

Andrés Bishop Ferretti

DEVELOPMENT OF A WEB APP FOR THE IMPACTRON AND THE POST-AWS DATA SETS

Supervised by Dr. Enric Aguilar Anfrons



UNIVERSITAT
ROVIRA i VIRGILI

Degree of Geography and Territorial Management
Department of Geography
Rovira i Virgili University

Vila-seca

2018

Dedicated to

My beloved wife and daughters

DEVELOPMENT OF A WEB APP FOR THE IMPACTRON AND THE POST-AWS DATA SETS

Andrés BISHOP FERRETTI

Degree of Geography and Territorial Management

June 2018

Abstract

Long, uninterrupted, and homogeneous climate data sets essential for climatological research, applications and user services make up large data sets that are not usually readily accesible. In this dissertation POSTIMPACT is presented, a Shiny web application that allows users to interact and visualize in various forms with climatological data from the IMPACTRON and POST-AWS networks in an ubiquitous, straightforward and flexible way, fostering collaboration within researchers and stakeholders in general. The application has been successfully tested performing the analysis of some statistics, specifically the behaviour of the mean Δt_x in the Ebro Observatory where 8 segments were found to form three groups according to the characteristics of their biases, and were the magnitude of the mean differences of these groups varies along the year with Winter and Autumn months showing the minimal mean differences. Additionally, the analysis of the mean of Δt_x and the mean of Δt_n for the whole IMPACTRON data set showed a bias characterized by a slight negative sign for Δt_n , due, in essence, to the negative sign of the spring and summer mean differences.

Acknowledgements

Due to the academic nature of this work I wish to thank in the first place Dr. Enric Aguilar who guided me in the topic election of my work, and helped me in many key points of its elaboration. I feel compelled also to recognize here his great abilities not only as a researcher but also as a teacher, abilities that I have had the fortune of experience not in one but in many key subjects of my studies.

But as this work is the culmination of a bigger endeavour I also would like to deeply thank all the people, professors and administrative staff, of the Department of Geography and of the Faculty of Tourism and Geography for their support and guidance in my academic endeavour. And also would like to thank the professors from the Los Andes university in Venezuela, where I began my studies of Geography and where I got my very first foundations in Geography.

Of course for all this to happen there have to be a solid base. This solid base is beyond any doubt my family. Without their loving support and comprehension I would not have written a single word. Thanks to my beloved parents, Walter and Rosa, and a thousand thanks to my beautiful daughters, Fátima and Amina and to my adorable and brilliant wife Caribay.

Contents

Abstract	iii
Acknowledgements	iv
1 Introduction	1
1.1 Objectives	1
1.2 Context	1
1.2.1 Transitions	3
1.2.2 Transitions from Manual (MAN) to Automatic Weather Stations (AWS)	5
1.2.3 The Automatic Weather Stations (AWS)	5
1.2.4 Parallel measurements	7
1.3 The IMPACTRON and POST-AWS projects	8
1.3.1 IMPACTRON	8
1.3.2 POST-AWS-temp	9
1.4 App value	10
2 Data and Methodology	12
2.1 Shiny	12
2.2 Database and statistics description	15
3 Results	20
3.1 POSTIMPACT	20
3.2 Selected statistics analyses	24
3.2.1 Analysis of station SP82380A: Ebro Observatory	25

3.2.2	Analysis of the median Δt_x and Δt_n (IMPACTRON tab). . .	29
4	Discussion	31

List of Figures

2.1	Select box widget with generic appearance	13
2.2	Stations in mainland Spain.	17
2.3	Stations in Canary Islands (Spain).	17
2.4	IMPACTRON and POST-AWS stations network.	18
3.1	Scatter plots tab	21
3.2	Box plots tab	22
3.3	Plotly menu	22
3.4	IMPACTRON tab	23
3.5	All stations tab. Selected options correspond to the default options .	24
3.6	Summaries tab	24
3.7	Map tab. Default view	25
3.8	Info tab.	25
3.9	Map tab. Ebro Observatory location	26
3.10	Ebro Observatory scatter graphic Δtx plot	27
3.11	Ebro Observatory Δtx segments box plot	28
3.12	Ebro Observatory Δtx Seasonal box plots	29
3.13	IMPACRON median Δtx box plot	30
3.14	IMPACTRON median Δtn box plot	30

List of Tables

2.1	Resume of render functions used	15
2.2	Spanish AWS of the IMPACTRON project	16
2.3	Resume of statistics in <i>statistics.txt</i> file	18
2.4	Selected statistics in <i>analysis_txtn_qc_segmented.txt</i> file	19
3.1	Summary for Δtx ($^{\circ}C$) in Ebro Observatory.	27

1 | Introduction

This document constitutes the dissertation thesis for the Degree in Geography and Territorial Management taught by the University Rovira i Virgili. In this introductory chapter the objectives, the context, IMPACTRON and POST-AWS networks, and value of the application, are presented.

1.1 Objectives

The following objectives are proposed:

1. To develop an application that allows the interactive visualization in a web client of the database of IMPACTRON and POST-AWS datasets.
2. To analyze the biases found between MAN and AWS stations in the IMPACTRON and POST-AWS datasets studies using this app.

1.2 Context

Climatological research, applications, and user services depend, in many cases, on the availability of long, uninterrupted, and homogeneous climate data sets of meteorological/climatological data (e.g., air temperature, precipitation, humidity, atmospheric motion, atmospheric pressure, evaporation, sunshine, and present weather) (WMO, 2011).

Climatic data have been being collected worldwide since the second half of the 19th century (or even before in a limited numbers of locations) by National Meteorological and Hydrological Services (NMHSs) as well as other institutions

(e.g., in Spain, Hydrographical Confederations or in Andorra Electrical Companies). However, the availability of centennial series, especially at the daily resolution and in digital format, is restricted, except for some regions mainly located in the northern hemisphere, to a period beginning at the last third of the 20th century (when most weather information began to be collected by NMS's or other institutions) being, thus, spatially and temporally limited (Brunet and Jones, 2011; WMO, 2011).

Homogeneous climate data sets have been defined succinctly as climate time series where variations are caused only by variations in weather and climate (Aguilar et al., 2003; Venema, 2012) and are free of other influences. However, most, if not all, long term climate data sets have been affected by many non-climatic factors, such as changes in the location or the environment of weather stations, changes of the instrumentation, of averaging methods, among many others, that have introduced discontinuities, or inhomogeneities, that can lead to erroneous interpretations of the studied climate and its fluctuations (Li-Juan and Zhong-Wei, 2012; Ribeiro et al., 2016; Venema, 2012).

The frequency of such inhomogeneities varies, but a study of homogenization of instrumental western climate records point to frequency around one homogeneity break every 20 years. The same study indicates a bias in temperature trends of half degree Celsius in the period between the 1870s to the 1980s a bias similar to the long-term global climate trends (Auer et al., 2007).

Also, interruptions in climatic data sets are frequent and in many cases prevent (or should prevent) researchers to use them. Whereas Western European and North American countries are not exempt of such data gaps (Staudt et al., 2007), the problem worsens in regions such as South America, Africa, Eastern Europe and parts of Oceania and Asia. Interruptions may include a whole region for long periods of time, or specific sub-regions during short to medium periods of time. At the Expert Team on Climate Change Detection and Indices (ETCCDI) Regional Workshop held in Guayaquil (Ecuador) in January 2011, a selection of data for more than 200 TX, TN, and RR daily data was subject to analysis and participants found that many time-series contained large data gaps or missing values that compromised its

suitability for many types of climatological research (Skansi et al., 2013). In most of Eastern Europe and Western Asia big data gaps appeared when an important number of stations stopped working in the 1990's after the dissolution of the USSR (Hu et al., 2014; Syrakova and Stefanova, 2009).

One source of inhomogeneities that is increasingly affecting climatic data sets is the transition from Manual (MAN) to Automatic Weather Stations (AWS). WMO (2008) report indicates how the ratio of AWS to conventional weather stations increases persistently as the capabilities of the former increases and their operation costs represent significant savings compared with those of the observer-staffed weather stations. However, it must be mentioned that the deployment of AWSs requires strict and costly maintenance costs, including calibration, to keep the system in good functioning. This kind of transition will be further discussed later on this chapter.

1.2.1 Transitions

Weather stations and in general climate monitoring networks are inevitably subject to gradual or abrupt changes through time. As previously stated, these changes are originated by numerous non climatic events and can lead to inhomogeneities. Jones et al. (1986) group the most important causes of inhomogeneities in a climatological data set in four classes:

- Changes in station environment
- Changes in station location
- Changes in observation times and calculation methods
- Changes in exposure, measurement techniques and instruments

The type of instruments, their exposure and the techniques used to measure the collected data for most of the weather stations and of course national programs have been subjected to numerous changes through time that affect the homogeneity of its series. In the case of air temperatures several examples can illustrate this kind of changes.

