WISDom

Flowrate time series processing tool

Version 1.1.0

User's manual

IPS Instituto Politécnico de Setúbal

TÉCNICO LISBOA

inesc id lisboa

FCT Fundação para a Ciência e a Tecnologia

WISDom project

The WISDom - Water Intelligence System Data - research project aims at the development of algorithms and models that allow extracting knowledge from the data, supporting the water utilities in the decision-making and, thus, improving the management of its systems by reducing water losses and ensuring the provision of drinking water in quantity and quality.

The project was coordinated by Instituto Politécnico de Setúbal and was developed in collaboration with Instituto Superior Técnico and Instituto de Engenharia de Sistemas e Computadores, Investigação e Desenvolvimento em Lisboa. Three public water utilities of different size sand features collaborated in the project's development by providing the necessary data.

The WISDom project started on January 2019 and was developed throughout 42 months. It included 5 main tasks, namely (1) exploratory data analysis; (2) predictive analysis; (3) spatial leak location; (4) advanced recognition of anomalous behaviour; and (5) prototype development and recommendation establishment for public water utilities.

Software tools were developed in each specific project tasks. Such software tools are included in the WISDom platform, available at wisdom.ips.pt.

Table of contents

Table of Figures

1 Introduction

The WISDom – Flowrate time series processing is a software tool that allows the processing of unevenly and evenly spaced flowrate time series for use in multiple engineering computer applications, namely, for creating early warning systems against failures, for the calibration of hydraulic models or for pipe bursts detection and location. The time series processing include the automatic identification of the most common anomalies in flowrate time series due to acquisition and transmission problems, the time step normalization and the time series reconstruction.

This document is intended for the various users of the software and consolidates the information necessary to use all the software's functionalities. It should be used as a direct query in situations where the interface does not clarify the user regarding the process in question.

The following standards were considered when writing this document:

- Common formatting
  - ***Module***, when referring to the different software modules;
  - ***Button***, when referring to a specific button in a module;
  - ***.EXTENSION***, when referring to file extensions.
- Common abbreviations:
  - ***.CSV***, when referring to a delimited text file that uses a comma to separate values and with each line representing a data record.
  - ***.ZIP***, when referring to a high compressed file.
  - ***.EXE***, when referring to an executable program file.
  - ***.PDF***, when referring to a portable document format file.

1.1 Software goals

Processing raw flowrate time series (prior to any use by engineering tools) is essential to guarantee reliable data that do not compromise the success of the multiple engineering applications. The current software implements the methodology proposed by Ferreira et al. (2022)[1] for the processing of unevenly (and evenly) raw flowrate time series. The methodology includes four main steps, namely, 1) the automatic identification of anomalous values, 2) time series reconstruction in short duration gaps, 3) the time step normalization, and 4) the time series reconstruction in long duration gaps. This methodology is a very useful tool for the daily management of water utilities, preparing the time series

[1] Ferreira, B., Carriço, N., Barreira, R., Dias, T., & Covas, D. (2022). Flowrate time series processing in engineering tools for water distribution networks. Water Resources Research, 58, e2022WR032393. https://doi.org/10.1029/2022WR032393

to be used in different engineering tools, namely, hydraulic simulation, model calibration or online burst detection.

1.2 Requirements for using the software

The WISDom software has minimum hardware and software prerequisites that must be met in order to use it.

Below are the minimum prerequisites for using the software:

- **Hardware prerequisites**
  - Computer or laptop
- **Software prerequisites**
  - Windows operative system

1.3 Data requisites

The WISDom software was developed to process raw flowrate time series. Such time series should be in a specific file format to be used into the software, namely, in the ***.CSV*** format. The ***.CSV*** file should include rows (measurements), each with two values (separated per commas), namely, the timestamp and the value. The timestamp should be in the "yyyy-MM-dd HH:mm:ss" format[2]. Figure 1depicts an example of a ***.CSV*** format file of a flowrate time series ready for input into the software.

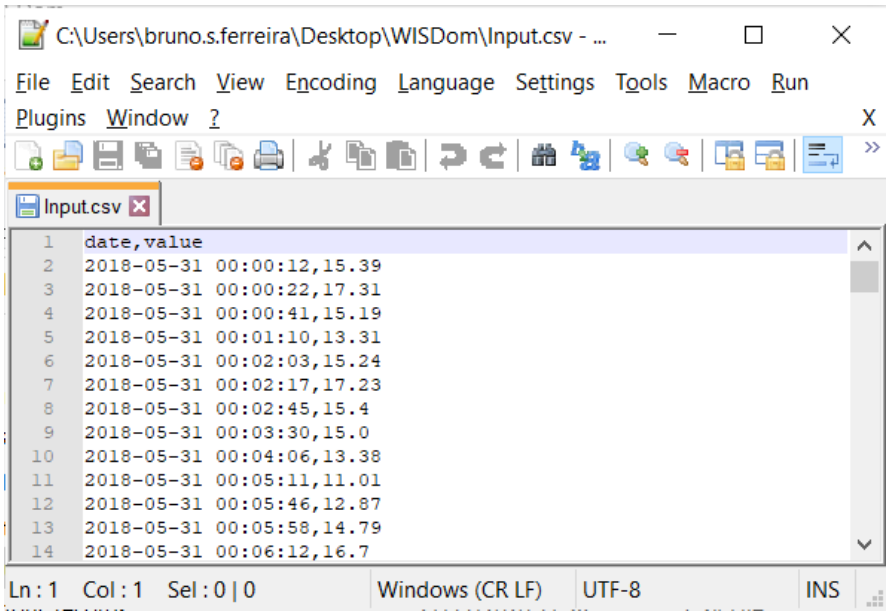

Figure 1 – Example of flowrate time series formated as input into the software.

[2] "2022-08-16 16:20:00" is an example of such timestamp format.

2 Available functionalities

The Flowrate time series processing tool presents a modular structure, divided into four processing modules, namely:

- **`Automatic identification of anomalous values`**
- **`Time series reconstruction in short duration gaps`**
- **`Time step normalization`**
- **`Time series reconstruction in long duration gaps.`**

The software allows the complete processing of the time series (using the four modules sequentially) or the exclusive use of a specific module.

`Multilanguage capabilities` are also implemented into the software, currently supporting English (by default) and Portuguese languages.

The following subchapters describe the required steps to set up the software and the main features in each module.

2.1 Set up

The required steps to access the WISDom flowrate time series processing tool are as follows:

1) Extract the contents in the ***`WISDom_1.1.0.ZIP`*** file into a folder.

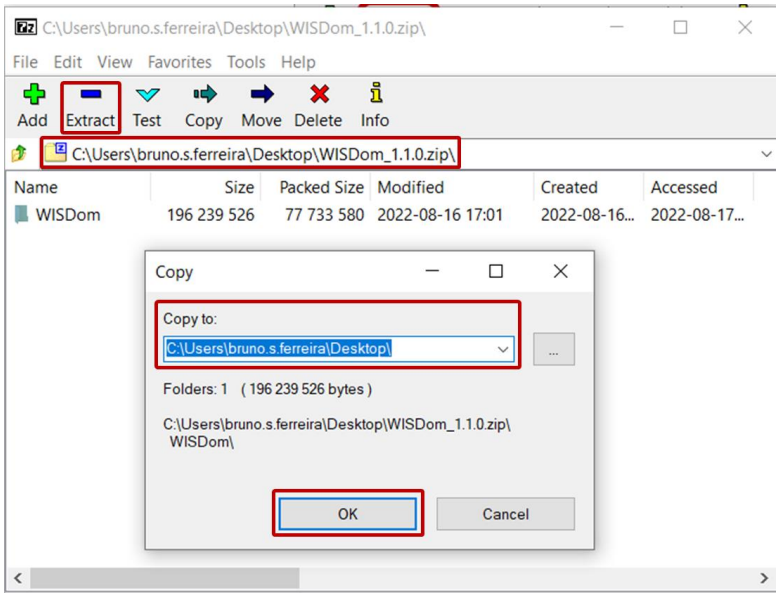

Figure 2 – Extracting the contents form the ***`.ZIP`*** file.

