

# Homogenization of Argentinean Temperature Data

Alexis Hannart\*, Jean-Philippe Boulanger+, Philippe Naveau\*, Olivier Mestre\*\*, Monica Marino\*\*  
 \*CNRS, +IRD, \*\*Météo France, ++Servicio Meteorológico Nacional

## Introduction

### Context

- Long climate records are affected by artificial shifts (inhomogeneities) due to changes in the measurement conditions.
- Non anecdotal issue, as inhomogeneities have the same magnitude as the climate signal : direct analysis of the raw data series can lead to wrong conclusions.
- Issue is now extensively treated : many methodologies and results are available.
- In Argentina, many climate change surveys are based on the weather service data, which has never been homogenized.

### Objectives

- Adapt state-of-the art homogenization techniques to Argentinean data specificities (temporal and spatial scarcity, lack of metadata).
- Apply methodology to the data (detect and correct shifts).
- Analyse most obvious implications on existing results on climate change.

### Data

- Overall 249 stations (1 per 10<sup>4</sup> km<sup>2</sup> on average) from 1959 onwards.
- Only 45 stations have uninterrupted record (24 years recorded on average).
- Preliminarily checked for inconsistencies.

## Method

### Homogenization is a three steps process :

- Removal of climate signal based on difference of candidate series with neighbouring series (one by one or averaged into a reference series).
- Detection of shifts on yearly data, based on statistical methods combined with metadata.
- Correction on monthly data, based on dates of detected shifts.

### Challenges posed by Argentinean data :

- Data scarcity ⇒ difficult to infer and remove local climate signal ⇒ shifts are blur, decision is hard to objectiveize.
- Lack of metadata ⇒ limited posterior validation of detected shifts.
- High number of stations ⇒ too time consuming to process manually.

### Methodological options retained to address challenges :

- Pairwise difference of neighbouring stations ⇒ avoids building weak reference series, makes the most of local information to retrieve climate signal.
- Bayesian multiple change point detection ⇒ mitigates the lack of metadata through prior characterization of shifts amplitude and frequency / objectivizes decisioning through cost minimization.

### Other innovations of proposed method :

- Bayesian framework allows fully objectivized decisioning ⇒ no need to calibrate/ chose thresholds.
- Simplifying assumptions of the bayesian model make it explicitly resolved and light to implement ⇒ reduced computing time (minutes).
- Bayesian framework provides posterior distributions of jump amplitude and positions ⇒ confidence intervals on estimators can be derived.
- Fully automated attribution of jumps based on the use of posterior probabilities rather than visual inspection ⇒ limits complexity and subjectivity, considerably reduces human time investment.

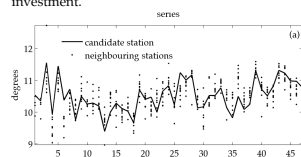


Figure 1a: Candidate series consists of daily minimum temperature at Salta Aero (plain line). Ten neighbouring series are selected based on distance, correlation and length of record (dots for clarity). Detection is performed on the ten series obtained by difference between candidate and neighbours.

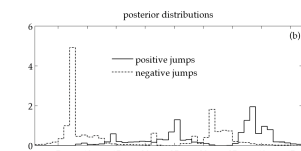
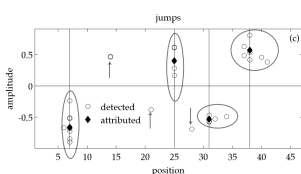


Figure 1b and 1c:

- Bayesian detection algorithm outputs:
  - A set of jumps detected and estimated from the ten difference series, plotted in a position x amplitude diagram (1c) (white circles).
  - Posterior distributions of the position of each jump detected, which are grouped and summed up by jump sign (1b).

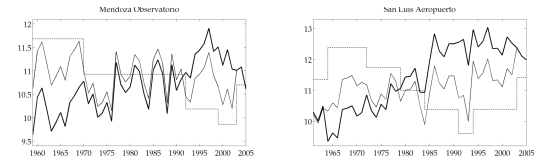


The attribution algorithm estimates jumps, in the candidate series (1c, dark diamonds), from grouping and weighted averaging of jumps detected on difference series (1c, white circles). Grouping is performed using an objective criterion derived from posterior distributions (1b), instead of visual inspection as is usually the case.