The fact that the air is a poor conductor of heat have made the issue of the exposure and its changes through time one of the main concern of researchers. The necessity to protect the measurement instrument from direct or indirect radiation with the use of a screen, which in turn creates the problem of a micro climate developing within the screen due to the lack of ventilation, have lead to a plethora of screen designs in the last two centuries.

In the 19th century, instruments were typically placed on building walls facing north and protected from direct sun radiation with a metal screen. When it was realized that the temperature of the building affected the temperature of the screen, various types of screen were introduced and installed away from buildings (Venema, 2012). Among these open-air screens one can mention the introduction of, among many others, the so called French or Montsouri shelter or the Glaisher (also known as Greenwich) stand, both provided with a roof and vertical panels but fully open at the front and the bottom.

Later, the Stevenson screen was introduced, with its many variants such as the ‘Cotton Region Shelter’ in the United States. The newly introduced screen was the fruit of the design of the Scottish engineer Thomas Stevenson, who, in 1886, introduced the double louvred approach and the front an bottom closures in his design. Although the Stevenson screens coexisted with many other models, they soon became the de facto standard in many countries (Burt, 2014). In Spain, the change from the aforementioned two models of screens to the Stevenson screen was carried out, for most of the stations of the network at the time, in the mid-1910’s (Brunet et al., 2011).

The transitions between different screen systems clearly have the effect of introducing non climatic discontinuities in climatic series. Parker (1994) reported that in several comparisons between Montsouri and Stevenson screens carried out in several European locations during the 19th and 20th centuries warm biases were found. More recently Brunet et al. (2011) made an attempt to adjust the bias with respect to modern readings in a comparison study between Montsouri and Stevenson screens for Spanish stations where about 6 years of daily paired temperature observations for daily maximum (T_x) and minimum (T_n) temperature

data recorded at two experimental sites were analyzed. The study found a strong warm bias in Tx data of about 1 °C at annual scale with clear higher seasonal values in summer.

1.2.2 Transitions from Manual (MAN) to Automatic Weather Stations (AWS)

Very often changes from manned to automated systems take place in a whole network and to evaluate the biases they introduce widely used approaches has been to compare nearby stations, conduct a series of side by side measurements and/or perform instruments intercomparisons. These approaches have also produce valuable datasets, suitable for the study of the shape and size of the potential inhomogeneities that would affect a climate time series experiencing changes equal or similar to those studied.

But before detailing these approaches a few concepts relating to AWS are presented next.

1.2.3 The Automatic Weather Stations (AWS)

Since the introduction of the first sealed liquid-in-glass thermometer around 1641 (Middleton, 1969), many developments and improvements have been made including the recent adoption of the electrical resistance temperature sensor used in Automatic Weather Stations (AWS). This kind of sensor (in the case of temperature readings) is only one of many components of the AWS which is defined by WMO (1992) as a meteorological station at which observations are made and transmitted automatically. Thus, in AWS many components must work together not only to record observations automatically but also to transmit them automatically. The WMO (2011) groups these components into two large subsystems: the AWS software and the AWS hardware.

The AWS software is a costly component, especially when ill designed code that is no flexible enough to easily cope with changes, is used. It have two main components: the application software, build upon algorithms used to process the

data in accordance to the user needs (usually includes functions for conversion of sensor output to meteorological data, manual entry of observations, quality control, transmission and display, among others) and, the system software which manage the processor and the application programs and is usually put into the Central Process Unit (CPU) as a firmware, being accessible only to the manufacturer (although this is changing as users are increasingly developing their own software).

As for the AWS hardware, it is grouped into three main components:

1. A set of sensors: their requirements are not very different from those of sensors at MAN stations and in general any sensor with electrical output (analogue, digital or "intelligent") is suitable. For temperature, pure metal resistance thermometers are the most common types, being the platinum resistance thermometer (100 Ω at 0C) the preferred for its good long-term stability.
2. A CPU: although in a majority of cases it consists of one microprocessor based system installed near the sensors, depending on different requirements and circumstances the CPU functions may be executed by different units. Its main functions are data acquisition, data processing, data storage and data transmission.
3. Peripheral equipment: These are basically the power supply (designed to provide high stability and interference-free operation), the real-time clock (powered by a battery for functioning even during power outages), the built-in test equipment (monitoring automatically vital components) and, the local display and terminals (allowing observations to be entered or edited manually)

As Gilabert (2016) points out, the introduction of AWS has been accompanied with a large increase in the spatial and temporal resolution of the observations and in the real-time data availability but not without introducing systematic biases that compromise homogeneity and reduce robustness and reliability of temperature series. The differences between MAN and AWS systems or among different AWS

types could be associated with several factors (some of them not directly related to automation itself) as the technological differences among automatic sensors, screen changes, station relocation, or different response-time related to different atmospheric conditions.

1.2.4 Parallel measurements

A common approach used to evaluate changes that take place in the whole network, as is the case with the introduction of AWS, has been to perform a series of parallel measurements and/or conduct instruments inter comparisons, procedures that as a by side product leave valuable data sets.

In the parallel approach the objective is to set the new configuration maintaining as much as possible the old set-up, documenting the elements that have changed, and keeping the operations of the old set-up for an overlap period (Baddour and Kontongomde, 2007).

As this approach will require maintaining two set-ups at each site undergoing the change, its associated costs are high. Nevertheless references to comparisons are found as early as 1864 and it is identified as a good practice by the WMO (Aguilar et al., 2003).

Alternatively, paired observations can be selected as an approach when parallel observations are not feasible or possible. This approach consists in reproducing, as much as possible, the old set-up and take paired observations using a new set-up for a period of time, ideally covering the full range of climatic seasons(Baddour and Kontongomde, 2007).

1.3 The IMPACTRON and POST-AWS projects

In Spain, the State Meteorological Agency (AEMET), currently has more than 850 automatic weather stations throughout Spain and the Meteorological Service of Catalonia, currently has 187 automatic stations (METEOCAT, 2017). IMPACTRON and POST-AWS are projects that aim to better understand the inhomogeneities caused by the transition from MAN to AWS weather stations by compiling a database with parallel measurements and analysing it.

1.3.1 IMPACTRON

The Assesment of the Impact Over Air Temperature Series of the Transition Between Observation System (IMPACTRON hereafter) was a scientific network which lasted between 2015 and 2017 and funded by the Spanish Ministry of Economics and Competitiveness (MINECO), which involved the Spanish State Metereological Agengy (AEMET) , the Meteorologic Service of Catalonia, the University of Zaragoza, the Cantabria Univerity, the Ebro Observatory, and the Rovira i Virgili University (Aguilar, 2015).

Its main objective was to document and analyze the effects over the air temperature series of the most relevant transitions of observing systems in the Spanish observational network. To accomplish this, firstly an identification and analyzes of relevant parallel measurements was done, followed by "a comprehensive description of the effect of each transition over the statistical distribution of air temperatures" and an evaluation of "its impact over the climate series at different resolutions".

The IMPACTRON network allows to reduce the effects of transitions over the Spanish Network temperature series making improvements for its use in climatic analyzes, contributing thus to improve the description of the climatic change and variability. It does so by offering an objective assessment of the bias introduced by transitions in climatic series, enabling the inclusion of such biases in benchmarking processes, providing a complementary approach to purely statistical homogenization, making possible more adjusted climatic analyzes, and by

contributing with indications to the meteorological instrument manufacturers and the managers of the observational networks about the impact of the changes introduced.

1.3.2 POST-AWS-temp

The Parallel Observations Science Team (POST) is an independent project that aims to compile a global database with parallel measurements, in order to better understand inhomogeneities that distort the climate signal and make the assessment of trends and variability more difficult. As the specific transitions depends on local climate, answers to questions about the systematic large-scale biases produced by transitions need a large global parallel dataset. POST have several ongoing studies:

- POST-AWS-temp. Studying the influence of automation on temperature.
- POST-AWS-precip. Studying the influence of automation on precipitation.
- POST-early. Versing about the temperature change due to the transition of early screens to Stevenson screens.
- POST-move. On the influence of relocation on temperature.

POST-AWS-temp has gathered up until now data from from 10 countries: Argentina (9 pairs), Australia (13), Brazil (4), Israel (5), Kyrgyzstan (1), Peru (31), Slovenia (3), Spain (46), Sweden (8), USA (6). POST-AWS-temp, and in general POST, expect to gather enough data to asses station settings impact and also understand whether different climates produce different biases or not.

1.4 App value

Static graphical representation can pose serious limitations as the amount of data to be represented grows in number and complexity. It presents the caveat of predefined graphics that often do not satisfy the needs of users, or do not allow them to choose specific data to be represented. On the other hand, dynamic ways of representing big data sets allow users to select and interact, at different levels, with such data in more efficient and enriching ways.

One common way to provide dynamic representations has been the building of applications that provide online portals, but this approach usually requires complex data interactions and a combination of, at least, a development of HTML, CSS, or Java tools (Ellis and Merdian, 2015). More recently, *Shiny*, a technology created by RStudio that only requires knowledge in the R programming language, permits the move from static to dynamic graphic representation in a relatively straightforward way.