2) Access the ***`WISDom.exe`*** file in the recently created folder with extracted content.

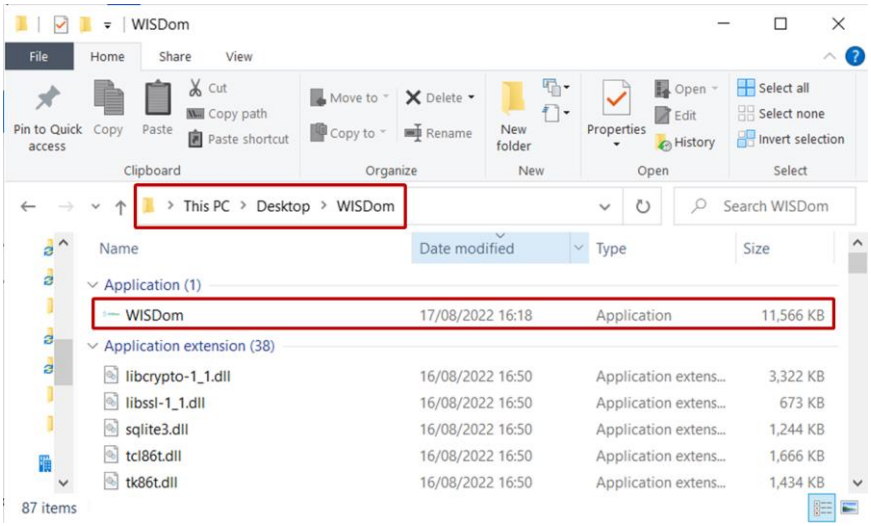

Figure 3 – Running the executable program file.

3) The software should open in a new window.

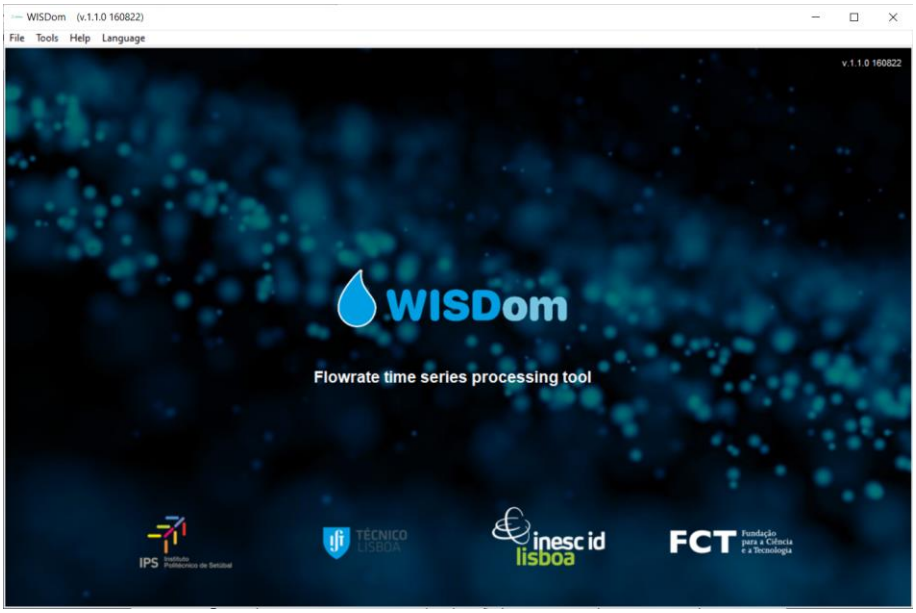

Figure 4 – Software front page.

2.2 Change language

The software language can be changed by using the ***Language*** button in the front page (See Figure 5). This opens a cascade menu with the supported languages. Note that the software language can only be altered in the front page menu and not during the processing.

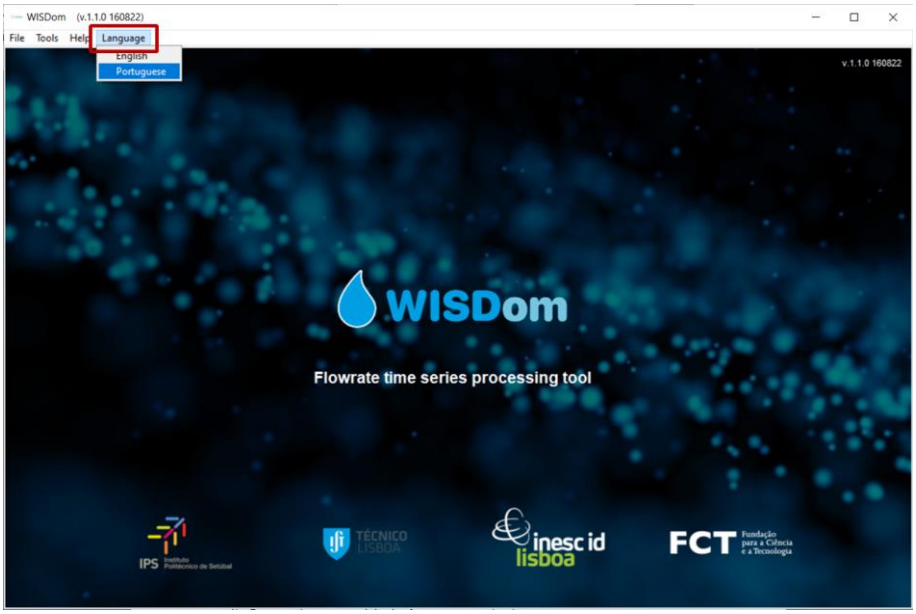

Figure 5 – Change software language.

2.3 Access to manual, research paper and about section

The original research paper with further description of each module can be directly accessed in the software. This can be done by using the ***Help*** button and by clicking in the ***Research paper*** in the cascade menu. The ***User's Manual*** and an ***About*** section can also be accessed form the same menu (see Figure 6).

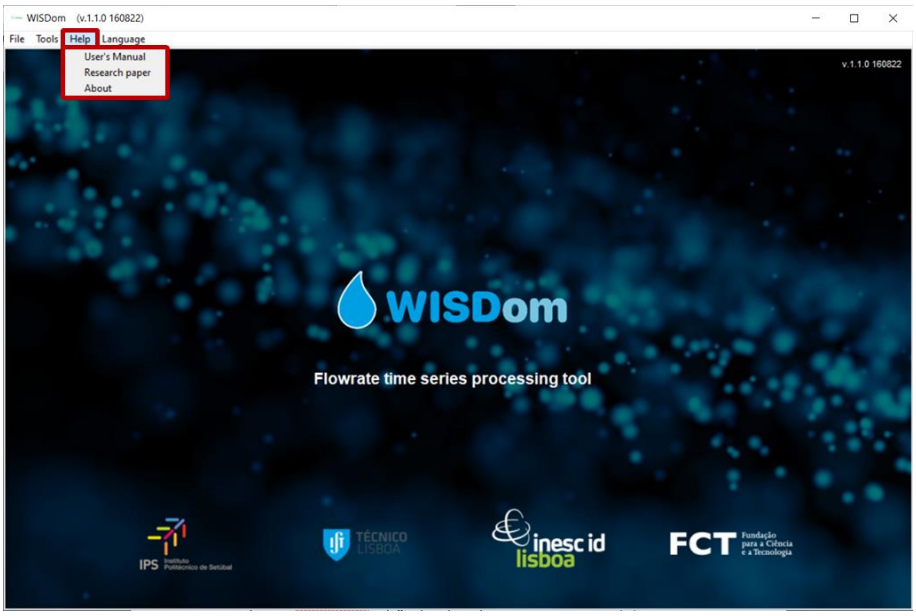

Figure 6 – Access to the Research paper, User's manual and About section.

2.4 Navigation between modules

As noted earlier, the processing tool presents a modular structure and is divided into four processing modules. The navigation between such modules can be done using the **`Tools`** button and by selecting the desired module in the cascade menu (see Figure 7).

