) file in the *Raw* folder (see section 2.4). Weather data for a given station won't be downloaded for the years for which a data file already exist in the *Raw* folder. Detailed information about the downloading process are printed in the console area located at the bottom of the interface (see section 1.4). The downloading process can be stopped at any time by clicking on the stop icon that appears in the toolbar as soon a downloading task is started.

3.1.4 Formatting the weather data

WHAT automatically formats the data as soon as they have been successfully downloaded for a given weather station. To do this, data from each annual file are put together end to end in chronological order. Only the data related to air temperature (mean, max and min) and total precipitation are kept. In addition, days with missing data in the dataset are filled with a NaN (not a number) value. Finally, information on the number of days with missing data for each meteorological variable are displayed in the right side-panel. Alternatively, it is possible to open and format previously downloaded weather data files by clicking on the *Load* button in the right side-panel and selecting the desired files from the dialog window that will open.

By default, WHAT will automatically save the formatted data in a single tsv (tabular-separated values) file in the *Input* folder (see section 2.4). The automatic saving of the formatted data series can be disabled by unchecking the *Automatically save concatenated data* option. From the right side-panel, it is then possible to navigate through the datasets that were formatted over the course of a given session using the left-right arrows and save any dataset manually by clicking on the save button.

3.2 Filling the gaps in daily weather records

3.2.1 Introduction

Climate data are useful in several fields of Earth sciences, including hydrology, hydrogeology and agronomy. However, climate datasets are, most of the time, incomplete. This can represent a major hindrance in various applications, such as for hydrological or hydrogeological model simulations that heavily depend on these data.

Filling the gaps in weather datasets can quickly become a tedious task as the size of the data records and the number of stations increase. Moreover, it can also be rather complex when aspects such as time-efficiency of the method, accuracy of the estimated missing values, and preservation of the statistical properties of the weather time series (probability distribution and normals) are taken into account. This is particularly true for the estimation of missing daily precipitation data because of their high spatial and temporal variability (Simolo et al., 2010).

WHAT addresses this issue by providing an automated, robust, and efficient method to quickly and easily fill the gaps in the daily weather datasets from the files that were produced as described in Section 3.1.4. It is also possible to fill the gaps in weather datasets from files that were not produced with WHAT, provided that the data are formatted in the right format (see Appendix B). In addition, WHAT includes the framework to easily validate and assess the uncertainty of the estimated missing values with a cross-validation resampling technique. These features are available in the *Fill Data* tab shown in Fig. 3.3. This tab consists in a side-panel to the left where the gapfilling procedure can be managed and configured and an area to the right where various outputs are displayed.

3.2.2 Loading the weather data files

When starting WHAT or when a project is opened, the content of the *Input* folder is automatically scanned for weather data files. The results are displayed in a list of weather stations, located under the label *Fill data for weather station*. A summary of the number of days with missing data for each dataset is also produced and displayed in the tab *Missing Data Overview* of the display area, to the right. The icon with the circular arrows, located next to the list of stations, can be clicked to re-scan the *Input* folder for new weather data files to update the list of stations and the summary.

