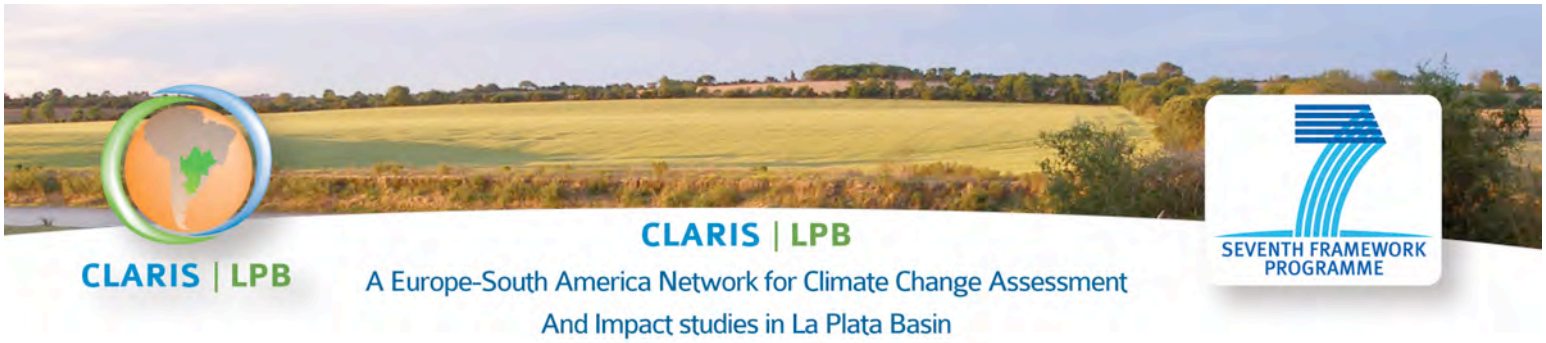


Instrument: **SP1 Cooperation**

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

D7.4: Statistical downscaling tool

Due date of deliverable: Month 24

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

Organisation name of lead contractor for this deliverable: P13-CONICET

Deliverable No	Deliverable title	WP	Lead beneficiary	Estimated indicative person-months (permanent staff)	Nature	Dissemination level	Delivery date
D.7.4	Tailored software for climate model output analysis and downscaling	WP7	P13-CONICET	26,00	R	PU	24/48

D7.4: Statistical downscaling tool

Ariel D'Onofrio (UBA/IRD) and Jean-Philippe Boulanger (IRD)

The objective of D7.4 was to create of a tool for statistical downscaling to help researchers to create and analyze statistical downscaling data. CHAC software is the final result, written entirely in Matlab, the manual and software are freely downloadable from http://sourceforge.net/projects/chac/.

CHAC offers a user-friendly and flexible tool for statistical downscaling; it has many features, which allows users to generate their own results in a straightforward basis. The software is self-contained and there is no need of any knowledge in Matlab programming, though a little knowledge of Matlab environment is recommended.

Users are required to introduce their station database, reanalysis and model information.

Scientific basis for precipitation downscaling (D'Onofrio et al., 2010):

CHAC, a weather pattern classification method is applied and calibrated to Argentine daily weather stations in order to predict daily precipitation data. The clustering technique for weather pattern classification is based on k-means and is applied to a set of 17 atmospheric variables from the ERA-40 reanalysis covering the 1979-1999 period The set of atmospheric variables represents the different components of the atmosphere (dynamic, thermal and moisture). Different sensitivity tests are applied to optimize (i) the number of observations (weather patterns) per cluster, (ii) the spatial domain size of the weather pattern around the station and (iii) the number of members of the ensembles. All the sensitivity tests are compared using the ROC (Relative Operating Characteristic) Skill Score (RSS) derived from the ROC curve used to assess the performance of a predictive system. First, we found the number of observations per cluster to be optimum for values larger than 39. Second, the spatial domain size (~ 4°x4°) was found to be closer to a local scale than to a synoptic scale, certainly due to a dominant role of the moisture components in the optimization of the transfer function. Indeed, when reducing the set of variables to the subset of dynamic variables, the predictive skill of the method is significantly reduced, and the domain size must be increased. A potential improvement of the method may therefore be to consider different domains for dynamical and non-dynamical variables. Third, the number of members per ensembles of simulations was estimated to be always two to three times larger than the mean number of observations per cluster (meaning that at least all the observed weather patterns are selected by one member). The skill of the statistical method to predict daily precipitation is found to be relatively homogeneous all over the country for different thresholds of precipitation. The same results were found in various countries around the world.

Scientific basis for temperature downscaling (D'Onofrio et al., 2011):

We study the utilization of tapped-feed forward network (TDFFN) as a method for statistical downscaling of daily maximum/minimum temperature on short-range forecast models. The study is made over 7 stations of Argentina. TDFFN allows finding a statistical transfer function that takes into account the previous values of tmax/tmin and precipitation as well as model variables in order to make a tmax/tmin forecast. First, we compare the skill of TDFFN trained with Bayesian Regularization method, with the more standard multi-layer perceptron (MLP) trained with early stopping criteria, using ERA-40

reanalysis variables as predictors. Second, on the basis of the knowledge gain over the study of ERA-40 prediction we develop a method for daily forecast prediction for tmax/tmin over the Global Forecasting System model (GFS). The method is based on training different TDFFN for tmax/tmin and precipitation since it's showed from the results of the first study that different predictions variables must be used for different predictand variables to archive optimal prediction, also precipitation prediction that feed the final model must be made in a different way due to its stochastic nature, this different models are then used in a cascade fashion to feed the final TDFFN that generates the prediction. Results shows, both for ERA-40 and GFS, that a skill improvement is achieved if past values are taking into account. Also Bayesian regularization method allows not only predicting a conditional mean daily value but also predicting the variance of the error made in the prediction, given the opportunity to use this information to generate errors bars, confidence intervals, and impact analysis.

Technical features.

Easy plotting of key information

Map of stations database.

Spatial map of users stations data.

Time series of stations data divided in different periods of the year.