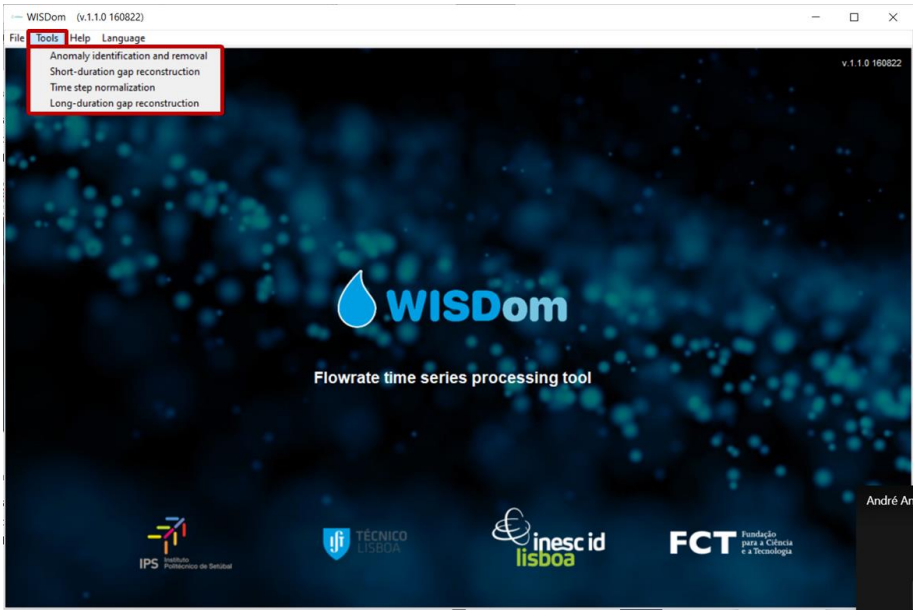

Figure 7 – Navigation between modules.

2.5 Automatic identification of anomalous values

The automatic identification of anomalous events module can be accessed in two distinct ways. The first way is through from the front page, by using the ***`File`*** button and by selecting the ***`Import`*** function (See Figure 8).

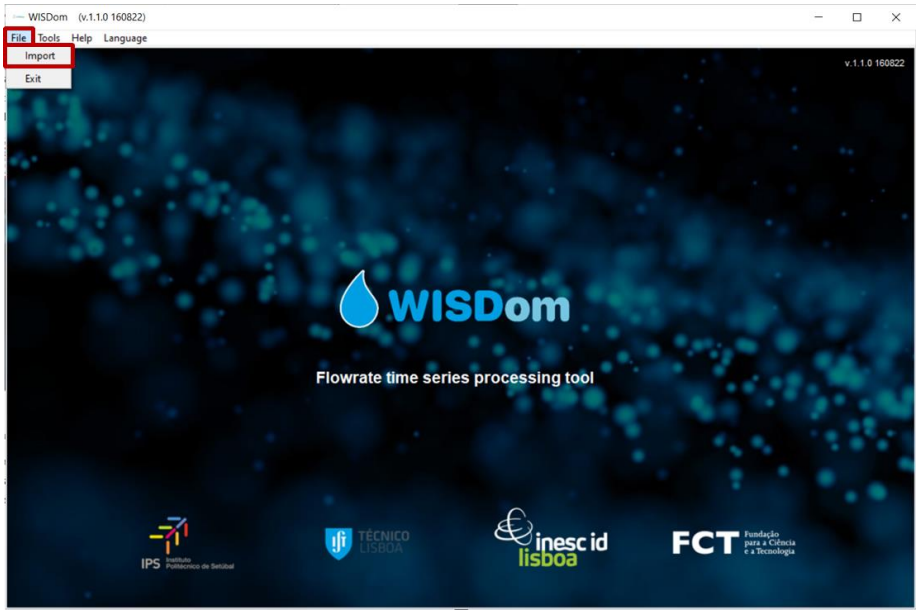

Figure 8 – Importing raw flowrate timeseries from the front page.

Alternatively, the user may change into the ***Automatic identification of anomalous values*** module as explained in 2.4. Subsequently, the file should be imported by using the ***File*** button and by selecting the ***Import*** function (see Figure 9).

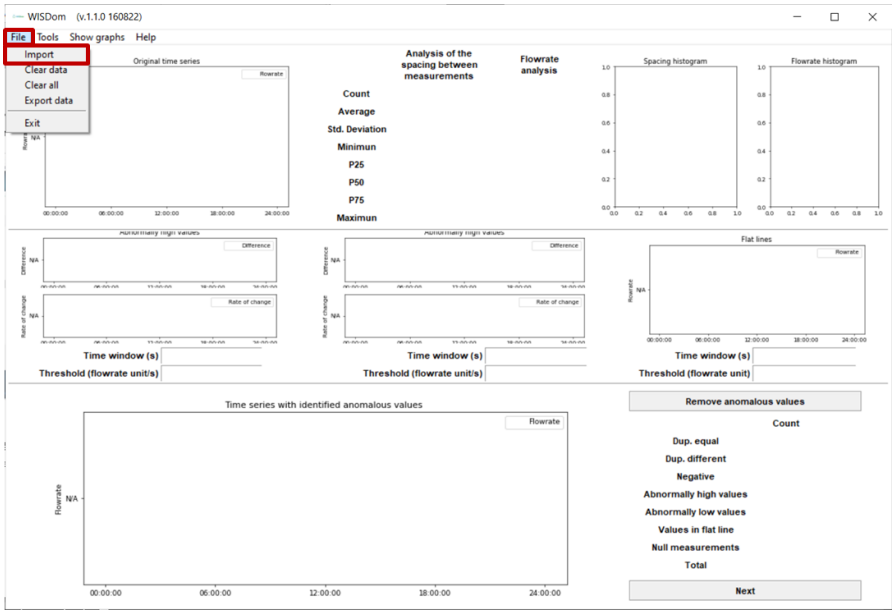

Figure 9 – Import raw time series file in the automatic identification of anomalous values module

This will prompt a file explorer menu, in which the user should select the raw time series to be processed. In this example, the Input.CSV (available in the WISDom folder) is used (see Figure 10).

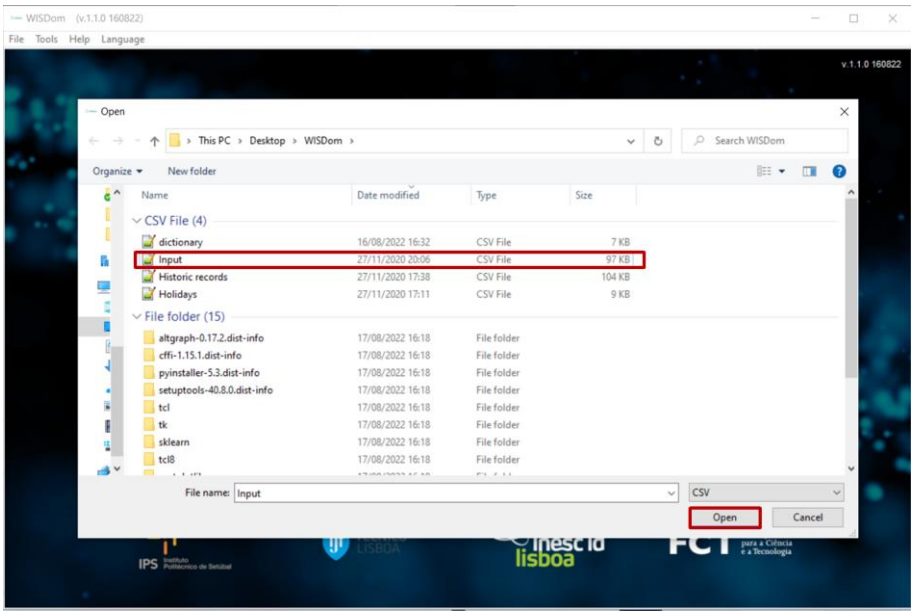

Figure 10 – Importing raw flowrate time series from the front page.