While *Shiny* has been mostly used in the fields of statistics, biology, and even phsicology, to process, analyze and graphically convey data or to create educational tools (Bahar et al., 2017; Doi et al., 2016; Ellis and Merdian, 2015), it uses also extend to the field of Geography and Environmental Sciences, with applications made for geo-spatial analyzes of patient data (Moraga, 2017), mapping of species (Feng et al., 2017) or building population pyramids for educational purposes(Zaragozí et al., 2016) to cite some examples.

In Climatology studies, one can find some examples of web applications tools, some of them developed with Shiny, as is the case of WebXTREME, an online tool for the evaluation of indices of climatic extremes (Klein et al., 2017), or the Terrestrial Precipitation Analysis (TPA) a set of web tools packages that allows researchers to calculate precipitation statistics (Lemoine et al., 2016).

In this work we present the POSTIMPACT App, an interactive web application. Implemented using the Shiny framework, POSTIMPACT has the ability to let users easily create various types of plots to represent the data available from the project without the need of running any command in R, making it suitable for the use of

people without any programming skills.

This facilitates the assessment of such data and make it available globally on any web browser where users can navigate through data creating interactive visualizations. Additionally, the various libraries available to enhance the plots represent a significant added value, as is the case of the *plotly* library, that allows users to interact with the graphics in various ways, facilitating visual analyses. This and other libraries used will be further discussed in the Data and Methodology chapter.

2 | Data and Methodology

2.1 Shiny

POSTIMPACT was built using *Shiny* (Chang et al., 2017), an elegant, simple and powerful extension to the R programming Language designed to create interactive applications oriented to the analysis and visualization of data. One of the main advantages of Shiny is that in order to build applications there is no need of any knowledge of HTML, CCS, or JavaScript (although enhancements can be done, if desired, through the use of those languages). Another great advantage is that *Shiny* can make use of all the powerful functions of R.

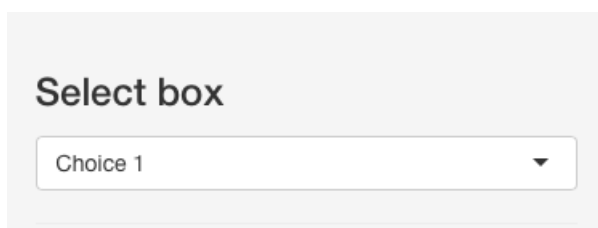
Shiny applications can be executed locally and can be hosted either locally or in an external server. To host and run locally, the user have to install the *Shiny* package in the R environment. Installing the *Shiny* package is also the way to develop any Shiny application. To execute an application from a server the user just need a web browser. Besides the *Shiny* package itself, numerous other packages add functionalities to *Shiny*, and are either specifically designed to be used with it or are packages that add general enhancements to R.

Any *Shiny* application must have at least two components: the file *ui.R* and the file *server.R*. These two files must be contained in the same directory along with other documents that may be necessary for the application to work, such as data sets or images. The file *ui.R* contains the instructions that define the user interface of the application, and therefore controls its design and its appearance. The file *server.R* contains the instructions that the computer needs to build the application, that is, the instructions that will make the necessary calculations and graphs.

The user interface (*ui.R*) explores the different data processing options encoded

in the controller (*server.R*). To build the user interface, Shiny offers a wide range of widgets or control elements (such as buttons, lists, sliders, radio buttons, check boxes, text, file up loaders, action buttons, dates and select boxes, among others), which automatically give the desired appearance and are required for the handling and visualization of the data. POSTIMPACT makes use of just one of such widgets: select box (figure 2.1) which opens a drop-down list with several options where only one can be selected.. It also provides different layouts, so tabs and navigation menus can be implemented and the output window can be adjusted to the desired appearance specifying, with the use of output functions, the location of each of the output element (e.g. tables, graphs).

Fig. 2.1. Select box widget with generic appearance



The outputs can be anything R alone could output, like tables, plots or plain text. However, many packages extend the range of possible outputs available. Some of them make JavaScript implementations work in R, allowing users to interact with the outputs (e.g. selecting a point in a plot) or create maps. POSTIMPACT makes use of two of such libraries: *plotly*, and *leaflet*.

The *plotly* package connects R to the open source JavaScript graphing library allowing with just one line of code the composing, editing, and sharing of interactive data plots. *Plotly's* R library is free and open-source.

Leaflet is a very popular open-source JavaScript library for interactive maps (Cheng et al., 2018). The R *Leaflet* package makes it easy to integrate and control *Leaflet* maps in R. Such maps have powerful features as:

1. Controls and interactions: overview, zoom slider, zoom in / out buttons, scale bar, map shift, zoom and rotation, selection of entities.
2. Definition of styles and personalization: styles of entities: points, lines,

polygons and icons, customizable appearance of the controls using CSS3.

3. Events : great flexibility to create pop up descriptions and markers
4. Overlays: linking of listener functions to attend events on the map

Leaflet basemap in POSTIMPACT uses a slightly modified *Mapbox* (a large provider of custom online maps) *NorthStar* tiles style.

The file `server.R` works in a similar way to the input. Within each output functions (related to a corresponding output function in *ui.R*), the server code defines the outputs to be displayed and, through the previous inputs, executes the code and displays in the application window the pertinent output corresponding to these inputs.

The exchange of data between the *ui.R* and the *server.R* files is done automatically, so when any value of the menu is modified (update input) the part of the code that deals with it is rerun and the new output is showed; this is what is referred to as "reactive programming" (Wan and Hudak, 2000) a feature that allows programmers to build, trough the use of reactive functions, an interface that responds immediately to any change that the user makes. This implies that any part of the code which needs to rerun due to interaction, will have to be within a reactive object.

The functions to be used are decided according to the type of output that is wanted and they have to be adapted to the definitions of inputs and outputs in the *ui.R*. Normally the renders (table `ta1`) are used, without the need to create a reactive concrete variable, since, being reactive functions, they allow to update the outputs directly. However, in some cases a reactive variable is declared and then used within the render function.

FUNCTION	DESCRIPTION
<i>renderPlot</i>	generates a plot
<i>renderPlotly</i>	generates a <i>plotly</i> plot
<i>renderPrint</i>	prints output and/or return a printable R object
<i>renderUI</i>	returns a dynamically generated object
<i>renderDataTable</i>	returns a data frame or a matrix
<i>renderLeaflet</i>	returns a <i>leaflet</i> map object

Table 2.1: Resume of render functions used

2.2 Database and statistics description

For the POSTIMPACT, data gathered for the IMPACTRON project is used. The data is divided in three *.txt* files: *AWSstations.txt*, *statistics.txt* and, *analysis_txtn_qc_segmented.txt*. All of them have rows for observations and variables for columns. The first row of each file contains the name of the variables and the data is separated by blank spaces (one or more spaces, tabulators, ...), and the decimal separator is a point.

The *AWSstations.txt* file contains the station identifiers, the latitude and longitude (table 2.2) of 51 Spanish stations that under the IMPACTRON project have made parallel measures between AWS an MAN stations (Figures 2.2 and 2.3). Figure 2.4 shows a World map displaying the location of the AWS in 10 countries for which data data is available for the POST-AWS project.

The *statistics.txt* file have the daily data of each station for several statistics along with data for the year, month, day, segment and, season. Table 2.3 resume the statistics contained in this file.

Finally, the *analysis_txtn_qc_segmented.txt* have the data of various statistics grouped by station, season and segment. Table 2.4 shows the statistics contained in this file.

STATION	LAT	LON	STATION	LAT	LON
SP81810A	41.28	2.07	SP81840B	41.68	2.80
SP80530A	42.57	-6.60	SP81840C	42.32	2.10
SP81480A	41.77	2.48	SP81840D	42.38	3.00
SP81810F	41.84	2.81	SP80110A	43.57	-6.03
SP81410A	41.63	-4.75	SP80020A	43.30	-8.37
SP82100A	40.65	-4.67	SP81810B	41.70	2.20
SP82130A	40.94	-4.12	SP827200	39.89	-4.03
SP81300A	41.52	-5.73	SP826100	39.47	-6.33
SP82020A	40.95	-5.50	SP81840B	42.26	2.37
SP82320A	40.84	-1.87	SP838300	37.27	-6.90
SP81810C	42.32	2.17	SP841900	37.18	-3.78
SP81810D	42.11	2.21	SP848700	36.84	-2.35
SP81810E	41.93	2.25	SP84330B	37.77	-0.80
SP83480A	38.99	-3.92	SP828000	39.00	-1.85
SP84100A	37.84	-4.84	SP823500	40.35	-1.12
SP83910A	37.42	-5.87	SP828400	39.48	-0.47
SP84510A	36.75	-6.05	SP828600	39.95	-0.07
SP84330A	37.78	-0.80	SP808000	42.87	-2.73
SP84300A	37.98	-1.12	SP808500	42.77	-1.65
SP83590A	38.37	-0.48	SP81710B	41.63	0.90
SP82310A	40.07	-2.13	SP81760B	40.96	0.32
SP80840A	42.45	-2.33	SP81760C	41.25	0.73
SP81710A	42.35	1.43	SP60010A	28.63	-17.75
SP81710B	41.35	0.67	SP60015A	28.47	-16.32
SP81750A	41.15	1.17	SP60001A	27.82	-17.89
SP81710C	41.69	1.20			

Table 2.2: Spanish AWS of the IMPACTRON project

Fig. 2.2. Stations in mainland Spain.