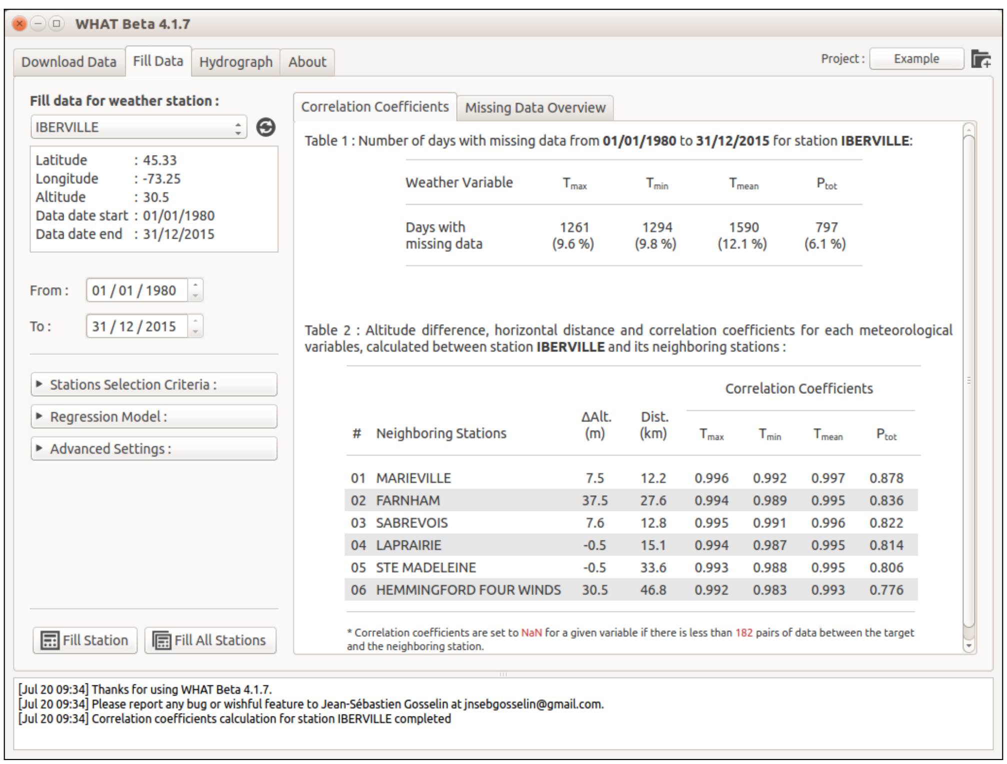


Figure 3.3: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Fill Data* tab.

3.2.3 Setting the model parameters

The first step is to select the station for which missing values in the dataset need to be filled. This is done from the drop-down list located under the *Fill data for station* section shown in Fig. 3.4. Under this list are automatically posted information about the currently selected weather station. It is also possible to define the period for which the data of the selected station will be filled by editing the dates fields located next to the *From* and *To* labels. By default, dates are set as the first and the last date for which data are available for any of the stations of the list.

The method used to estimate the missing data for the selected weather station consists in the generation of a multiple linear regression (MLR) model, using synchronous data from selected neighboring stations from the list. The neighboring stations are selected mainly on the basis of the correlation coefficients computed between their data and those of the selected weather station. The values of these coefficients are automatically displayed in the table located in the right side of the interface when a new weather station is selected from the list. Moreover, among the selected neighboring stations, the ones with the highest correlation coefficients have more weight in the model than those with weak correlation coefficients. For this reason, correlation coefficients that fall below a value of 0.7 are shown in red in the table, as a guidance for the user. There are several settings that can be used to control the selection of the neighboring stations, the generation of the MLR model, and the outputs of the gapfilling procedure. An overview of these settings is presented below.

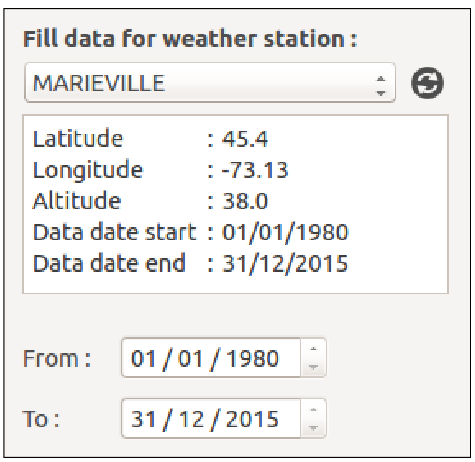


Figure 3.4: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Fill data for station* section. For this setup, missing values in the datasets of the selected weather station (MARIEVILLE) would be filled for the 01/01/1980 to 31/12/2015 period.