Upon importing, the software will update the ***Automatic identification of anomalous values*** module (see Figure 11), namely, with the graph of the raw time series, statistical information and histograms regarding the spacing between measurements and the flowrate values (in red rectangles). Additionally, a set of parameters (necessary for the processing) is automatically calculated and suggested, by default, to the user, which has the possibility to accept or to change the value of each parameter (in green rectangles). The anomalous values removal can be achieved by using the button ***Remove anomalous values***, depicted in Figure 11 as a blue rectangle.

Figure 11 – Automatic identification of anomalous values module after file import.

Upon removing anomalous values, the software will update with the graph of the processed time series with identification of the anomalous values (see red rectangle in Figure 12). Additionally, a table containing the count of each type of anomalies is updated (see green rectangle in Figure 12). At this point, the user has the possibility to either change the parameters values and to redo the anomalous values removal, or to proceed to the next processing module. The latter can be achieved by using the ***Next*** button, as depicted by a blue rectangle in Figure 12.

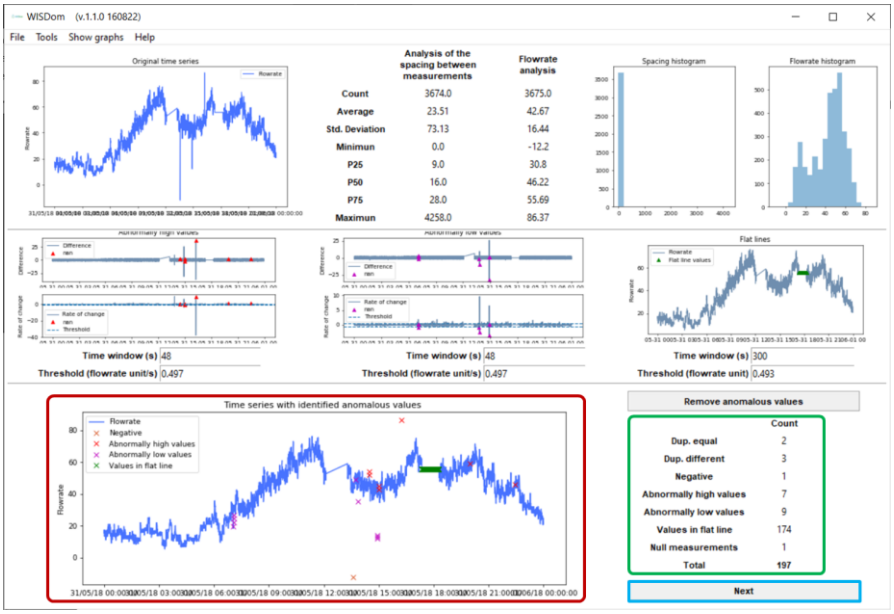

Figure 12 – Result of automatic identification of anomalous values

2.6 Time series reconstruction in short duration gaps

The time series reconstruction in short duration gaps module can be accessed in two distinct ways. The first way is through the previous module (***Automatic identification of anomalous values***) by using the ***Next*** button (as explained in the previous subchapter).

Alternatively, the user may directly change into the ***Time series reconstruction in short duration gaps*** module as explained in 2.4. Subsequently, the file (with already identified anomalous values) should be imported by using the ***File*** button and by selecting the ***Import*** function (see Figure 13).

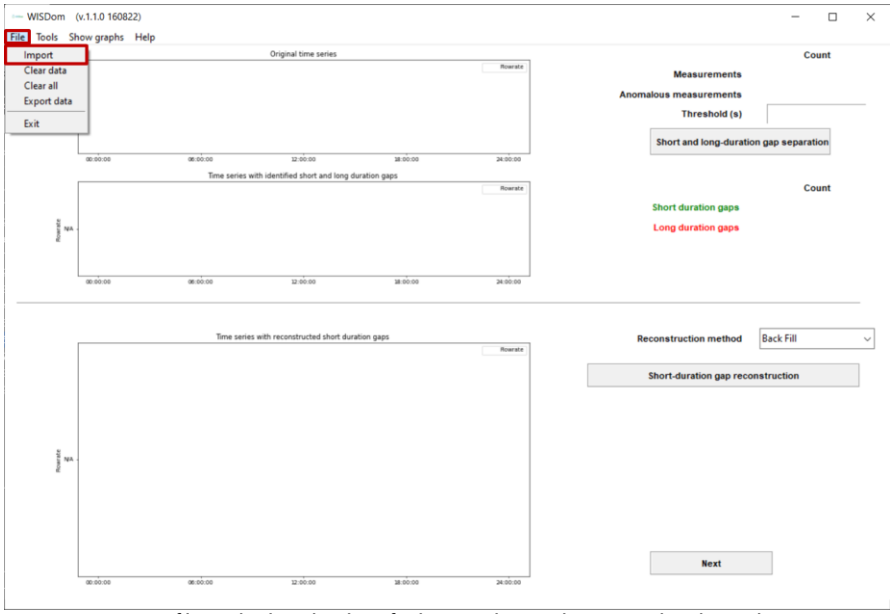

Figure 13 – Import time series file with already identified anomalous values into the short-duration gap reconstruction module.

Upon importing (or by clicking the ***Next*** button in the previous page), the software will update the ***time series reconstruction in short duration gaps*** module (see Figure 14), namely, with the graph of the time series with identification of the existing gaps of non-validated measurements (in red rectangle). The count of the total number of measurements and the number of anomalous measurements are also made available to the user (see green rectangle).

The user has the possibility to separate the existing gaps of non-validated measurements into short and long duration gaps. The former at reconstructed in this module, whilst the latter are reconstructed after the time step normalization process. To this end, a threshold should be defined by the user (default of 900 s), followed by using the button ***Short and Long-duration gap separation*** (see blue rectangle in Figure 14).

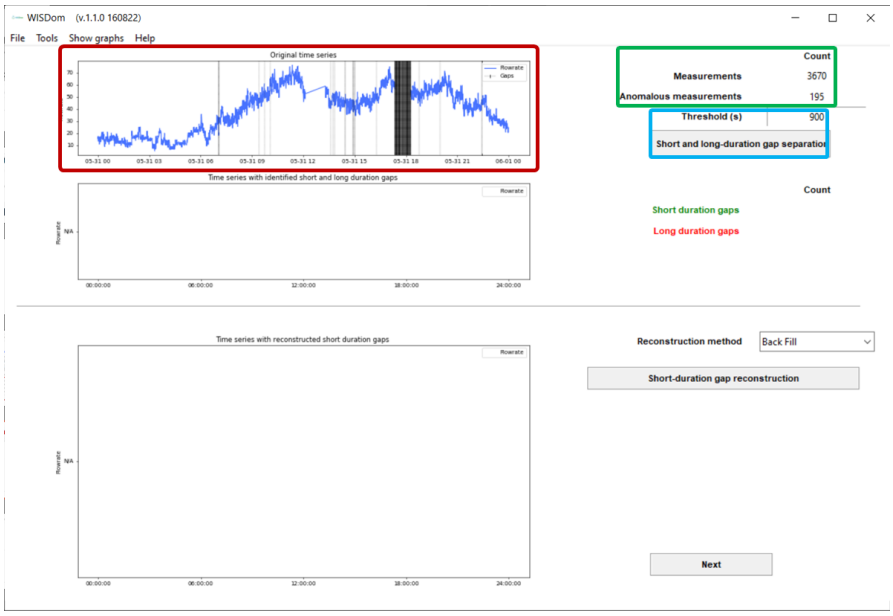

Figure 14 – Short and long-duration gap separation module prior to gap separation.

Upon separating gaps, the software will update with the graph of the processed time series with identification of the anomalous values (see red rectangle in Figure 15). Additionally, a table containing the count of each type of gaps is updated (see green rectangle in Figure 15). At this point, the user has the possibility choose the reconstruction method for short duration gaps by choosing an option from the dropout menu (see blue rectangle in Figure 15). Available methods for short gap reconstruction include Backfill (as default), Frontfill, Linear inteprolation and Polynomial interpolation. Once the method has been choosen, the reconstruction can be achieved by using the ***Short-duration gap reconstruction*** button, as depicted by a blue rectangle in Figure 15.