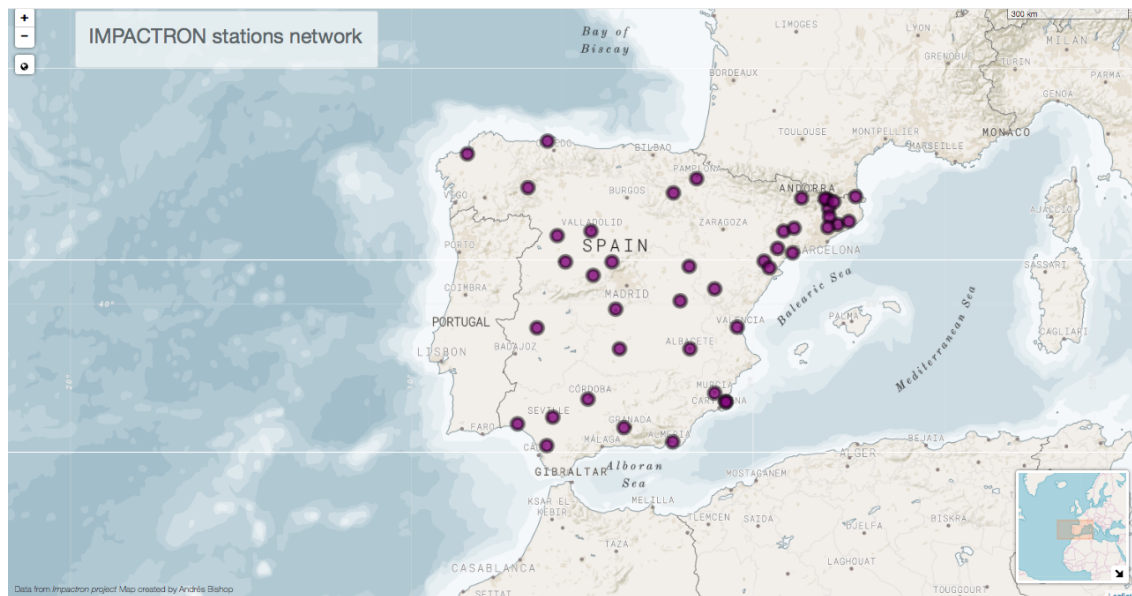


Fig. 2.3. Stations in Canary Islands (Spain).

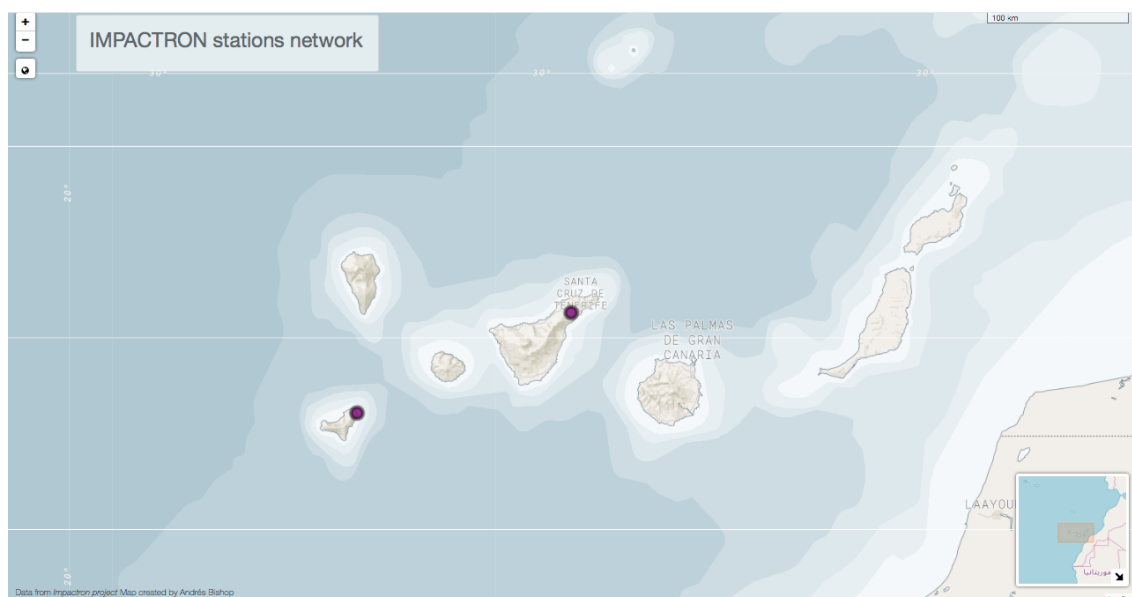
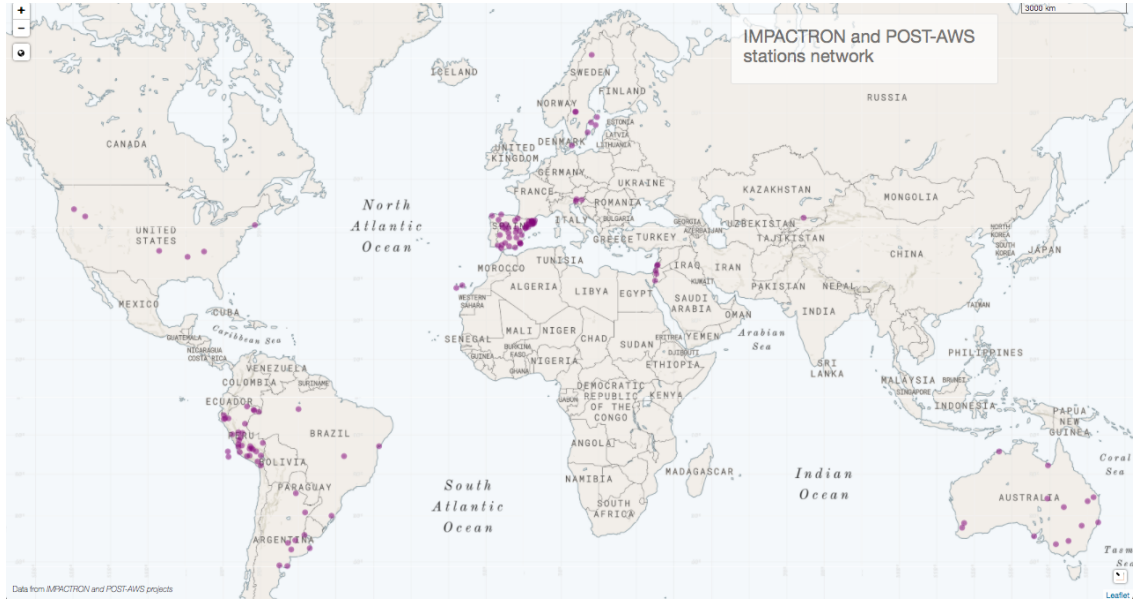


Fig. 2.4. IMPACTRON and POST-AWS stations network.

NAME	CONTENT	NAME	CONTENT
tx	maximum temperature	qcdtn	quality control
qctx	quality control	tm	mean temperature
(Δ tx)	difference in Tx	(Δ tm)	difference in Tm
qcdtx	quality control	pc	precipitatioN
tn	minimum temperature	qcpc	quality control
qctn	quality control	(Δ rn)	difference in Rn
(Δ tn)	difference in Tn	qcder	quality control
rn			

Table 2.3: Resume of statistics in *statistics.txt* file

NAME	CONTENT
bias	corresponding bias
rmse	Root Mean Sq. Error
p10	percentile 10
p25	percentile 25
median	corresponding median
p75	percentile 75
p90	percentile 90
IQR	Interquartile range

Table 2.4: Selected statistics in *analysis_txtn_qc_segmented.txt* file

3 | Results

3.1 POSTIMPACT

POSTIMPACT allows users to visualize data from each individual station of the IMPACTRON and POST-AWS projects through scatter plots and box plots, presents a box plot graphic for the mean of all stations seasonally, shows the data of all the statistics of every station and, displays a map where users can locate all of the stations and select any of them to visualize a basic box plot of it. All of these features can be seen simply by clicking on four tabs and selecting items from various drop-down lists. Furthermore, the application allows user interactions such as map padding and zooming, table filtering, and plots download, among many others.

The application consists of seven tabs:

(1) a "Scatter plots" tab (figure 3.1) where user can select from a drop-down list named "Choose a station to display" a particular station and also can select from an other drop-down list named "Choose a variable to display" the statistic to be shown in the form of a scatter plot. Once these two arguments are chosen, an scatter plot is drawn. The x axis of the graph represents the data through time, from the first measure taken to the last, while the y axis represent the value of the statistic for each each single measure taken.