In the example, 24 jumps are detected on difference series, out of which 3 are dropped out (arrows), the remainder being grouped (large circles) into 4 jumps finally attributed to Salta Aero.

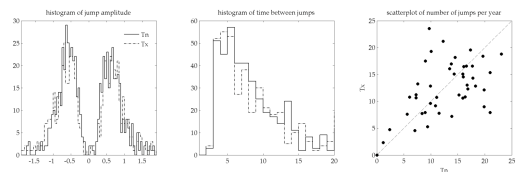
## Results

Method is applied to series of daily minimal and maximal temperature (Tn, Tx) leading to detection and correction of respectively 645 and 625 jumps.



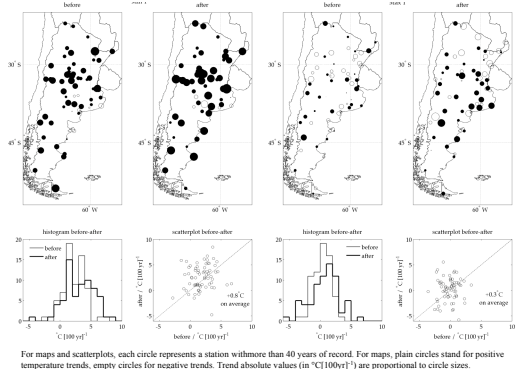
### Main jumps characteristics

- Consistency with a priori information from past studies on both frequency (8.5 yr for Tn, 8.3 yr for Tx) and mean amplitude (0.75°C for Tn, 0.78°C for Tx).
- Marked asymmetry : negative jumps are more frequent (+15% for Tn, +3% for Tx) and have larger amplitudes (+15% for Tn, +7% for Tx).
- Consistency between shifts detected on Tn and Tx, with respect to shifts position (59% correlated by year).
- No obvious temporal nor spatial pattern on number of jumps.



### Main implications on climate trends:

- Globally, past results revisited with homogenized data show that warming trend on Tn is amplified (+2.5°C/century vs +1.7°C/century), and the slight cooling trend on Tx becomes a slight warming trend (+0.2°C/century vs -0.1°C/century).
- Locally, the dispersion around these globally averaged « warming effects » of instrumental correction is important, and homogenization sometimes result in a « cooling effect » at the station level.



For maps and scatterplots, each circle represents a station with more than 40 years of record. For maps, plain circles stand for positive temperature trends, empty circles for negative trends. Trend absolute values (in °C/100yr<sup>2</sup>) are proportional to circle sizes.

## Conclusion

- New method proposed seems to handle properly data scarcity and lack of metadata, which were most challenging issues.
- It has thus been possible to identify and correct a large number of artificial shifts in the Argentinean climate record.
- Frequency and amplitude of those inhomogeneities match usually observed levels.
- Trend analysis performed on homogenized temperature series lead to an increased estimate of warming overall, as compared to previous studies.
- Next steps : monthly and daily database correction, homogenization of precipitation database, further implications on extreme values and trends.

## References

- Caussinus H., Mestre O. (2004), Detection and correction of artificial shifts in climate series. J. Roy. Statist. Soc. Series C: Applied Statistics 53, 405-425.
- Hannart A., Naveau P. (2008), A Bayesian homogenization method: Sequential detection based on prior characterization of change points. Water Resour. Res. (submitted)
- Rusticucci M., Barrucand M. (2004), Observed trends and changes in Temperature Extremes over Argentina. J. Climate vol 17, No. 20, 4099-4107.

## Acknowledgements

We wish to thank the Institut Pierre-Simon Laplace and the European Union for funding this research, IRD and CNRS for supporting it, the CIMA and DCAO for hosting us while working on this study, the SMN and Météo France for providing comprehensive data and fruitful insights.