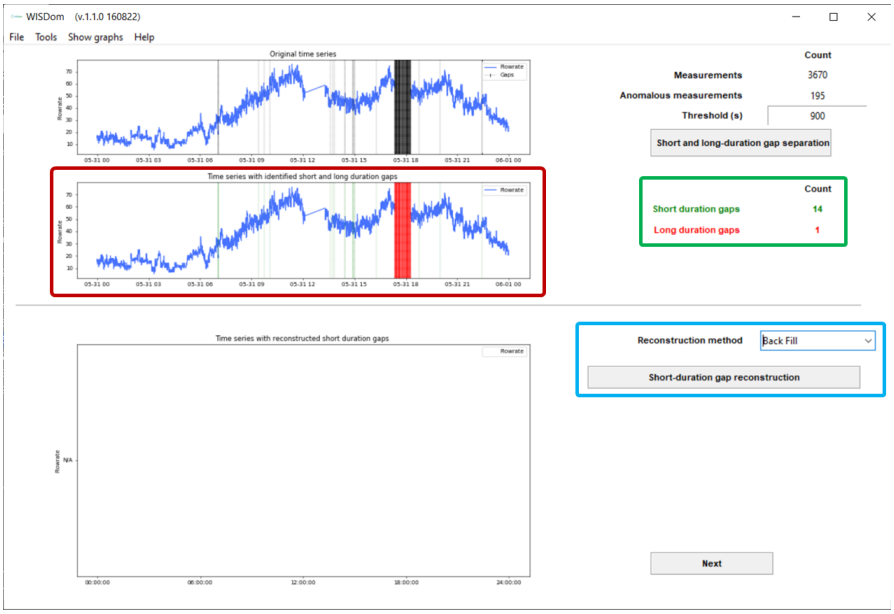

Figure 15 – Short and long-duration gap separation module after gap separation.

Upon reconstructing short-duration gaps, the software will update with the graph of the processed time series with identification of the reconstructed values and of long-duration gaps (see red rectangle in Figure 16). At this point, the user has the possibility to either change the reconstruction method and to redo the reconstruction process, or to proceed to the next processing module. The latter can be achieved by using the ***Next*** button, as depicted by green blue rectangle in Figure 16.

Figure 16 – Short and long-duration gap separation module after short-gap reconstruction.

2.7 Time step normalization

The ***Time step normalization*** module can be accessed in two distinct ways. The first way is through the previous module (***Time series reconstructin in short duration gaps***) by using the ***Next*** button (as explained in the previous subchapter).

Alternatively, the user may directly change into the ***Time step normalization*** module as explained in 2.4. Subsequently, the file (with short-duration gaps already reconstructed) should be imported by using the ***File*** button and by selecting the ***Import*** function (see Figure 17).

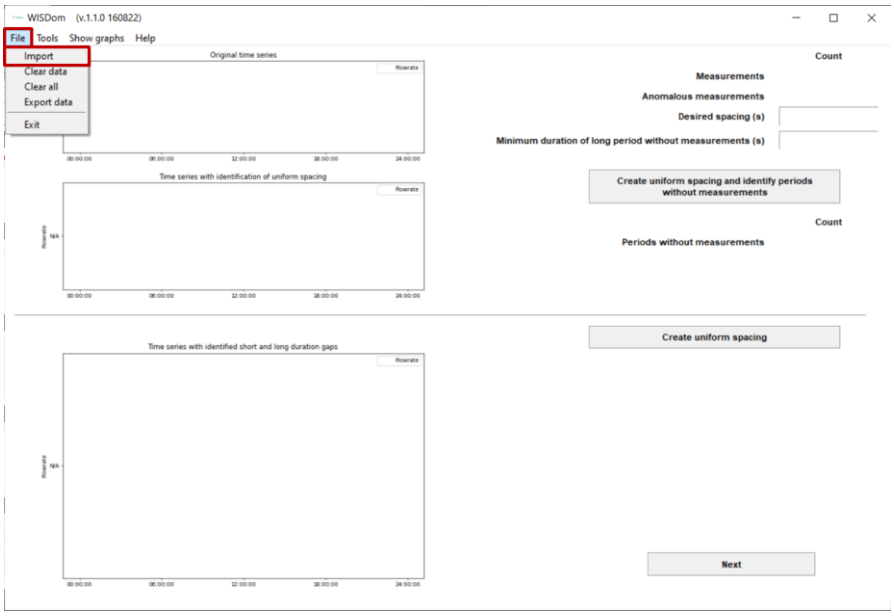

Figure 17 – Import time series file with short-duration gaps already reconstructed into the time series normalization module.

Upon importing (or by clicking the ***Next*** button in the previous page), the software will update the ***time series normalization*** module (see Figure 18), namely, with the graph of the time series with identification of the existing long-duration gaps (in red rectangle). The count of the total number of measurements and the number of anomalous measurements are also made available to the user (see green rectangle).

The user has the possibility to isolate the existing long periods without measurements. To this end, a threshold (minimum duration of long period without measurement) should be defined by the user (default of 900 s), as depicted in blue rectangle in Figure 18. Aditionally, the desired time step after the normalization process can be defined in the same area (default of 900 s). After the definition of the required parameters, the normalization process can be previewed by using the button ***Create uniform spacing and identify periods without measurements*** (see blue rectangle in Figure 18).

Figure 18 – Time step normalization module prior to the normalization process.

This step updates the graph of the time series with identification of long periods without measurements, as well as vertical lines representing the uniform spacing (see Figure 19). Note that this step is a preview and not necessary for the overall normalization process. In large data sets, this step should be skipped as it can be time-consuming. The time step normalization process can be carried by using the ***Create uniform spacing*** button (see blue rectangle in Figure 19).

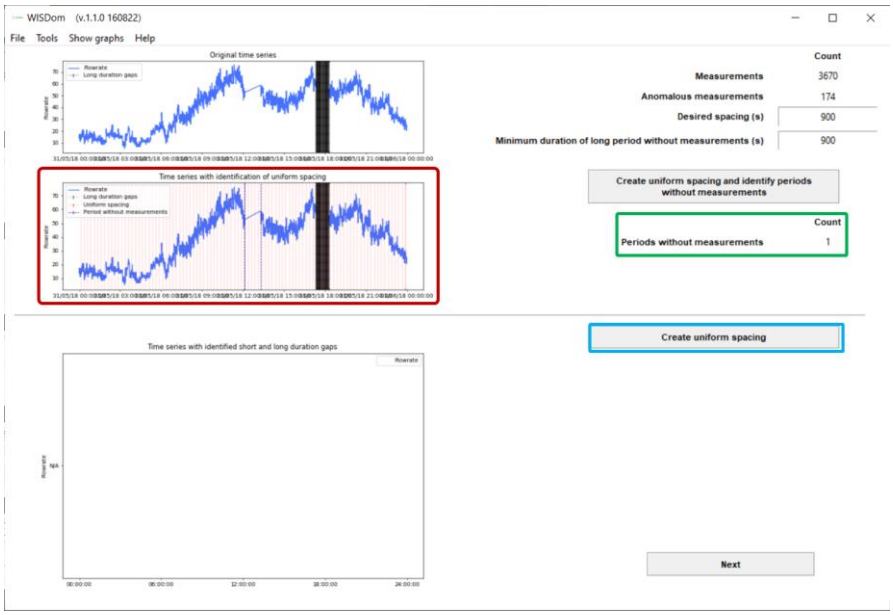

Figure 19 – Time step normalization module with preview of the normalization process.

Upon normalizing the time step, the software will update with the graph of the processed time series with normalized time-step (see red rectangle in Figure 20). At this point, the user has the possibility

to either change the time step normalization parameters and to redo the normalization process, or to proceed to the next processing module. The latter can be achieved by using the ***Next*** button, as depicted by green blue rectangle in Figure 20.