The plot also shows a black dotted line across the zero value of the y axis and several red dotted lines (its number depending of the number of segments of the selected station) across the mean of each each segment. Each of the data values represented by a point can be either a quality control accepted, suspicious or very suspicious value (blue, orange and red dots respectively).

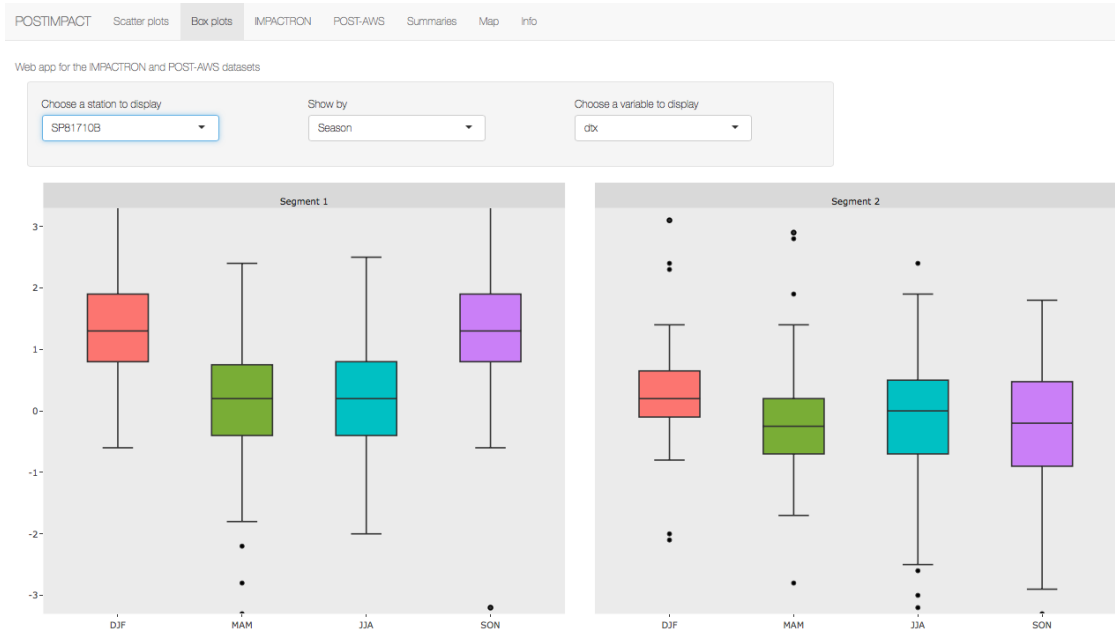
Below the scatter plot a statistical summary of each of the segments gives information about the minimal value, the first quarter value, the median, the mean, the third quarter and the maximum value.

Fig. 3.1. Scatter plots tab



(2) a Box plots tab (figure 3.2) that presents the user with box plots of the selected station data. As in the previous tab, the selection is made by means of drop-down list, in this case labelled "Choose a station to display", "Show by" and "Choose a variable to display". Here the "Show by" drop-down list presents three choices: month, season or segment. Each one of them allows the selected station data to be grouped accordingly. Worth mentioning is that for the box plots in this tab, only (Δtx) , (Δtm) , (Δtn) and (Δrn) as these statistics are at the key tools of the analyses.

Also as in the previous tab a black dotted line across the zero value of the y axis is present. Thanks to the *plotly* library, there are some added functionalities in the plots. When the cursor is rolled over the presented graphics, a menu (figure 3.3) appears at the top right corner. Within this menu the user can click on different icons to: download plot as *png*; zoom; pan; zoom in; zoom out; auto scale; reset axes; and toggle spike lines. Additionally, the minimal value, the first quarter value, the median, the mean, the third quarter and the maximum value are show for each

Fig. 3.2. Box plots tab

box when the cursor is placed over it.

Fig. 3.3. Plotly menu

(3) an IMPACTRON tab (figure 3.5) representing the data not by station this time but by statistic for the whole set of stations. The resulting box plot has similar functionalities to the box plots of the previous tab.

(4) a POST-AWS tab (figure 3.5) representing the data of the POST-AWS by country and by the mean of the following variables: (Δtx) , (Δtn) , (Δtm) and (Δtr) . The resulting box plot has similar functionalities to the box plots of the previous tab.

(5) a Summaries tab (figure 3.6).Presents a table that displays the information of biases for each segment of the selected station. The user can filter the information by performing a search in any of the columns or in the entire table. It is also possible to sort the rows into ascending or descending order of the values of any of the columns.

(6) a Map tab (figure 3.7). This tab shows a modified version of the *Mapbox North Star* style zoomed to the area of study and made with Leaflet. At the initial view, 6 clusters of stations can be seen: five over the Iberic Peninsula and one over

Fig. 3.4. IMPACTRON tab



the Canary Islands. At the right bottom a collapsible mini map displays the relative position and at the top left three buttons allow to zoom in or zoom out. Next to these buttons there is a floating panel labelled "Station explorer" with a drop-down list widget labelled "Choose a variable to display" and below a box plot of the variable selected for the active station.

The user can click on any of the station clusters, seen as coloured circles with a number, or zoom in, and as the map change its scale, each individual station appears in the form of purple dots. If these purple dots are clicked, a pop up message with the code of the station emerge and, in the floating window, the box plot actualizes showing the data for the new station.

(7) an Info tab containing information about the application use (figure 3.8).

Fig. 3.5. All stations tab. Selected options correspond to the default options

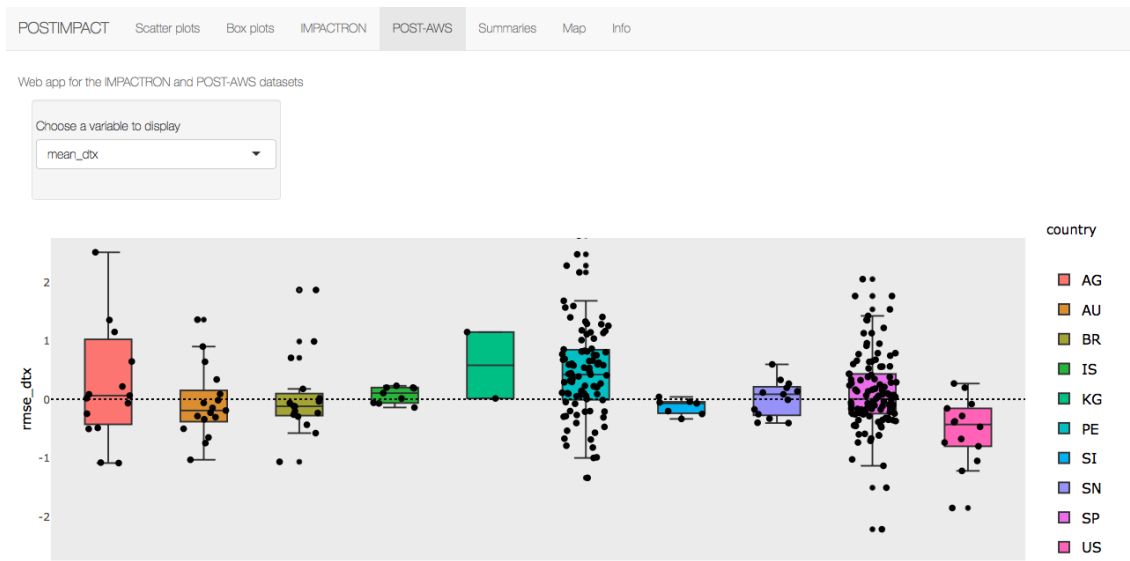
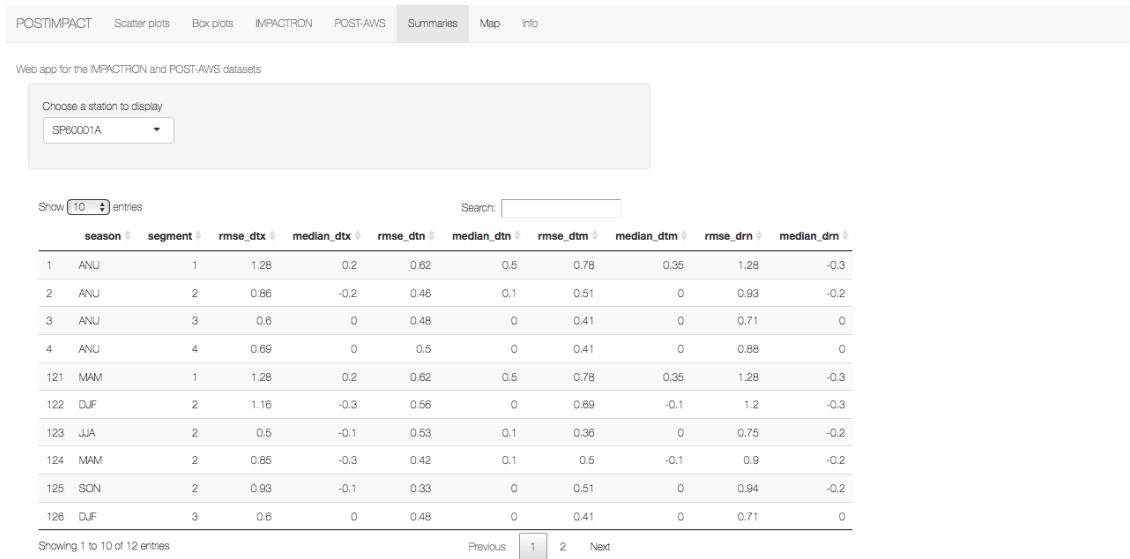


Fig. 3.6. Summaries tab



3.2 Selected statistics analyses

To illustrate the use of the POSTIMPACT an analysis of station SP82380A is presented followed by an analysis of the median differences in the maximum temperature (Δtx) for the whole of the stations of the IMPACTRON network.