Access to a set of different methodologies

The system allows to choosing different standard methods such as:

- Multilayer feed forward networks.
- K-means.
- Self-organizing maps.
- Linear regression.
- CHAC method, both for daily precipitation and temperature.

Assessment plots of the downscaling system skills

QQ-plot, PP-plot, ROC curves and Roc skill score for each station or a map.

Capacity-building and dissemination activities:

In order to disseminate the CHAC software to the scientific community, a set of workshops were organized during the development phase of the tool, these experiences were successful and provided feedback for fine-tuning CHAC in order to meet researchers' needs.

The workshops entitled: **"CHAC: An atmospheric pattern classification method for precipitation and temperature downscaling"** were organized in:

- Tambo Real Hotel, Quito, Ecuador, October 2-12, 2010
- LOCEAN/IPSL, Paris, France. June 16-18, 2008.
- University of Buenos Aires, Buenos Aires, Argentina. August 20-22, 2008.
- SENAMHI, Lima, Perú, December 2-6, 2009.

Peer-reviewed publication

"CHAC: A weather pattern classification system for statistical for regional climate downscaling for daily precipitation, A. D'onofrio, J.-P. Boulanger and E. C. Segura , Volume 98, Numbers 3-4, 405-427, DOI: 10.1007/s10584-009-9738-4

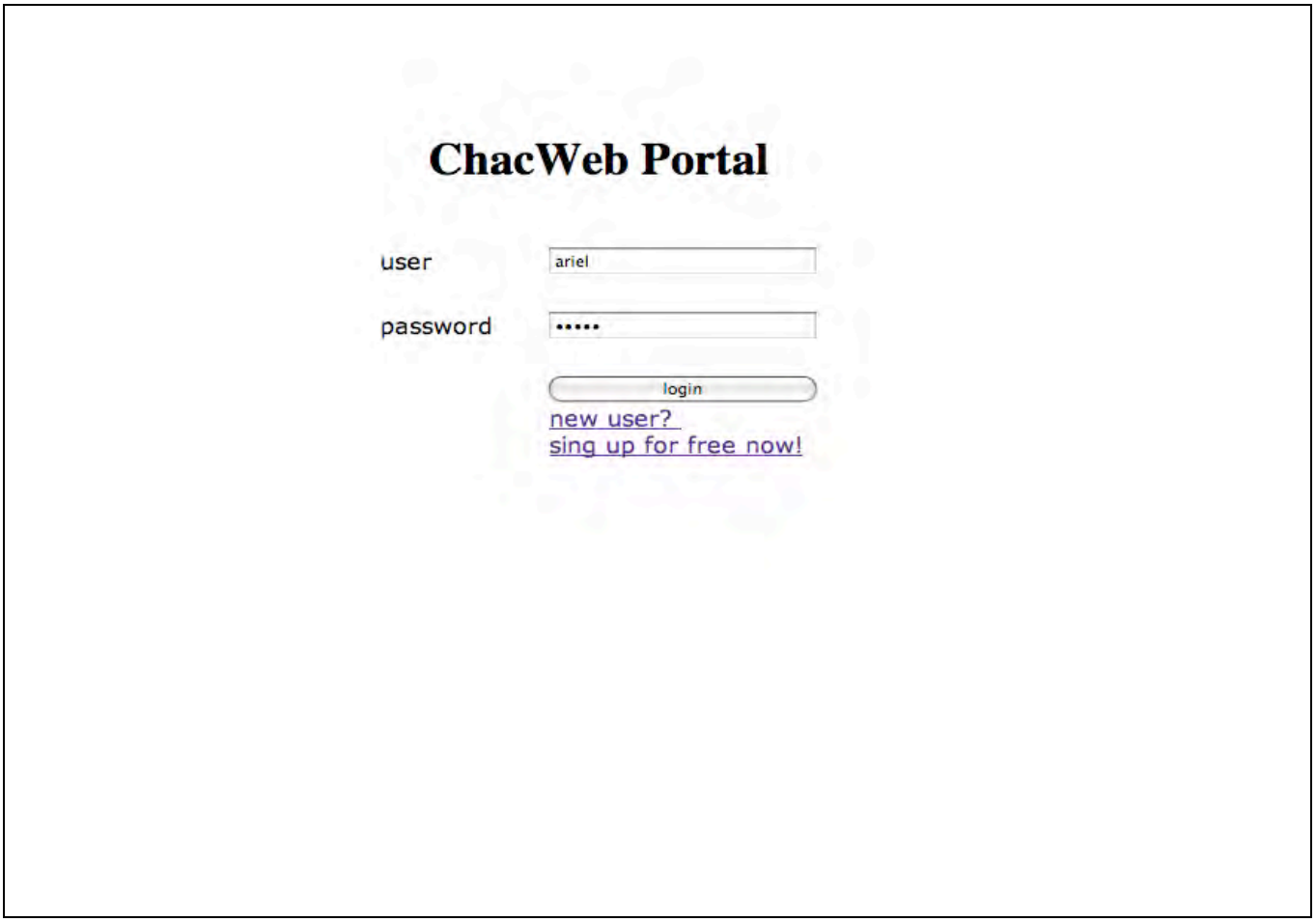
Future publications

"Using tapped-delayed feed-forward networks for daily maximum temperature short term forecast", A. D'onofrio, J.-P. Boulanger and E. C. Segura

