CLARIS | LPB

CLARIS | LPB

A Europe-South America Network for Climate Change Assessment

And Impact studies in La Plata Basin

SEVENTH FRAMEWORK PROGRAMME

www.claris-eu.org

Deliverables

Instrument: **SP1 Cooperation**

Thematic Priority: **Priority Area 1.1.6.3 "Global Change and Ecosystems"**

FP7 Collaborative Project – Grant Agreement 212492

CLARIS LPB
A Europe-South America Network for Climate Change Assessment and Impact Studies in La Plata Basin

DELIVERABLES

D3.1: Homogenised daily station data for precipitation, temperature and other variables

Due date of deliverable: Month 18

Start date of project: **01/10/2008** Duration**: 4 years**

Organisation name of lead contractor for this deliverable: P1-IRD

Deliverable No	Deliverable title	WP	Lead beneficiary	Estimated indicative person-months (permanent staff)	Nature	Dissemination level	Delivery date
D3.1	Homogenised daily station data for precipitation, temperature and other variables	WP3	P1-IRD	12,70	O	CO	18

D3.1. Homogenized daily station data for precipitation, temperature and other variables

Andrés Farall, Jean-Philippe Boulanger and Liliana Orellana

The study of climate variability is crucial for understanding the potential human impact on global weather. The evaluation of changes in secular trends and the detection of extreme events occurred in the past, as well as their potential evolution in the context of climate change, have similar relevance. Thus, a good quality control procedure of climate data is needed to guarantee the quality of the data. The lack of quality in climate data is mostly the result of measurement errors and arbitrary changes in the monitoring conditions, both situations usually referred to as lack of homogeneity.

Our work focuses on two different aspects of the homogenization problem
- Error detection in daily climate time series
- Inhomogeneity detection in daily climate time series

For both topics, we developed new statistical methodologies with practical implementation based on the open source statistical package R.
The error detection algorithms are already finalized and implemented on the CLARIS-LPB database. The inhomogeneity detection method is in its last stage of development.

Error detection in daily climate time series

Loosely speaking, for our method, a temperature observation $T_{i,d}$ is considered suspicious (for a particular station i and for a particular day d) if that observation is far away from the expected behaviour of $T_{i,d}$ using information of that station as well as information of surrounding correlated stations. The notion of distance used is robust, meaning that outliers (those we want to detect) do not affect the results.

Our methodology assumes that all stations could have erroneous information. Then, for every station i and day d, a robust distance D will be calculated taking into account the statistical relation existing among the temperatures of all involved stations. The error detection methodology comprises four main steps:
1. Measuring robust historical correlations between stations.
2. Defining the influence set of stations associated to each station.
3. Selecting optimal temporal bandwidths.
4. Calculating the robust Mahalanobis distances for daily observations.

The Algorithms Implementation

The methodology was carried out with the following characteristics
- Algorithms Programmed in R
- WEB Interface Implemented in JAVA
- Methods Available from the Internet
- Accessible Worldwide (Authorized Users)