Fig. 3.7. Map tab. Default view

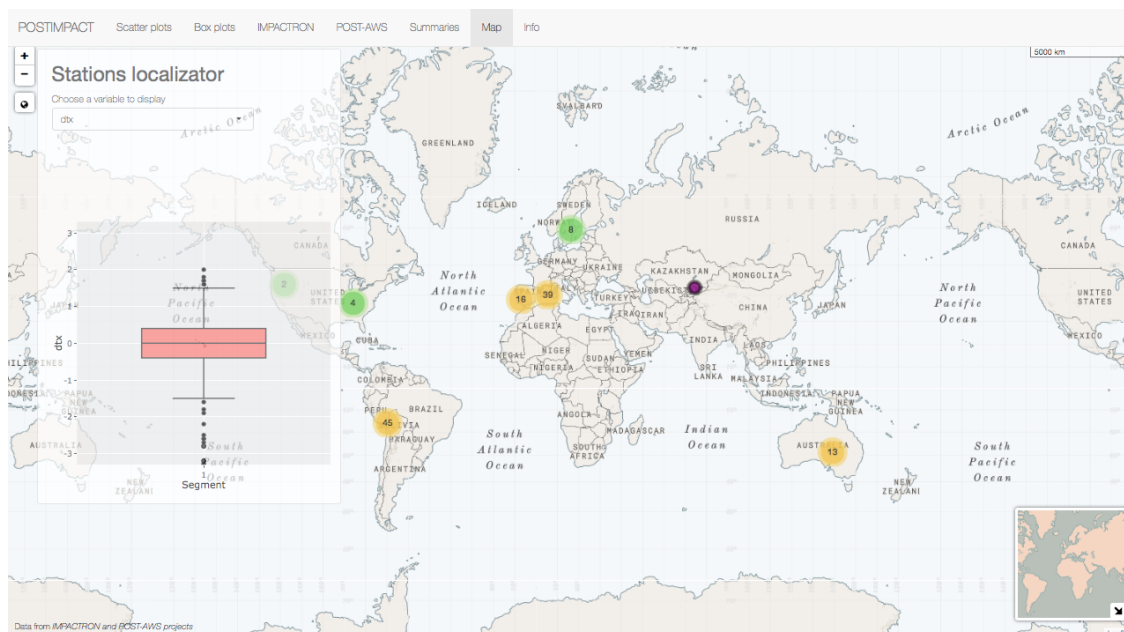
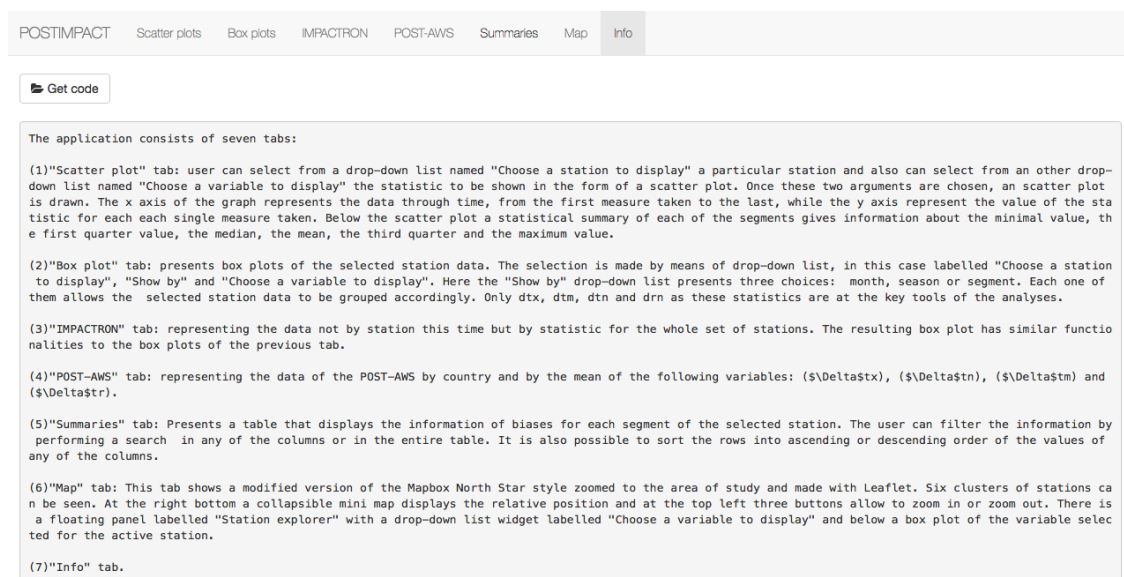


Fig. 3.8. Info tab.

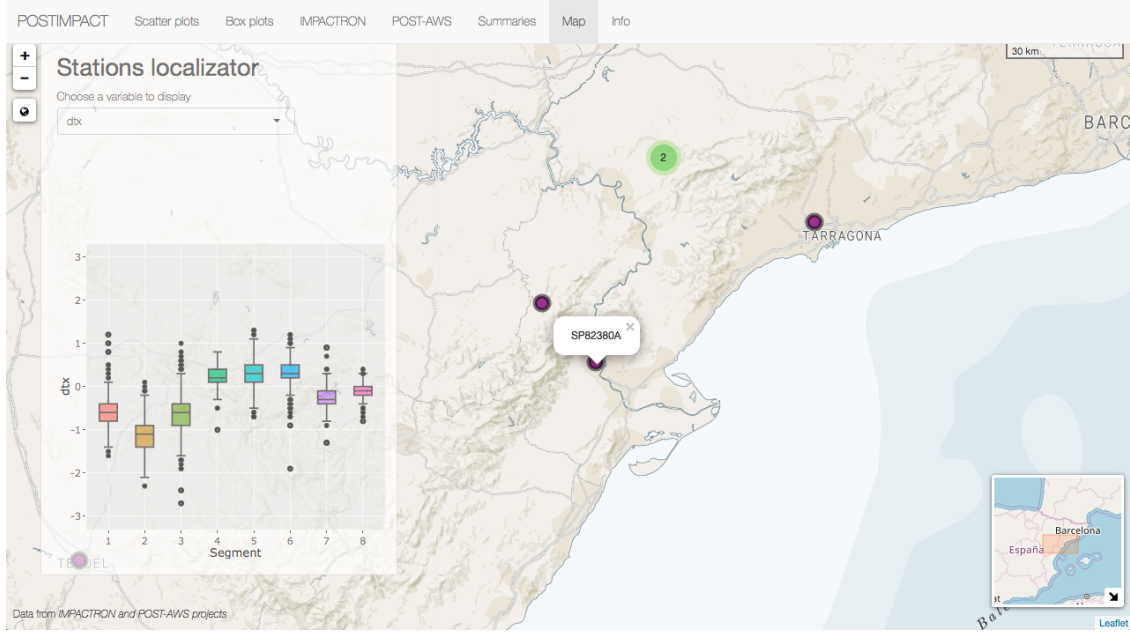


3.2.1 Analysis of station SP82380A: Ebro Observatory

The Ebro Observatory, represented as station SP82380A in POSTIMPACT, is located near the Spanish town of Roquetes (approx. 8.000 inhabitants), at latitude 40.82 N and longitude 0.49 W, on a hilltop surrounded mainly by Mediterranean

vegetation and have a continuous climate record since 1880 (from 1905 onward at the same site) (Gilabert, 2016). Figure 3.9 show its location and a preview of its corresponding box plots for the differences in daily maximum temperature (Δt_x) series between AWS and MAN in the Station explorer floating window.