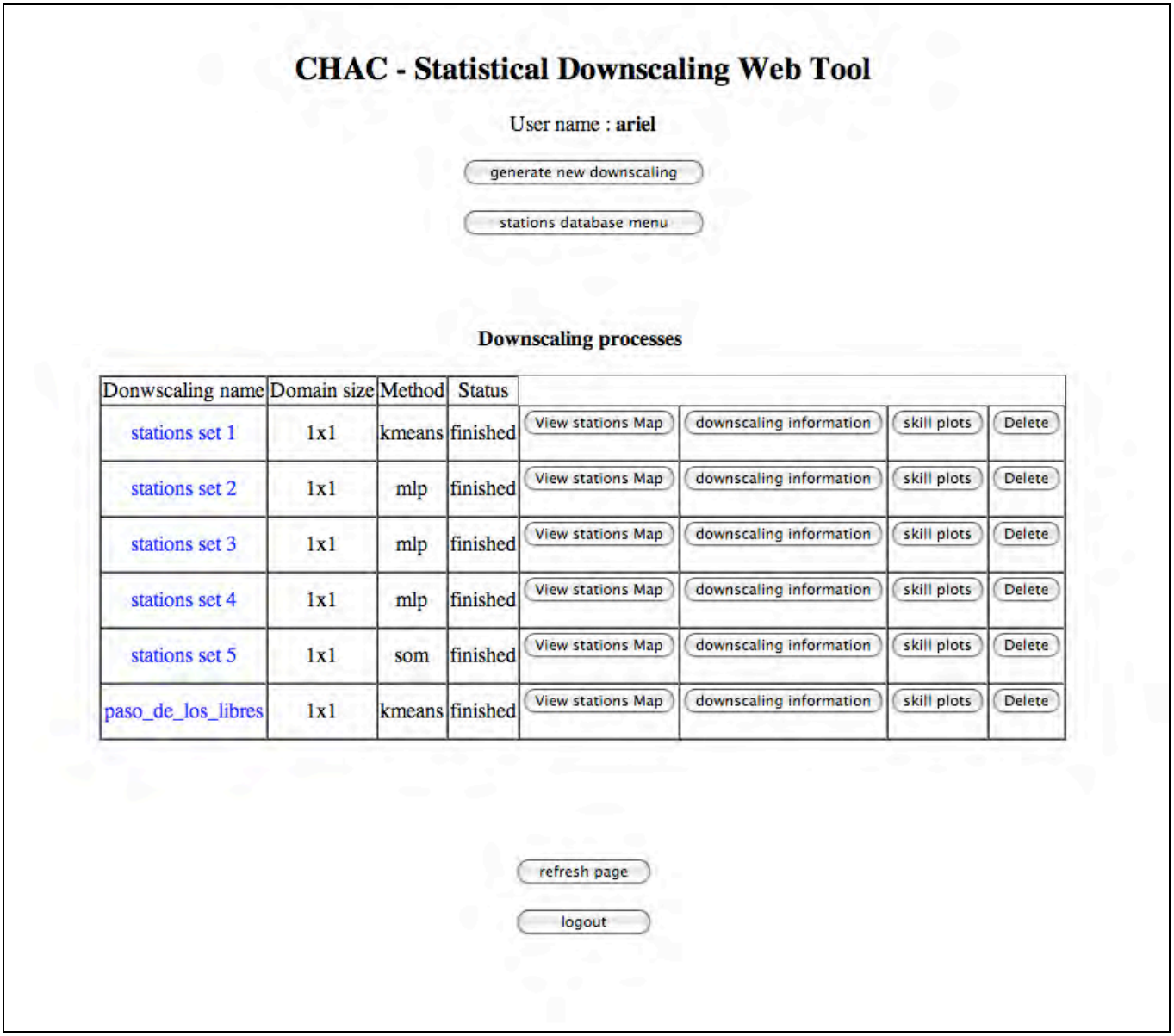
"Assessment of future climate models scenarios for south America, using Köppen classification and statistical downscaling", A. D'onofrio, J.-P. Boulanger and E. C. Segura

Web page overview

CHAC web page is hosted in a dedicated CLARIS-LPB server located at CIMA (Centro de Investigaciones de la Atmosfera y los Océanos, UBA).
It can be accessed from http://chac.cima.fcen.uba.ar:9090/chac/login.jsp.



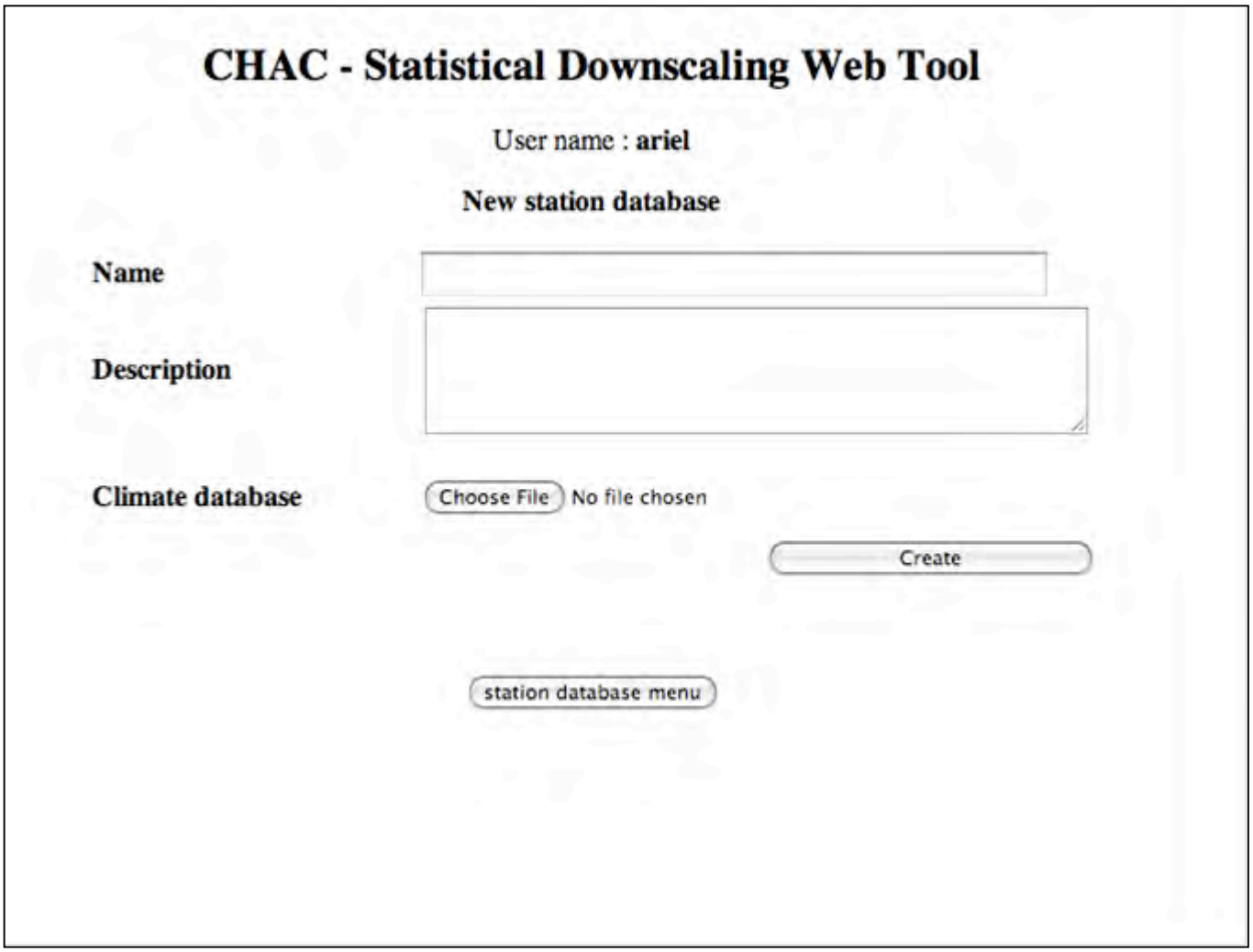
A user name and mail (optional) is required to get access to the system. Once a user has registered and logged in the system will show the following main screen.