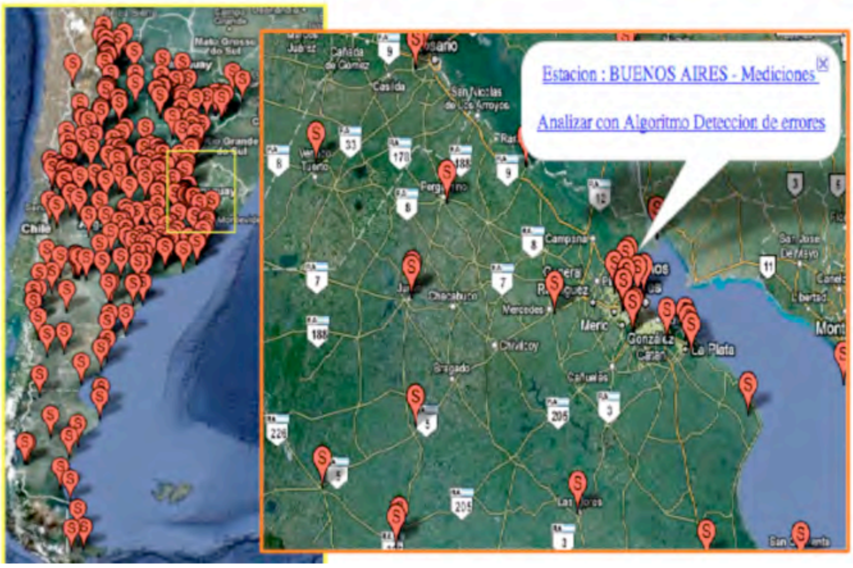


Figure 1 – Map of monitoring stations selection.

The user interface involves two simple steps:

1) Selection of the target station (Figure 1).
2) Setting of the algorithm parameters (Figure 2).

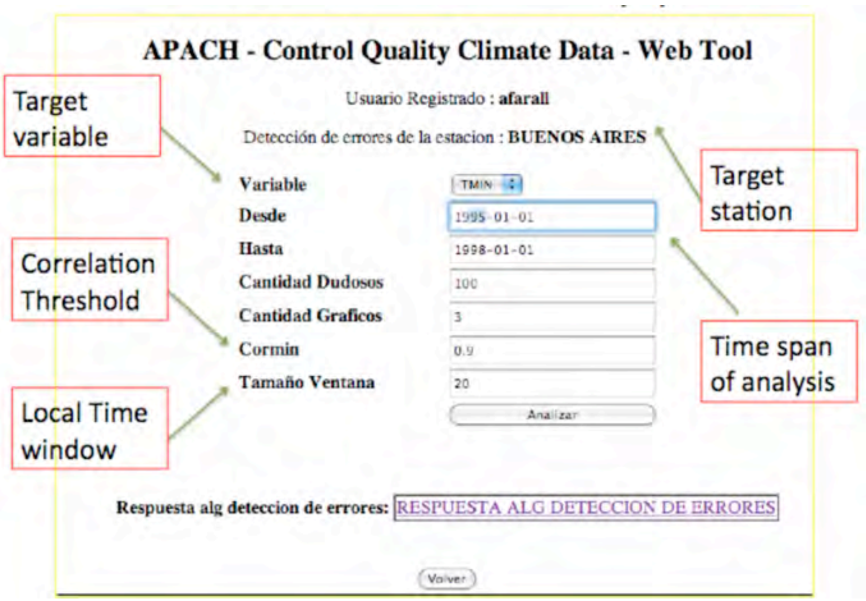


Figure 2 – User interface control panel.

The outcome of the method comprises both a graphical output and a numerical one. The graphical output (Figure 3) shows temperature trajectories of the doubtful station as well as those of the related stations.