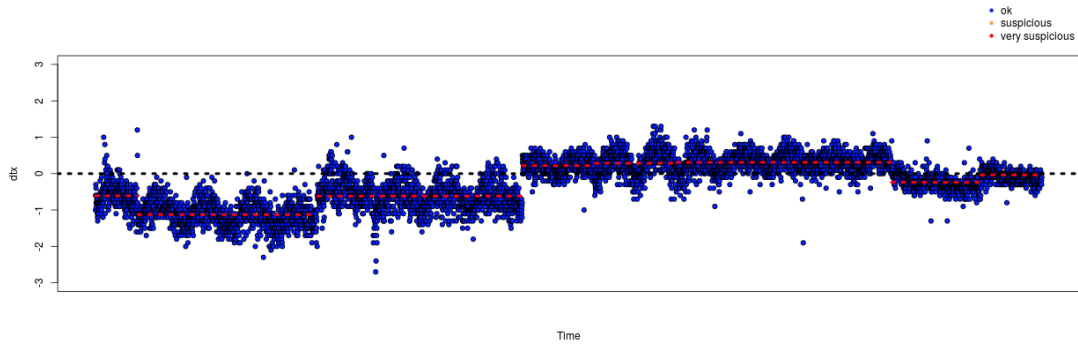
Fig. 3.9. Map tab. Ebro Observatory location



In the IMPACTRON data set The Ebro Observatory has 8 segments and 7,085 valid observations, that is, observations that have passed a quality control (QC). The time span of the overlapped observations included in the data goes from 07-01-1991 to 04-14-2015, that is roughly 24 years. According to Gilabert (2016) during this period, AEMET installed three different AWS: SEAC, SOSS and ESOS AWSs, and after statistical break points detection and visual inspection 8 validated break points that could be partially explained by the known changes in the AWS instrumentation were detected: 2 for SEAC, 3 for SOSS, and 3 for ESOS AWS's.

Figure 3.10 correspond to the plot presented by the Scatter plot tab for Ebro Observatory Δt_x . The horizontal red dotted lines indicate the AWS-MAN Δt_x mean for each segment. Dots represent observations that have undoubtedly passed QC (blue dots), or with some degree of doubt (yellow and red dots).

This plot presents a first visual approach to the magnitude and direction of the biases as it shows how, of the eight segments, five present a negative mean Δt_x and

Fig. 3.10. Ebro Observatory scatter graphic Δtx plot

SEGMENT	Min	1st Qu	Median	Mean	3rd Qu	Max
1	-1.60	-0.80	-0.60	-0.61	-0.40	1.20
2	-2.30	-1.40	-1.10	-1.13	-0.90	0.10
3	-2.70	-0.90	-0.60	-0.62	-0.40	1.00
4	-1.00	0.10	0.20	0.22	0.40	0.80
5	-0.70	0.10	0.30	0.28	0.50	1.30
6	-1.90	0.20	0.30	0.31	0.50	1.20
7	-1.30	-0.40	-0.30	-0.24	-0.10	0.90
8	-0.80	-0.20	-0.10	-0.04	0.00	24.60

Table 3.1: Summary for Δtx ($^{\circ}C$) in Ebro Observatory.

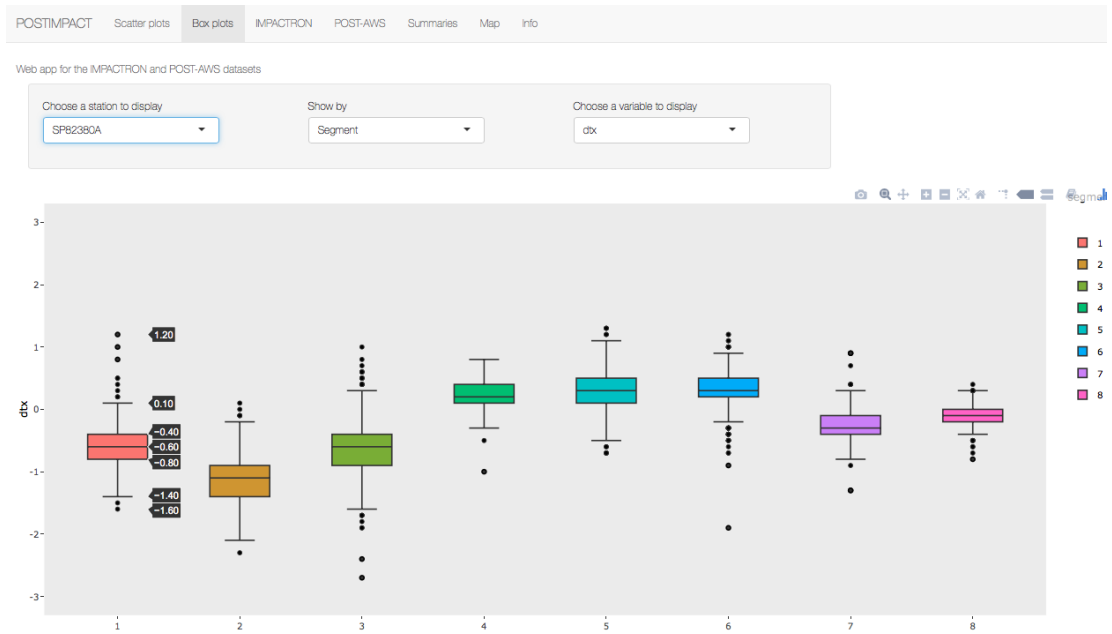
the other three a positive mean Δtx .

The summaries for each of the segments (presented in table 3.1) give more detail of the biases. It can be observed that segments can be grouped according to size and sign of their Δtx 's mean. For segments 1, 2 and 3, the mean is negative and always bigger than 0.6. Segments 4, 5 and 6 have a mean difference of positive sign that goes from 0.22 $^{\circ}C$ (segment 4) to 0.28 $^{\circ}C$ (segment 6). Finally, segments 7 and 8 have negative differences but in a smaller magnitude (-0.24 $^{\circ}C$ for segment 7 and -0.04 $^{\circ}C$ for segment 8).

Figure 3.11 shows the Box plot tab for Ebro Observatory, with the "Segment" option in the drop down list "Order by" selected and the cursor hovering over the segment 1 which triggers the apparition of the same information of table 3.1), this

time only for the selected segment. All in all, this box plot allow users to see in a graphical way the behaviour of Δtx . Worth noting is that in the box plot the distribution of the differences can be appreciated. In the case of segment 2, for example, is evident that the number of differences falling between the 1st quarter and the mean is greater than those falling between the mean and the 3rd quarter.

Fig. 3.11. Ebro Observatory Δtx segments box plot



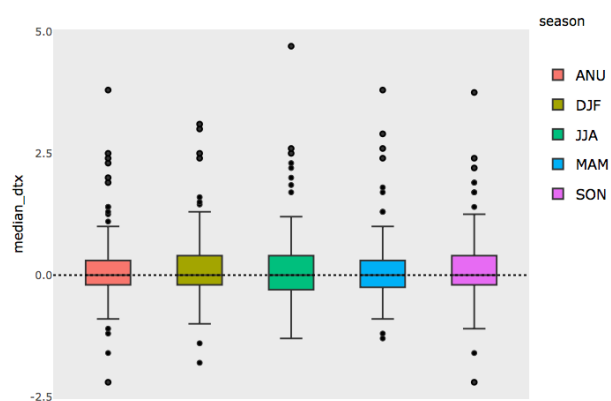
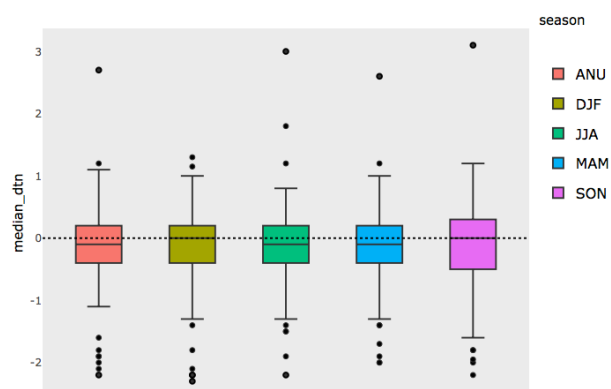
The seasonal box plots, also in the Box plot tab (figure 3.12), show that differences varies along the year (DJF winter, MAM spring, JJA summer and SON autumn). For segments 1,2 and 3, the negative AWS bias is larger in magnitude for Summer months (June-August) and the largest difference in mean temperature occurs in segment 3 (-1,30 °C). Winter and Autumn months show the minimal mean differences. Segments 4, 5 and 6 all exhibit a positive AWS bias is positive in a lower magnitude than for the aforementioned segments. Winter months show the larger AWS bias. The last two segments (7 and 8), present a much more smaller difference in Δtx and a much less clear yearly variation.

Fig. 3.12. Ebro Observatory Δtx Seasonal box plots

3.2.2 Analysis of the median Δtx and Δtn (IMPACTRON tab).

In this case the bias of individual stations are summarized with box plots. Figures 3.13 and 3.14 show the box plots for the differences in the median Δtx and in the median Δtn respectively. The plots are for annual (anu), winter (DJF), spring (MAM), summer (JJA) and autumn (SON) periods.

For the differences in the median Δtx the plots for all periods show in average no biases in the data set, although it seems to be quite a lot of outliers in the positive side. As for the differences in the median Δtn , the box plot for the annual data has a slight negative sign as a consequence of the negative sign of the spring and summer plots and the neutrality of the autumn and winter plots.