Station Selection Criteria : A MLR model is generated for each day for which a data is missing in the dataset of the selected station. This is done because the number of neighboring stations with available data can vary in time. Therefore, for a given date with missing data in the dataset of the selected station, the neighboring stations are selected in decreasing order of their correlation coefficients. Neighboring stations that also have a missing data at this particular date are excluded from the selection process. The maximum number of station that are selected for the generation of the MLR model can be specified in the *Nbr. of stations* field, located in the *Stations Selection Criteria* menu shown in Fig. 3.5. The number of neighboring station that is selected by default is 4. If for a given date, all the neighboring stations have missing data synchronously with the selected station, a NaN value is kept in the dataset at this particular date.

Moreover, the correlation between the data of two stations will, in general, decreases as the distance and the altitude difference between them increase. Therefore, the fields *Max. Distance* and *Max. Elevation Diff.* allow to specify thresholds for the distance and altitude difference.

Neighboring stations exceeding either one of these thresholds will not be used to fill the gaps in the dataset of the selected station. The default values for the distance and altitude difference are set to 100 km and 350 m, respectively, based on a literature review (Simolo et al., 2010; Tronci et al., 1986; Xia et al., 1999). The horizontal distances and elevation differences calculated between the selected station and its neighbors are shown in the table to the right, alongside the correlation coefficients. The values that exceed their corresponding threshold are shown in red.

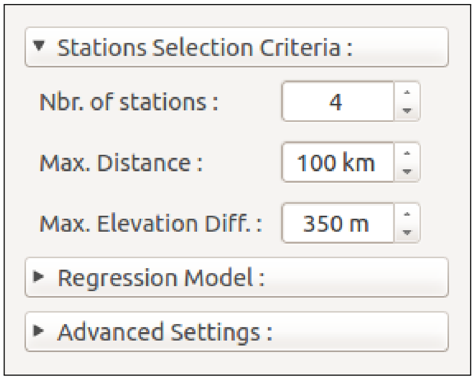


Figure 3.5: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Advanced Settings* menu.

Regression Model : It is possible to select whether the MLR model is generated using a Ordinary Least Squares (OLS) or a Least Absolute Deviations (LAD) criteria from the *Regression Model* menu shown in Fig. 3.6. A regression based on a LAD is more robust to outliers than a regression based on a OLS, but is more expensive in computation time.

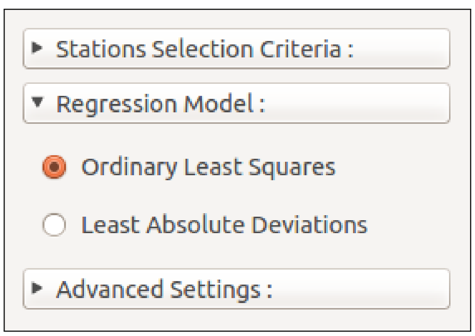


Figure 3.6: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Regression Model* menu.

Advanced Settings : It is possible to automatically estimate and add the daily Potential Evapotranspiration (ETP) to the output data file produced at the end of the gapfilling procedure of the selected station. This option is enabled by checking the *Add ETP to data file* option in the menu *Advanced Settings* shown in Fig. 3.7. The daily ETP is estimated with a method adapted from Thornthwaite (1948), using the daily mean air temperature time series of the selected station. Alternatively, it is possible to add manually the ETP to an existing weather data file by clicking on the open file icon located next to the *Add ETP to data file* option.

The *Full Error Analysis* option can be checked to perform a cross-validation resampling analysis during the gapfilling procedure. The results from this analysis can be used afterward to estimate the accuracy of the method. This option is discussed in more details in Section 3.2.5.

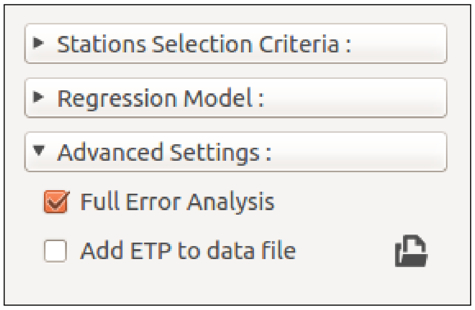


Figure 3.7: Screenshot of WHAT v4.1.7-beta in Ubuntu-Linux 15.04 showing the *Advanced Settings* menu.