In here, a list of downscaling processes previously executed by the user is shown along with options for each process. Pressing on “View stations Map” the user can see where the stations of that downscaling process located in the earth map. From “downscaling information” a list of all the parameters of the downscaling process is available. “Skill plots” shows a set of plots generated for each station, and “delete” button erases de process from the system.

Uploading daily station information

In order to start a user has to upload at least one station database to use as an input for downscaling. The stations data must be entered in a .csv file. This information is only viewed by the user who uploads it, so privacy of data is guaranteed. To upload station database we enter to station database menu -> upload stations database.

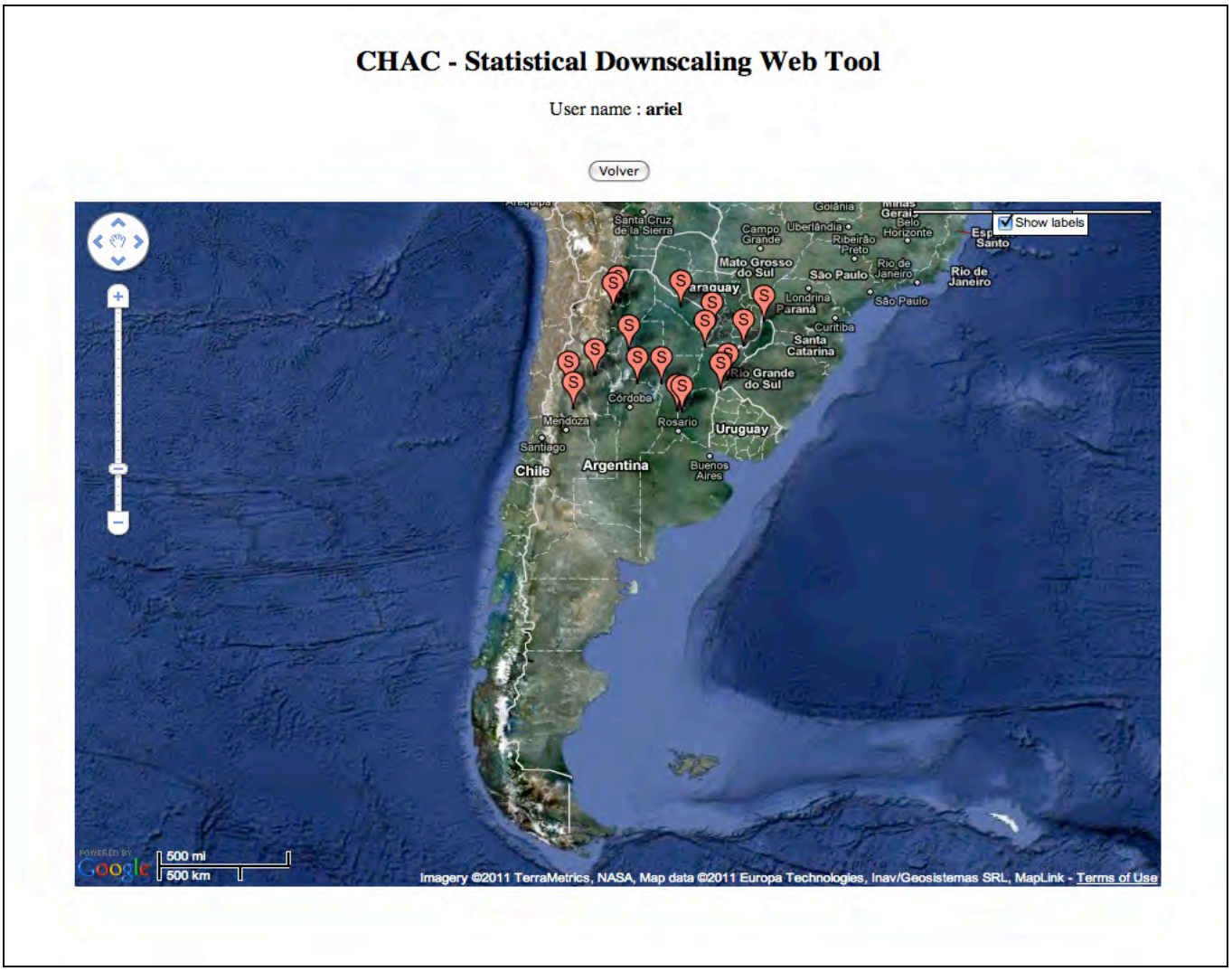


Users can upload as much stations databases as they need.

Viewing station information

Once the stations are uploaded the user can navigate through its data in the following ways:

A map of the stations present in the station database from stations database-> options -> View stations.



The set of stations belonging to a stations database can be listed from stations database->options->stations list.