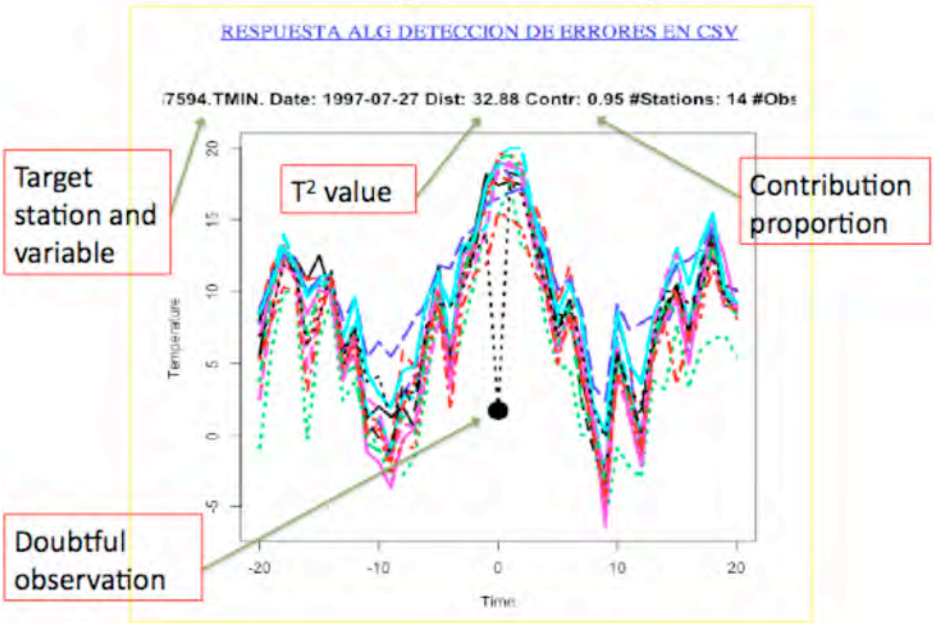
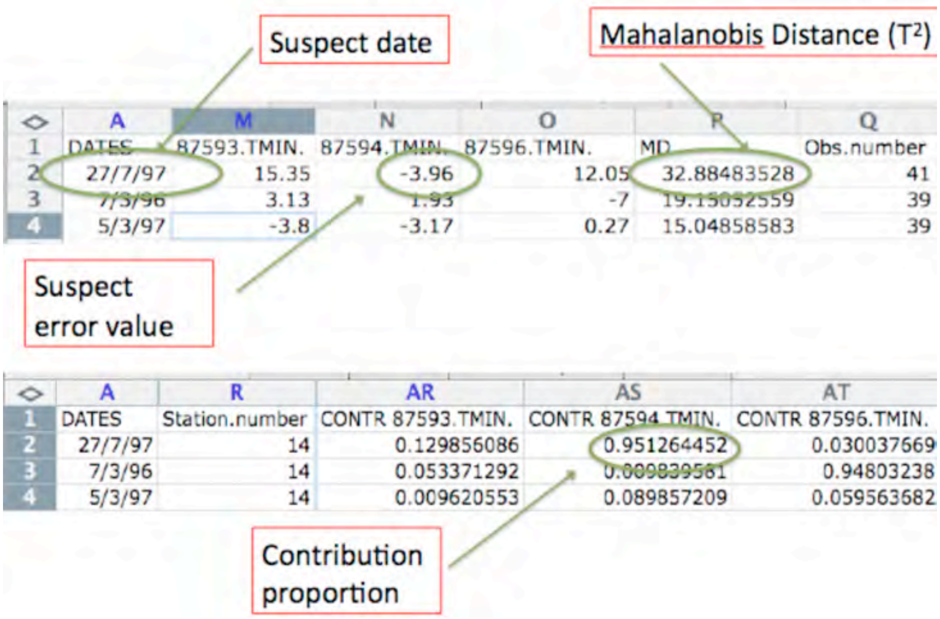


Figure 3 – Graphical output.

The numerical output (Figure 4) shows the relevant information needed to decide about the correctness of the suspect observation.



	A	M	N	O	P	Q
1	DATES	87593.TMIN.	87594.TMIN.	87596.TMIN.	MD	Obs.number
2	27/7/97	15.35	-3.96	12.05	32.88483528	41
3	7/3/96	3.13	1.93	-7	19.15052559	39
4	5/3/97	-3.8	-3.17	0.27	15.04858583	39

	A	R	AR	AS	AT
1	DATES	Station.number	CONTR 87593.TMIN.	CONTR 87594.TMIN.	CONTR 87596.TMIN.
2	27/7/97	14	0.129856086	0.951264452	0.030037669
3	7/3/96	14	0.053371292	0.009839581	0.94803238
4	5/3/97	14	0.009620553	0.089857209	0.059563682

Figure 4 – Numerical output.

The next step is to develop, -for the error detection problem-, an interface to let the expert user to flag observations (as good/doubtful/bad) and to correct wrong data at the same time. The storage of the flag information and the corrected data is done jointly with the raw database, thus allowing the future extraction of the "corrected" database without altering the raw data. The software and its interface will be presented at CLARIS LPB M36 Meeting