3.2.4 Filling the gaps in the data

The automated procedure to fill the gaps in the dataset of the selected weather station can be started by clicking the button *Fill* located at the bottom of the left side-panel. It is also possible to run this procedure in batch mode to fill the gaps in the datasets of the entire list of weather station by clicking on the button *Fill All Stations*. The parameters for the gap filling procedure will, however, be the same for all the stations.

Once the process is completed for a station, the resulting gapless daily weather dataset is automatically saved in a tsv (tabular-separated values) file with the extension ".out" in the *Output* folder (see Section 2.4). The file is named after the weather station name, climate ID, and first and last year of the dataset. For example, the resulting output file for the station MARIEVILLE in Fig. 3.3 would be "MARIEVILLE (7024627) 1980-2015.out". In addition, detailed information on the values estimated for filling the gaps in the data are saved in a file with the same name as the ".out" file, but with a ".log" extension. Information includes, the names of the neighboring stations, the values of the data used for the estimations, as well as the expected uncertainty of the estimates.

3.2.5 Uncertainty of the estimated values

By default, each time a new MLR model is generated to estimate a missing value in the dataset of the selected station, the model is also used to predict the values in the dataset that are not missing. The accuracy of the MLR model is then approximated by computing a Root-Mean-Square Error (RMSE) between the values estimated with the model and the respective non-missing observations in the dataset of the selected station. The RMSE thus calculated is saved, along with the estimated value, in the ".log" file.

When the *Full Error Analysis* option in the *Advanced Settings* menu is enabled, WHAT will also perform a cross-validation resampling procedure to estimate the accuracy of the model, in addition to fill the gaps in the dataset. More specifically, the procedure consists in estimating alternately a weather data value for each day of the selected station's dataset, even for days for which data are not missing. Before estimating a value for a given day, the corresponding measured data in the dataset of the selected station is temporarily discarded to avoid self-influence of this observation on the generation of the MLR model. The model is then generated and used to estimate a value on this given day and the corresponding observed data is put back in the dataset of the selected station. When a value for every day of the dataset has thus been estimated, the estimated values are saved in a tsv (tabular-separated values) file in the *Output* folder with the extension ".err", along with the ".log" and ".out" files described in Section 3.2.4. The accuracy of the method can then be estimated by computing the RMSE between the estimated weather data and the respective non-missing observations in the original dataset of the selected station.

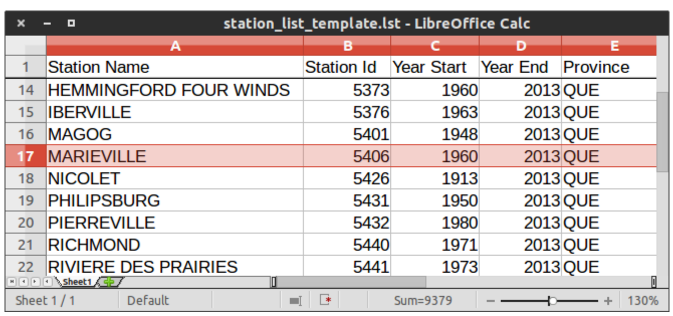
Activating this feature will significantly increase the computation time of the gap filling procedure, especially if the least absolute deviation regression model is selected, but can provide interesting insights on the performance of the procedure for the specific datasets used for a project.

A Creating a Weather Station List Manually

WARNING: this section needs to be updated and revised.

The stations information need to be saved in a tabular-separated values text file with an “lst” extension. A template of a station list (station_list_template.lst) is provided with the program in the Zip archive and an example is presented in Error: Reference source not found. The fields Station Name, Year, Start, Year End and Province do not need to match strictly with the station’s URL. These fields can be assigned any name/value by the user and are not directly used in the downloading process of weather data. The only field that is directly used in the download process is Station ID that is a unique number attributed to each weather station.