ChacWeb - Statistical Downscaling Web Tool

User name : **ariel**

Database name : **stations_of_ariel**

station name	OMM code	Latitude	Longuitude	Country			
JUJUY AERO	1	-24.3833	294.9167	BRASIL	View stations	Display data	Export to CSV
SALTA AERO	2	-24.85	294.5167	BRASIL	View stations	Display data	Export to CSV
LAS LOMITAS	3	-24.7	299.4167	BRASIL	View stations	Display data	Export to CSV
IGUAZU AERO	4	-25.7333	305.5333	BRASIL	View stations	Display data	Export to CSV
SANTIAGO DEL ESTERO AERO	5	-27.7667	295.7	BRASIL	View stations	Display data	Export to CSV
FORMOSA AERO	6	-26.2	301.7667	BRASIL	View stations	Display data	Export to CSV
CORRIENTES AERO	7	-27.45	301.2333	BRASIL	View stations	Display data	Export to CSV
POSADAS AERO	8	-27.3667	304.0333	BRASIL	View stations	Display data	Export to CSV
LA RIOJA AERO	9	-29.3833	293.1833	BRASIL	View stations	Display data	Export to CSV
VILLA MARIA DEL RIO SECO	10	-29.9	296.3167	BRASIL	View stations	Display data	Export to CSV
CERES AERO	11	-29.8833	298.05	BRASIL	View stations	Display data	Export to CSV
PASO DE LOS LIBRES AERO	12	-29.6833	302.85	BRASIL	View stations	Display data	Export to CSV

Station count : 12

Back

Pressing the button “Display data” the daily data is available.

CHAC - Statistical Downscaling Web Tool

Usuario Registrado : **ariel**

Data from station : **SANTIAGO DEL ESTERO AERO**

Date	Minimun Temperature	Maximun Temperature	Precipitation	Radiation	Wind
1956-02-01	17.2	29.0	NaN	NaN	NaN
1956-02-02	15.3	28.5	NaN	NaN	NaN
1956-02-03	18.2	31.7	NaN	NaN	NaN
1956-02-04	22.6	35.4	NaN	NaN	NaN
1956-02-05	13.8	30.2	NaN	NaN	NaN
1956-02-06	13.0	23.3	NaN	NaN	NaN
1956-02-07	15.7	27.0	NaN	NaN	NaN
1956-02-08	17.9	30.3	NaN	NaN	NaN
1956-02-09	20.4	33.5	NaN	NaN	NaN
1956-02-10	20.8	31.9	NaN	NaN	NaN
1956-02-11	19.1	32.6	NaN	NaN	NaN
1956-02-12	21.6	34.7	NaN	NaN	NaN
1956-02-13	19.0	33.3	NaN	NaN	NaN
1956-02-14	18.5	27.5	NaN	NaN	NaN
1956-02-15	15.2	21.8	NaN	NaN	NaN
1956-02-16	17.0	26.3	NaN	NaN	NaN
1956-02-17	14.8	30.5	NaN	NaN	NaN
1956-02-18	15.2	33.9	NaN	NaN	NaN
1956-02-19	18.8	36.5	NaN	NaN	NaN
1956-02-20	23.1	38.4	NaN	NaN	NaN