Fig. 3.13. IMPACRON median Δt_x box plot**Fig. 3.14.** IMPACTRON median Δt_n box plot

4 | Discussion

POSTIMPACT ¹ provides a solution to visualize and interact with IMPACTRON and POST-AWS data sets and does not require any knowledge of R nor any special software installation. It shows how dynamic visualizations tools remain a very useful alternative to display and communicate big and complex data sets, such as the ones used in climatological studies, in an ubiquitous, straightforward and flexible way, fostering collaboration within researchers and stakeholders in general.

The implementation of the app with R and Shiny was justified by the increasing usage of the first for statistical data analysis and the powerful, elegant and easy to implement web visualization programming environment the later have.

The application has been successfully tested performing the analysis of some statistics. Specifically the behaviour of the mean Δtx in the Ebro Observatory was carried out and differences of varied sign and magnitude were observed. Succinctly it can be said that according to the mean of Δtx , the segments can be classified in three groups according to the characteristics of the biases, and that the magnitude of the mean differences of these groups varies along the year with Winter and Autumn months showing the minimal mean differences.

Additionally the mean of Δtx and the mean of Δtn for the whole IMPACTRON data set were analyzed and a bias characterized by a slight negative sign for Δtn was observed, due, in essence, to the negative sign of the spring and summer mean differences.

Therefore, it can be said that the objectives set in chapter one have been achieved: An application for the interactive visualization in a web client of the

¹available at <https://bishopf.shinyapps.io/Impactron/>

database of IMPACTRON and POST-AWS datasets have successfully developed and, using this app, analyzes of the biases between MAN and AWS stations in the IMPACTRON and POST-AWS data sets were performed.

Furthermore, during the process of preparing this work, various transversal competences have been acquired or further developed by the author, such as: critical, logical and creative thinking, and an ability to innovate; autonomy, responsibility and initiative; and, an advanced knowledge of information and communication technologies, specifically of a fundamental programming language for climatological research.

POSTIMPACT can be extended and improved in future versions increasing its flexibility and enabling more options. Not pretending to be exhaustive, several enhancements for future versions are suggested: A more fluid progression from one tab to another, specially from the map tab to others; a more sophisticated table tab with more statistics to show and the possibility of downloading data; a dates selection options; options to select countries in the POST-AWS tab, and an interactive help and guidance tab.

Bibliography

- Aguilar, E. (2015). Evaluación del impacto en las series de temperatura de las transiciones entre sistemas de observación. Memoria científico técnica.
- Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J. (2003). GUIDELINES ON CLIMATE METADATA AND HOMOGENIZATION. Technical report, World Meteorological Organization, Geneva.
- Auer, I., Böhm, R., Jurkovic, A., Lipa, W., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Matulla, C., Briffa, K., Jones, P., Efthymiadis, D., Brunetti, M., Nanni, T., Maugeri, M., Mercalli, L., Mestre, O., Moisselin, J. M., Begert, M., Müller-Westermeier, G., Kveton, V., Bochnicek, O., Stastny, P., Lapin, M., Szalai, S., Szentimrey, T., Cegnar, T., Dolinar, M., Gajic-Capka, M., Zaninovic, K., Majstorovic, Z., and Nieplova, E. (2007). HISTALP - Historical instrumental climatological surface time series of the Greater Alpine Region. *International Journal of Climatology*.
- Baddour, O. and Kontongomde, H. (2007). GUIDELINES FOR MANAGING CHANGES IN CLIMATE. *Environment*, (62).
- Bahar, B., Tuncel, A. F., Holmes, E. W., and Holmes, D. T. (2017). An interactive website for analytical method comparison and bias estimation. *Clinical Biochemistry*, 50(18):1025–1029.
- Brunet, M., Asin, J., Sigró, J., Bañón, M., García, F., Aguilar, E., Palenzuela, J. E., Peterson, T. C., and Jones, P. (2011). The minimization of the screen bias from ancient Western Mediterranean air temperature records: An exploratory statistical analysis. *International Journal of Climatology*, (31).

- Brunet, M. and Jones, P. (2011). Data rescue initiatives: Bringing historical climate data into the 21st century. *Climate Research*, 47(1-2):29–40.
- Burt, S. (2014). What do we mean by ‘air temperature’? (Retrieved on 2018-04-15 from <http://blogs.reading.ac.uk/weather-and-climate-at-reading/2014/what-do-we-mean-by-air-temperature/>).
- Chang, W., Cheng, J., Allaire, J., Xie, Y., and McPherson, J. (2017). Shiny - Web Application Framework for R. R Package Version.
- Cheng, J., Xie, Y., and Karambelkar, B. (2018). leaflet: Create Interactive Web Maps with the JavaScript ‘Leaflet’ Library.
- Doi, J., Potter, G., Wong, J., Alcaraz, I., and Chi, P. (2016). No Web Application Teaching Tools for Statistics Using R and Shiny. *Technology Innovations in Statistics Education*, 9(1).
- Ellis, D. A. and Merdian, H. L. (2015). Thinking outside the box: Developing dynamic data visualizations for psychology with Shiny. *Frontiers in Psychology*, 6(DEC):1–6.
- Feng, X., Castro, M. C., Linde, E., and Papeş, M. (2017). Armadillo Mapper. *Tropical Conservation Science*, 10:194008291772413.
- Gilabert, A. (2016). *Assessment of the bias introduced by the automatisisation of climate records combining climatological and metrological approaches*. PhD thesis, Rovira i Virgili.
- Hu, Z., Zhang, C., Hu, Q., and Tian, H. (2014). Temperature changes in central Asia from 1979 to 2011 based on multiple datasets. *Journal of Climate*.
- Jones, P., Raper, S., Bradley, R. S., Diaz, H. F., Kelly, P. M., and Wigley, P. M. L. (1986). Northern Hemisphere Surface Air Temperature Variations: 1851-1984. *Journal of Climate and Applied Meteorology*, 25(162).

- Klein, T., Samourkasidis, A., Athanasiadis, I. N., Bellocchi, G., and Calanca, P. (2017). webXTREME: R-based web tool for calculating agroclimatic indices of extreme events. *Computers and Electronics in Agriculture*, 136:111–116.
- Lemoine, N. P., Sheffield, J., Dukes, J. S., Knapp, A. K., and Smith, M. D. (2016). Terrestrial Precipitation Analysis (TPA): A resource for characterizing long-term precipitation regimes and extremes. *Methods in Ecology and Evolution*, 7(11):1396–1401.
- Li-Juan, C. and Zhong-Wei, Y. (2012). Progress in Research on Homogenization of Climate Data. *Advances in Climate Change Research*, 3(2):59–67.
- METEOCAT (2017). Dades de la Xarxa d’Estacions Meteorològiques Automàtiques. Technical report.
- Middleton, W. E. K. (1969). *Invention of the meteorological instruments*. Johns Hopkins Press, Baltimore, Md.
- Moraga, P. (2017). SpatialEpiApp: A Shiny web application for the analysis of spatial and spatio-temporal disease data. *Spatial and Spatio-temporal Epidemiology*, 23:47–57.
- Parker, D. (1994). Effects of changing exposure of thermometers at land stations. *International Journal of Climatology*, 14.
- Ribeiro, S., Caineta, J., and Costa, A. C. (2016). Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth*, 94:167–179.
- Skansi, M. d. l. M., Brunet, M., Sigró, J., Aguilar, E., Arevalo Groening, J. A., Bentancur, O. J., Castellón Geier, Y. R., Correa Amaya, R. L., Jácome, H., Malheiros Ramos, A., Oria Rojas, C., Pasten, A. M., Sallons Mitro, S., Villaroel Jiménez, C., Martínez, R., Alexander, L. V., and Jones, P. D. (2013). Warming and wetting signals emerging from analysis of changes in climate extreme indices over South America. *Global and Planetary Change*, 100:295–307.

- Staudt, M., Esteban-Parra, M., and Castro-Díez, Y. (2007). Homogenization of long-term monthly Spanish temperature data. *International Journal of Climatology*, 27(April 2007):1809–1823.
- Syrakova, M. and Stefanova, M. (2009). Homogenization of Bulgarian temperature series. *International Journal of Climatology*, 29(December 2008):1835–1849.
- Venema, V. K. (2012). Benchmarking homogenisation algorithms for monthly data. (Retrieved online on 2018-04-15 from <http://variable-variability.blogspot.com.es/2012/01/new-article-benchmarking-homogenization.htm>) .
- Wan, Z. and Hudak, P. (2000). Functional reactive programming from first principles. *Proceedings of the ACM SIGPLAN 2000 conference on Programming language design and implementation - PLDI '00*, pages 242–252.
- WMO (1992). *International Meteorological Vocabulary*. Number 182. World Metereological Organization, Geneve, second edition.
- WMO (2008). Guide to Meteorological Instruments and Methods of observation. Technical report, Geneva, Switzerland.
- WMO (2011). Guide to climatological practises. Technical Report 100.
- Zaragozí, B. M., Torres-Prieto, J., Belda-Antolí, A., Rosado-Abad, A., Cortés-Samper, C., and Ramón-Morte, A. (2016). Small data para educación: más de un millón de pirámides de población de España. *XVII Congreso Nacional de Tecnologías de Información Geográfica*, pages 454–462.