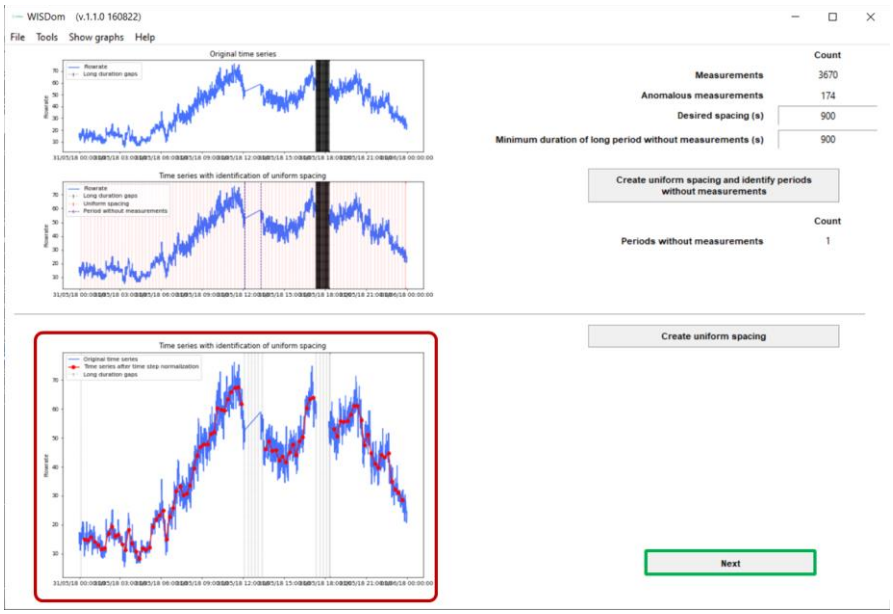

Figure 20 – Short and long-duration gap separation module after short-gap reconstruction.

2.8 Time series reconstruction in long duration gaps

The ***Time series reconstruction in long duration gaps*** module can be accessed in two distinct ways. The first way is through the previous module (***Time step normalization***) by using the ***Next*** button (as explained in the previous subchapter).

Alternatively, the user may directly change into the ***Time series reconstruction in long duration gaps*** module as explained in 2.4. Subsequently, the file (with normalized time step) should be imported by using the ***File*** button and by selecting the ***Import*** function (see Figure 21).

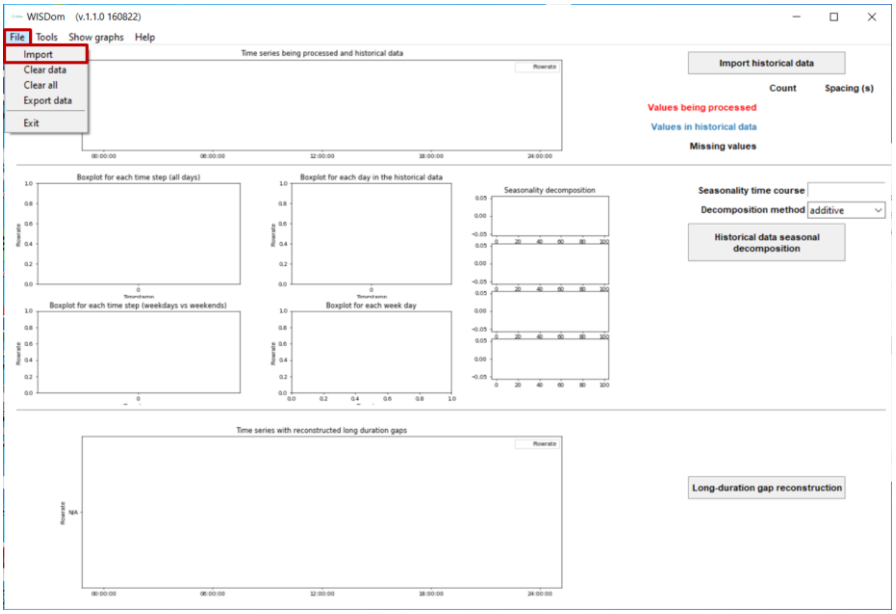

Figure 21 – Import time series file with normalized time step into the time series reconstruction in long duration gaps module.

Upon importing (or by clicking the ***Next*** button in the previous page), the software will update the ***Time series reconstruction in long duration gaps*** module (see red rectangle in Figure 22), namely, with the graph of the normalized time series with identification of the existing gaps (in red rectangle). The count of the total number of measurements is also made available to the user (see green rectangle in Figure 22).

Reconstructing long-duration gaps requires that the user imports already processed historical data of the same sensor. This can be done by using the ***Import historical data*** button (see blue rectangle in Figure 22).

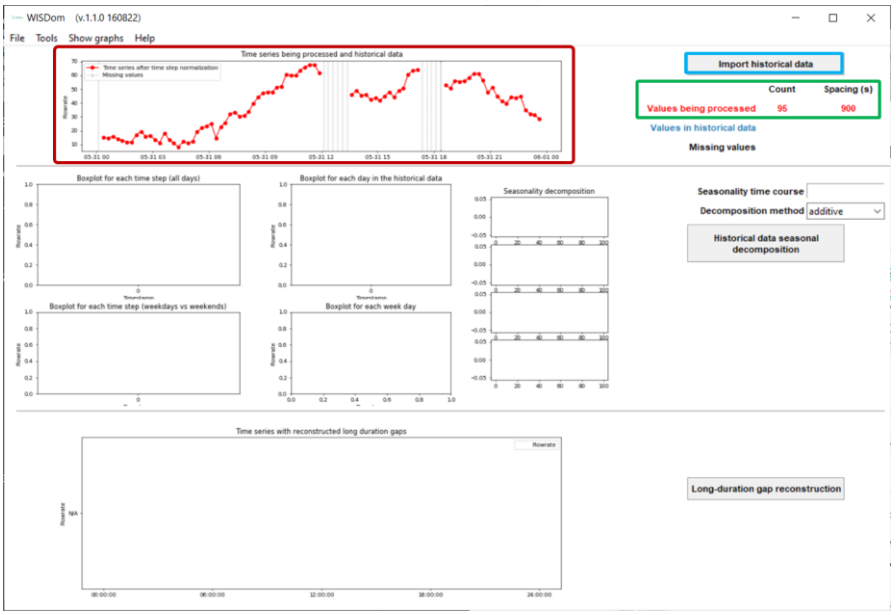

Figure 22 – Time series reconstruction in long durations gaps module prior to importing historical data

This will prompt a file explorer menu, in which the user should select the already processed historical time series to be imported. In this example, the Historic_records.CSV (available in the WISDom folder) is used (see Figure 23). The historical time series may not necessarily be on the same time step as the time series being processed. It needs, nonetheless, to be in an evenly spaced timestep.

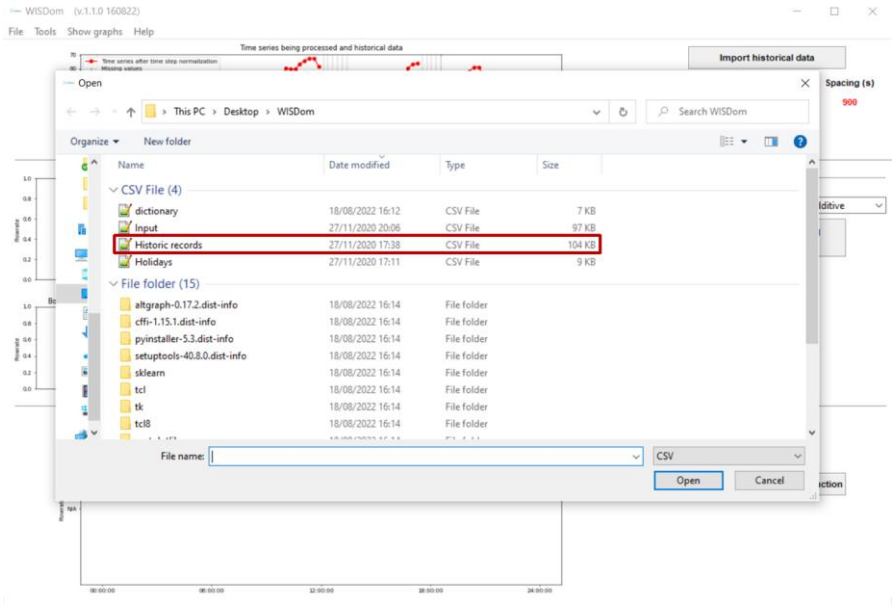

Figure 23 – Importing raw flowrate time series from the front page.