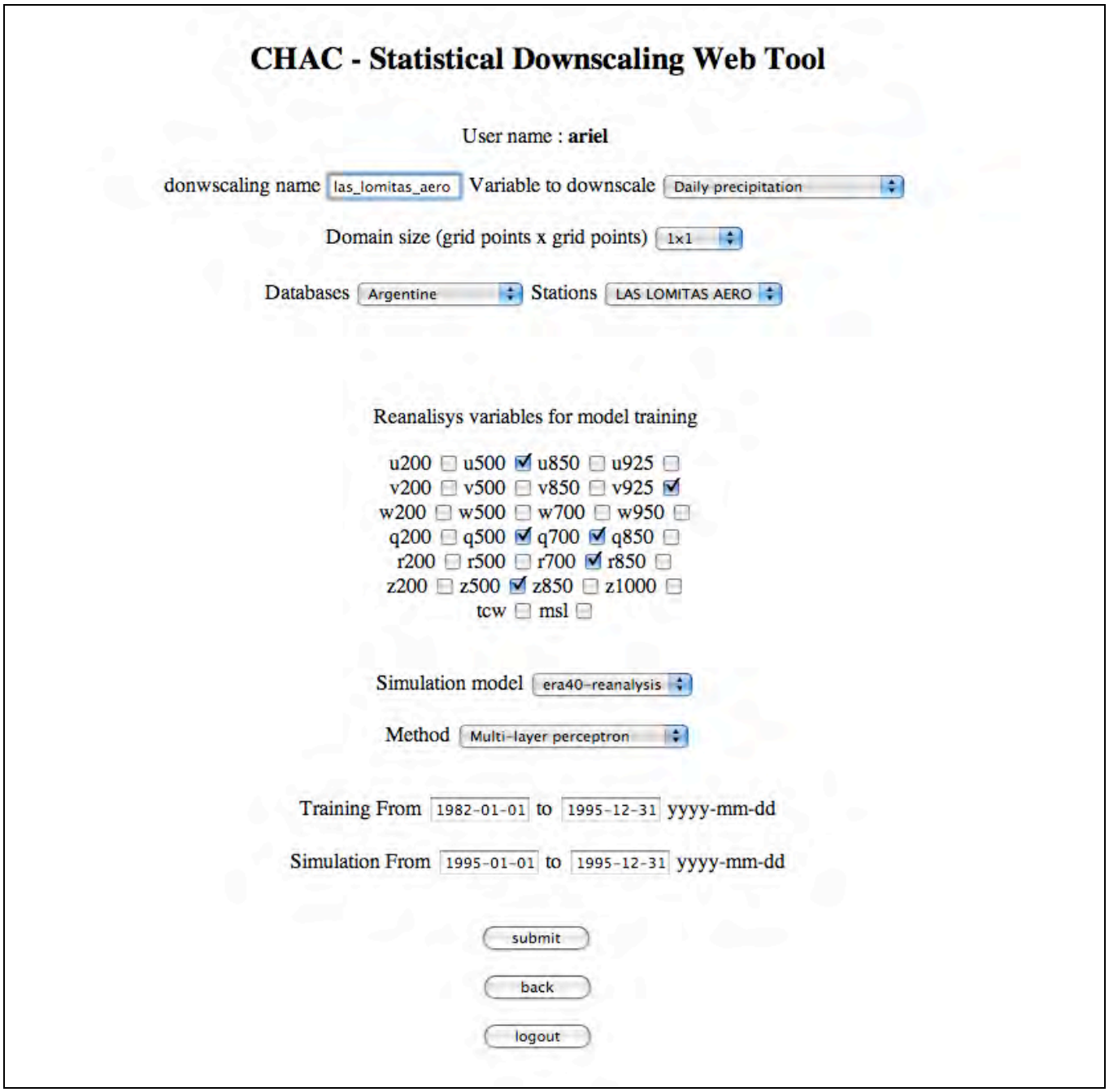
<< < ... 1 2 3 ... > >>

Number of rows : 17502

Back to stations list

Generating a downscaling process

Back to the main page, if the user presses the "generate downscaling" button a new page with the required information to fill is prompted.



In this form, all parameters regarding on how downscaling will be generated are set.

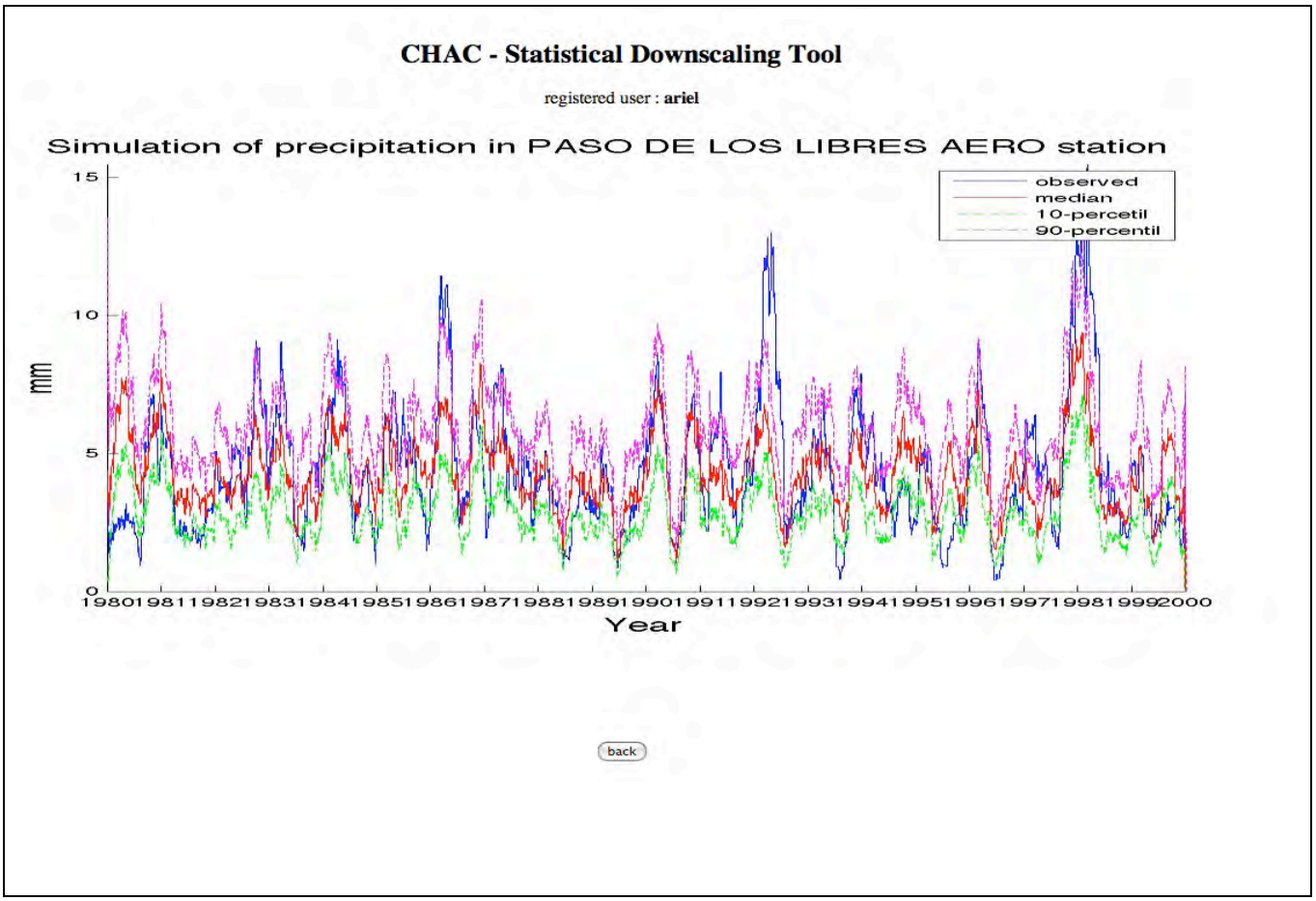
First, users must enter the downscaling process name, then the kind of local variable to simulate; at this moment CHAC allows precipitation, maximum and minimum temperature.
A window size of the weather pattern is needed; size ranges from 1x1 grid points to 10x10 grid points of ERA-40 reanalysis resolution (1.125 degree).
Next, a list of stations from a particular station database are selected, this stations are the one to be processed for downscaling.
Also, a set of predefined reanalysis variables from ERA-40 are needed along with which type of downscaling method to use and the beginning and end of the training and simulation.

Once the downscaling process is finished, an e-mail will be send to notify the user, also in the main page the downscaling process will be set as FINISHED.