Once a file containing a list of weather stations’ information has been created, it is possible to load it in Rainbird by clicking on the Load button located in the Fetch and Merge tab (Error: Reference source not found). The station list can be refreshed at any time by clicking on the Refresh button.


station_list_template.lst - LibreOffice Calc

	A	B	C	D	E
1	Station Name	Station Id	Year Start	Year End	Province
14	HEMMINGFORD FOUR WINDS	5373	1960	2013	QUE
15	IBERVILLE	5376	1963	2013	QUE
16	MAGOG	5401	1948	2013	QUE
17	MARIEVILLE	5406	1960	2013	QUE
18	NICOLET	5426	1913	2013	QUE
19	PHILIPSBURG	5431	1950	2013	QUE
20	PIERREVILLE	5432	1980	2013	QUE
21	RICHMOND	5440	1971	2013	QUE
22	RIVIERE DES PRAIRIES	5441	1973	2013	QUE

Figure A.1: Weather Station List (*.lst) Sample

Step 1 First go to www.climate.weather.gc.ca.

Step 2 At the bottom of the page, click on Advanced Search (see the red arrow in Figure A.2a).

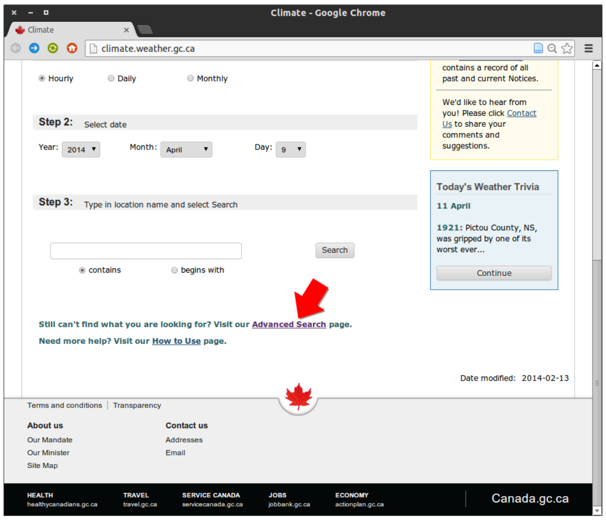
Step 3 Search for a station either by Province, Name or Proximity. For this example, the search was made using the Name of the station (see Figure A2).

Step 4 The research yielded one result (Figure A3). For Data Interval, select Daily and click Go. WARNING: The downloading and formatting of Hourly data is not currently supported in WHAT (but could be in a future version of the software).

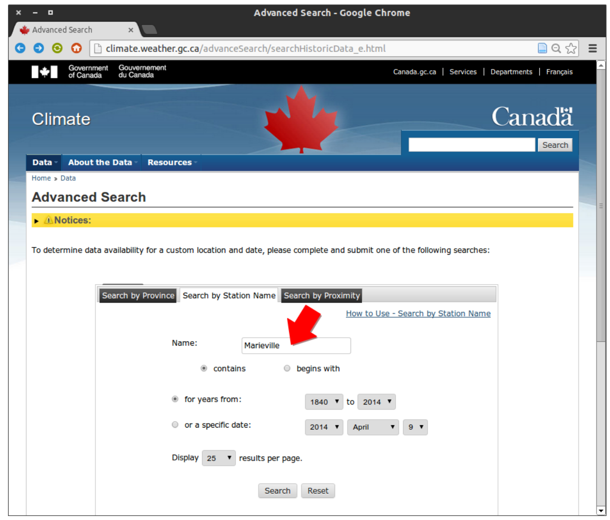
Step 5 The URL associated with Marieville weather station is:

http://climate.weather.gc.ca/climateData/dailydata_e.html?timeframe=2&Prov=QUE&StationID=5406&dlyRange=1960-06-01|2013-12-31&Year=2013&Month=12&Day=01.