Upon importing, the software will update the ***Time series reconstruction in long duration gaps*** module (see red rectangle in Figure 24), namely, with the graph of the time series being processed and the historical data. Additionally, the total number of measurements in each time serie and their spacing is made available to the user, as well the total number of missing values (see green rectangle in Figure 24).

Aditional features are included that do not directly contribute to the reconstruction process, but allow the extraction of usefull information from the historical data. Such features includes seasonality decompomposition and distinct boxplots for better visualization of the historical time series. They can be accessed by using the ***Historical data seasonal decomposition*** button (see blue rectangle in Figure 24). The software automatically identifies the number of seasons in a period (considered to be a day). In this example, 96 periods correspond to a complete day divided into 96 periods of 900 s. Furthermore, distinct seasonality decomposition methods can be used, namely, additive and multiplicative methods.

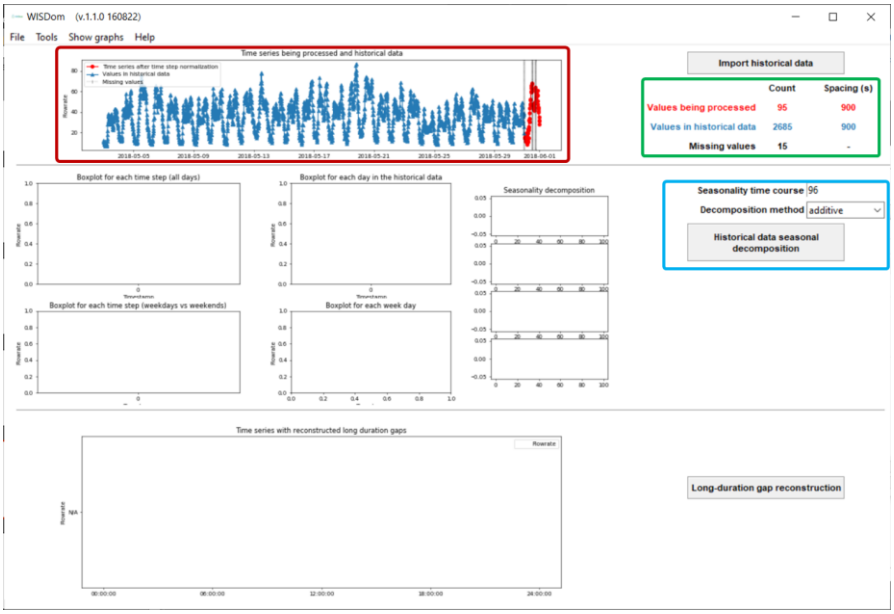

Figure 24 – Automatic identification of anomalous values module after file import.

This step updates the seasonal decomposition and box-plot graphs (see red rectangle in Figure 25). Note that this step does not directly contribute to the reconstruction process and can be skipped as it may be time-consuming.

The reconstruction of long-duration gaps can be carried by using the ***Long-duration gap reconstruction*** button (see green rectangle in Figure 25).

Figure 25 – Seasonal decomposition of historical records.

Upon reconstructing the long-duration gaps, the software will update with the graph of the processed time series (see red rectangle in Figure 26).

Figure 26 – Processed time series after the long-duration gap reconstruction.

2.9 File export

The WISDom software is prepared to export the time series being processed in any module and in any give time of the processing procedure. This can be done using the **`File`** button and by selecting the ***`Export data`*** option in the cascade menu (see red rectangle in Figure 27).

Figure 27 – Exporting processed data from any given module.

This will prompt a file explorer menu, in which the user should select the file name and destination. Note that the exported data will be on .CSV file format.

2.10 Clear data

Data can be cleared from the existing modules using two distinct functionalities. The first functionality clears the data from any given module and can be accessed by using the ***File*** button and by selecting the ***Clear data*** function (See red rectangle in Figure 28). Note that this function only clears the data from the module it was called upon.

Figure 28 – Clear data from any given module.

Alternatively, the user may use the second clear functionality, and in which all the data from the software is cleared (i.e., from all modules). To this end, the user should use the ***File*** button and selecting the ***Clear all*** function (See red rectangle in Figure 29). Note that this second functionality returns the user to the software's front page.

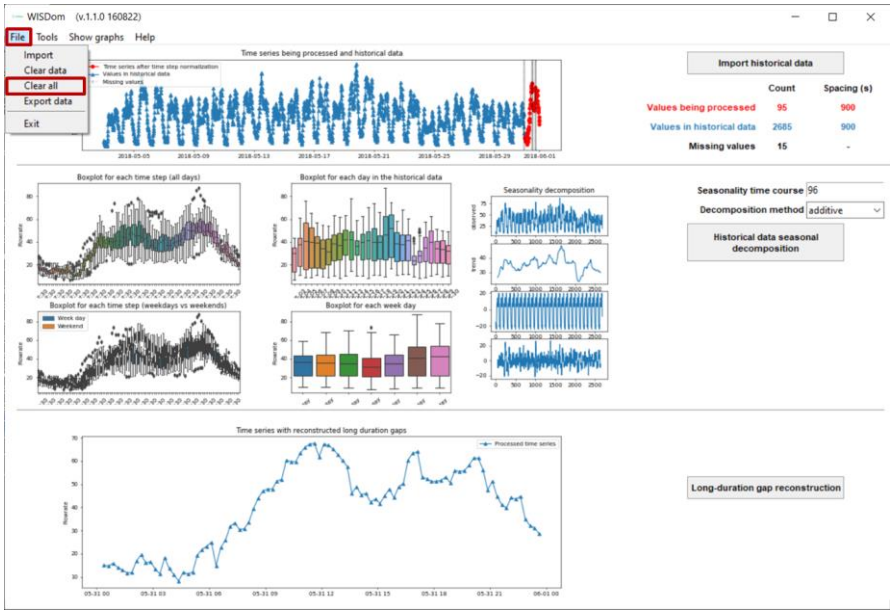

Figure 29 – Clear all data form the software.

2.11 Show and export graphs

In the WISDom software, any graph can be opened as a separate window. This way, it is possible to enlarged and export any graphs generated by the software. To open a graph on a separate window, the user should use the ***Show graphs*** function, which in turn opens a cascade menu with the module's available graphs (see red rectangle in Figure 30).

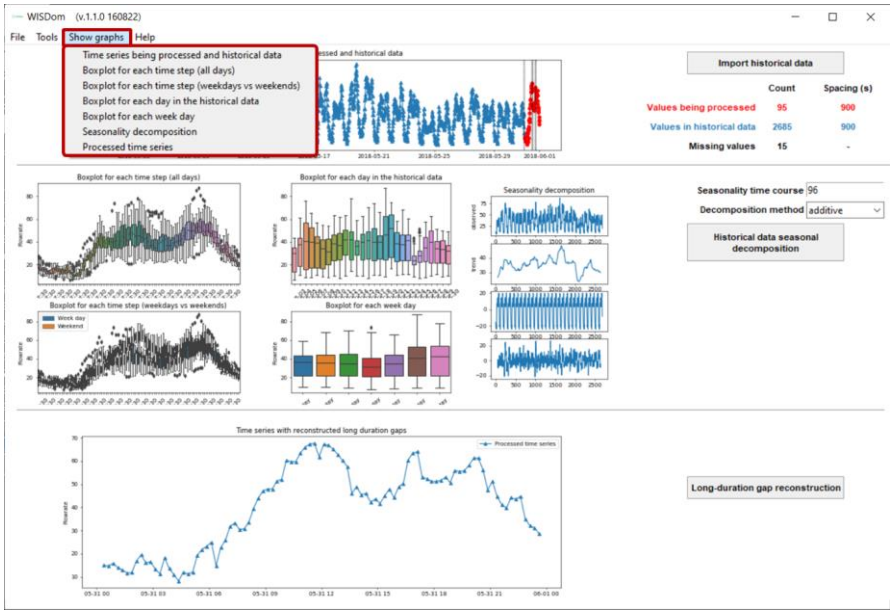

Figure 30 – Show graphs of a given module.