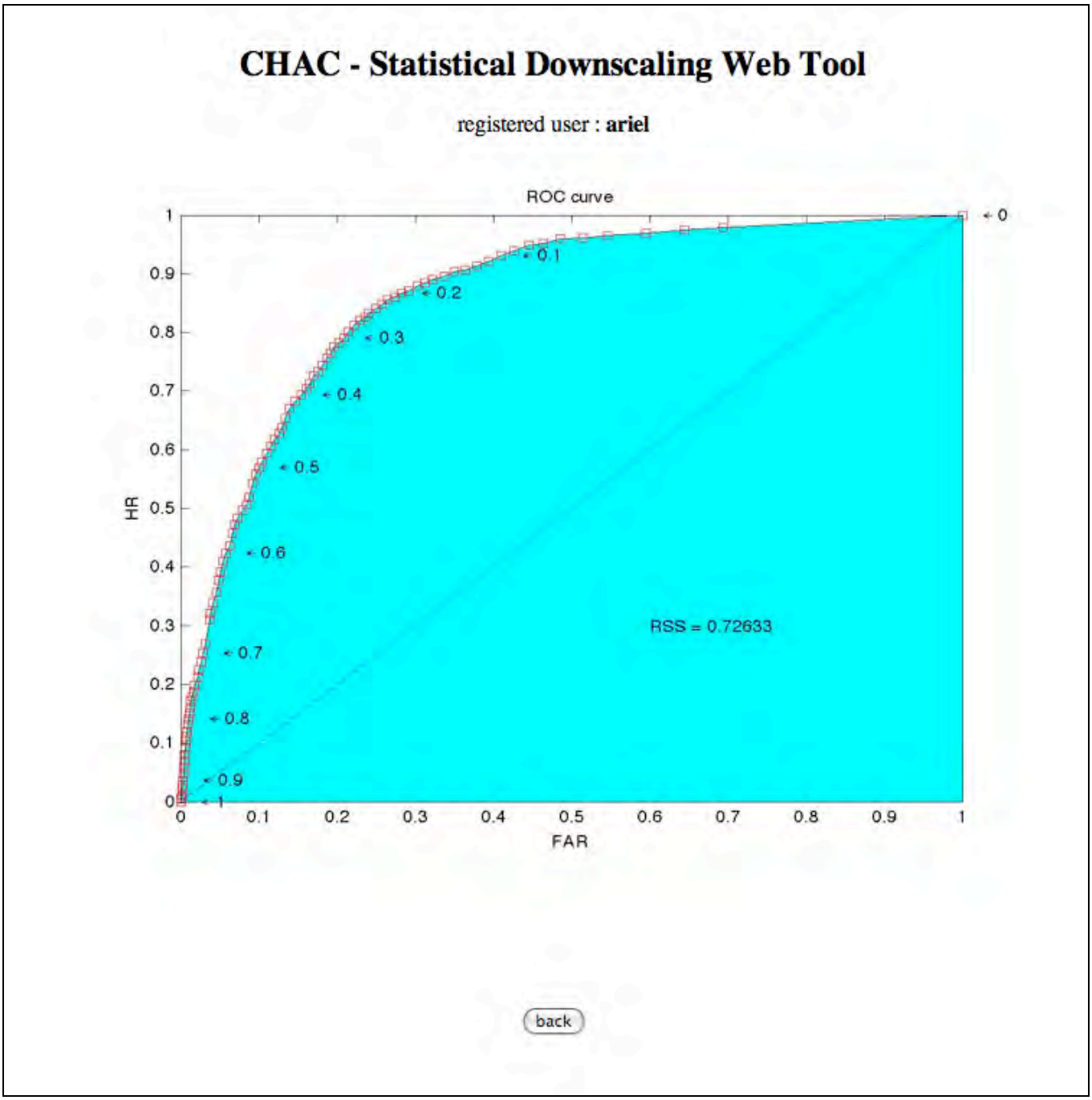
Viewing downscaling results

For every downscaling process a set of four predefined plots are available.

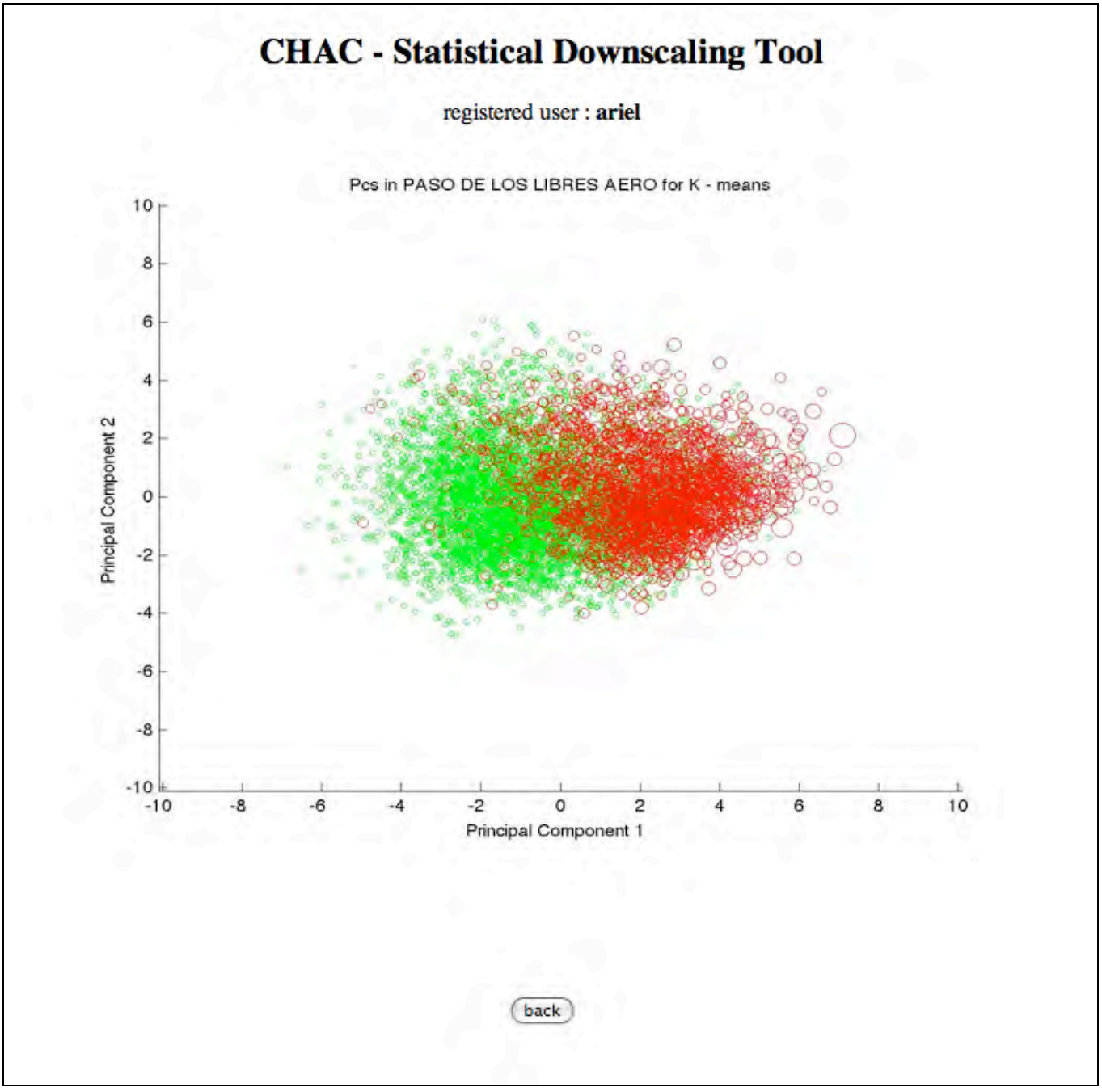
1) A time series plot of simulated vs. observed values of the simulating period. In case of an ensemble method (for instance k-means and SOM, 10^{th} and 90^{th} percentile will also be plotted).