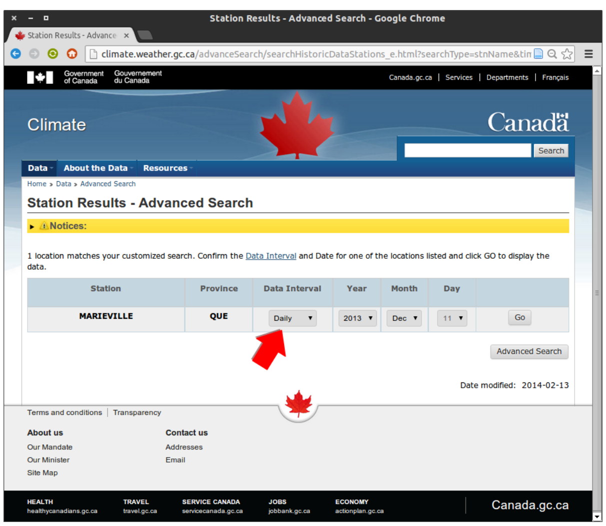
From this URL, we find that its Station ID is 5406 (see red circle in Figure A4), that is located in the province of Quebec (QUE) and that data are available from 1960 to 2013.



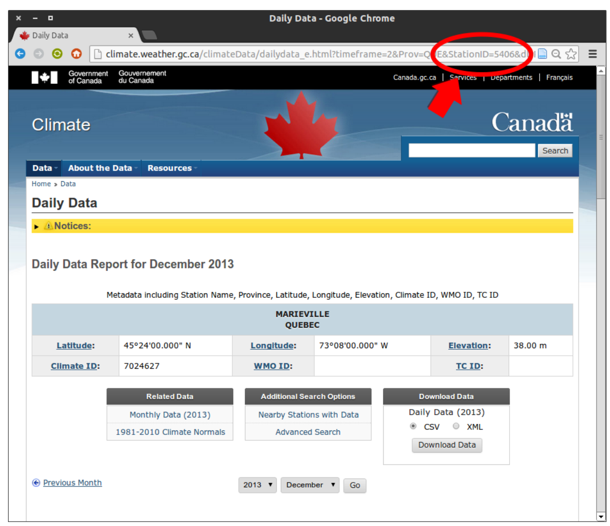
(a) Step 1



(b) Step 2



(c) Step 3



(d) Step 4

Figure A.2: Creation of a custom weather station list how-to

B Description of the weather data file format used in WHAT

It is possible to use weather data from any sources in WHAT, given the right format is used, either to fill the gaps in the weather time series and/or to interpret water level time series. For this purpose, it is recommended to use a copy of one of the sample files that are provided in the project example (distributed with the software) and fill the information and the data directly in it. The file must be kept in a text format using tab-separated values either with the extension ".csv" or ".out", depending if you want to fill the gaps in the weather time series or interpret water level time series. This can be achieved with any standard spreadsheet application such as Microsoft Excel or LibreOffice Calc. The format of the header must be faithfully observed for those files. In addition, "NaN" values must be entered where data are missing. 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 in the time-series. These missing blocks of data will be filled during the gap filling procedure or will be ignored for the plotting of the hydrograph.

Bibliography

C Simolo, M Brunetti, M Maugeri, and T Nanni. Improving estimation of missing values in daily precipitation series by a probability density function-preserving approach. *International Journal of Climatology*, 30:1564–1576, 2010. doi: 10.1002/joc.1992. D184.

C. W. Thornthwaite. An approach toward a rational classification of climate. *Geographical review*, 38:55–94, 1948. URL http://www.bcin.ca/Interface/openbcin.cgi?submit=submit&Chinkey=16399.

N. Tronci, F. Molteni, and M. Bozzini. A comparison of local approximation methods for the analysis of meteorological data. *Archives for Meteorology, Geophysics, and Bioclimatology, Series B*, 36(2):189–211, June 1986. ISSN 0066-6424, 1434-4483. doi: 10.1007/BF02278328. URL http://link.springer.com/article/10.1007/BF02278328. D218.

Y Xia, P Fabian, A Stohl, and M Winterhalter. Forest climatology: estimation of missing values for Bavaria, Germany. *Agricultural and Forest Meteorology*, 96:131–144, 1999. doi: S0168-1923(99) 00056-8. D182.