The desired graph will open in a separate window. From this new window it is possible to enlarge the graph or to zoom to a particular area of interest. To this end, the user should click on the 🔍

icon, as depicted in a red rectangle in Figure 31. Additionally, the user has the possibility to export the obtained graph by using the icon, as shown a green rectangle in Figure 31.

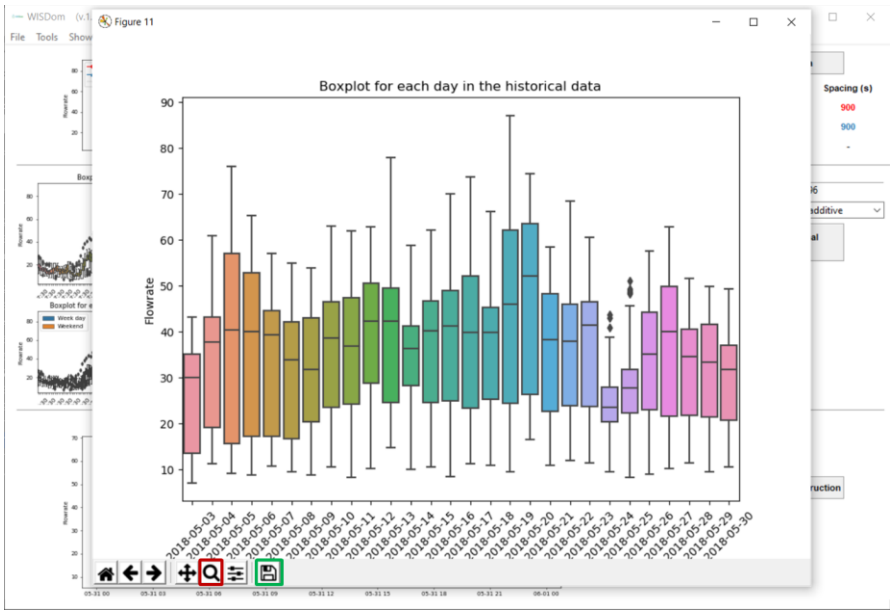

Figure 31 – Zoom and export graphs.

2.12 Editing holidays for reconstruction of long-duration gaps

The reconstruction of long-duration gaps is done using the method originally proposed by Quevedo et al. (2010)[3], with the improvements introduced in Ferreira et al. (2022)[4]. The reconstruction method takes into account distinct consumption patterns for weekdays, Saturdays and Sundays. Holidays occurring during weekdays are reconstructed using a consumption pattern of a Sunday. Thus, it is necessary to provide the software with a list of dates of Sundays.

Such list exists in the form of the ***Holidays.CSV*** file and contain the Portuguese national holidays ranging from 1970 until 2030. This file can be accessed and edited by the user using the following steps:

1. Access the WISDom folder and open the ***Holidays.CSV*** file.

[3] Quevedo, J., Puig, V., Cembrano, G., Blanch, J., Aguilar, J., Saporta, D., Benito, G., Hedo, M. and Molina, A. (2010) Validation and reconstruction of flow meter data in the Barcelona water distribution network. Control Eng Pract 18:640–651. https://doi.org/10.1016/j.conengprac.2010.03.003

[4] Ferreira, B., Carriço, N., Barreira, R., Dias, T., & Covas, D. (2022). Flowrate time series processing in engineering tools for water distribution networks. Water Resources Research, 58, e2022WR032393. https://doi.org/10.1029/2022WR032393

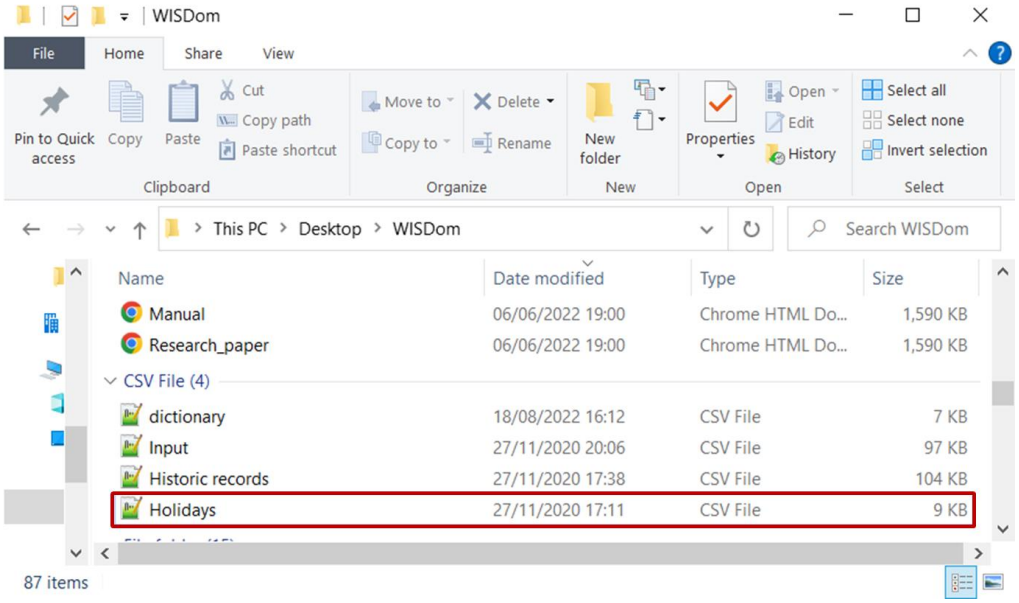

Figure 32 – Access to the ***Holidays.CSV*** file for editing.

2. Dates can be deleted, inserted or altered in the ***Holidays.CSV*** file. Note that dates should be inserted in the "yyyy-MM-dd" format[5]. After modifications, the file should be saved and the software should be reinitialized.

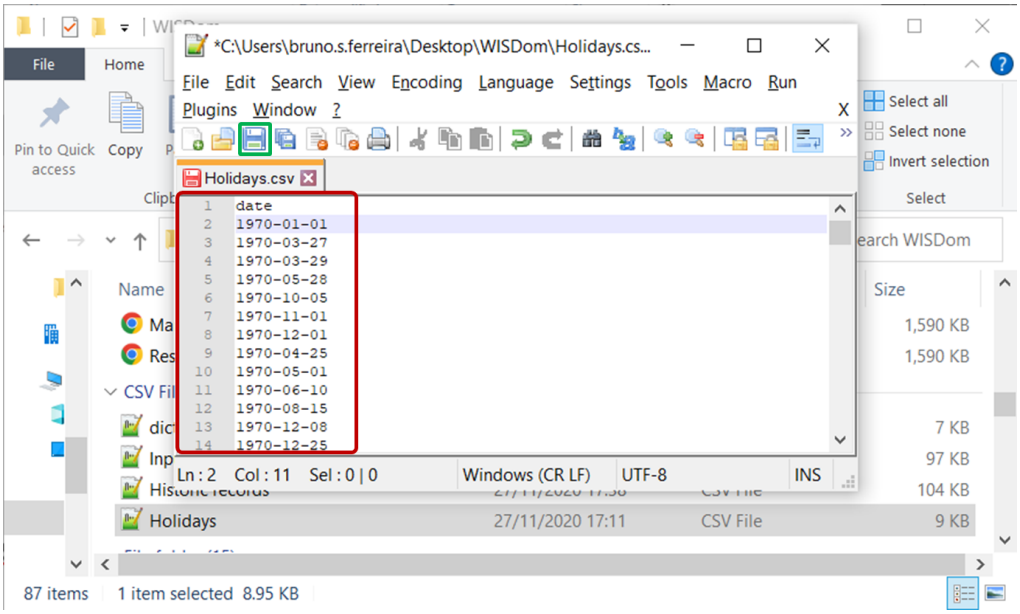

Figure 33 – Editing and saving the ***Holidays.CSV***.

[5] "2022-12-25" is an example of such timestamp format.