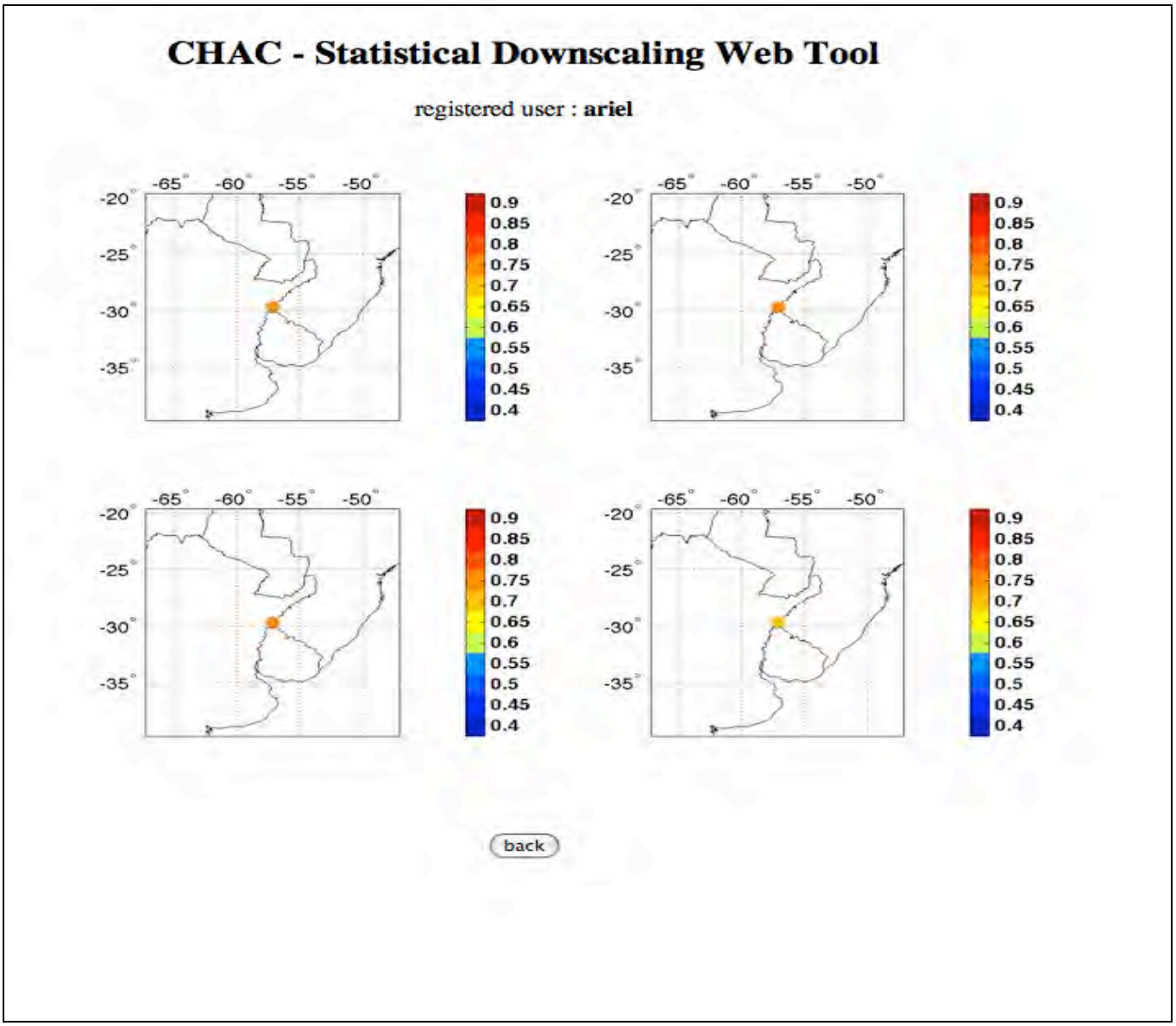
2) A ROC (Relative Operating Curve) of the simulated period.



3) A PCA (Principal Component Analysis) plot, showing the first two principal components of the training weather patterns, in case of precipitation, they are colorized in green to indicate absence of precipitation and red for presence of it..



Four plots indicating the ROC for four different events.


CHAC - Statistical Downscaling Web Tool
registered user : ariel
-65° -60° -55° -50°
-20° -25° -30° -35°
0.9 0.85 0.8 0.75 0.7 0.65 0.6 0.55 0.5 0.45 0.4
back