A regional quality-controlled database of historical climate data for southern South America

M. Skansi[1](mms@smn.gov.ar), H.Veiga[1], N. Herrera[1], F. Silva[2], L. Fortes[2], R. de Ruyver[3], M. Pastén[4], H. López[4], C. Martínez[5], M. Bidegain[5], G. Carrasco[6], Y.Castellon Geier[6], E. Carrasco[7], S. Rovere[8], G.Podestá[9]

1-SMN-AR , 2-INMET-BR, 3-INTA-AR, 4-DNMH-PY, 5-INUMET-UY, 6-SENAMHI-BO, 7-DMC-CL, UBA-AR, 9-UMRS-US

Summary

- A regional data base of daily climate data is being compiled for southern South America.
- The data base will be routinely updated every 10 days.
- A consensus Quality Control (QC) protocol was defined by participating countries.
- The climate data will be used to derive operational decision-relevant products to be disseminated by the Regional Climate Center for southern South America.

Motivation

There is active demand for reliable data and derived information on climate: long-term, continuous, well-calibrated observations are critical for defining the evolving state of the Earth's climate, and to support decisions in climate-sensitive sectors of society.

In response to this societal need, the Regional Climate Center for southern South America (CRC-SAS, the Spanish/Portuguese acronym) and a research project on climate services funded by the Inter-American Institute for Global Change Research (IAI) are compiling a database of daily weather variables.

The data base

Stations: The data base includes, for now, first-order meteorological stations operated by meteorological and some research agencies of Brazil, Paraguay, Uruguay, Bolivia, Chile, and Argentina (Figure 1).

Weather variables: During the first stage, only daily maximum and minimum temperatures and rainfall will be contributed by all participating countries. We will subsequently include other variables necessary to compute decision-relevant diagnostics.

Temporal coverage: 1961 to present.

Metadata: For now, only basic metadata are included (station location, elevation, observation frequency). A more thorough treatment of metadata is awaiting new standards to be released by the World Meteorological Organization (WMO).

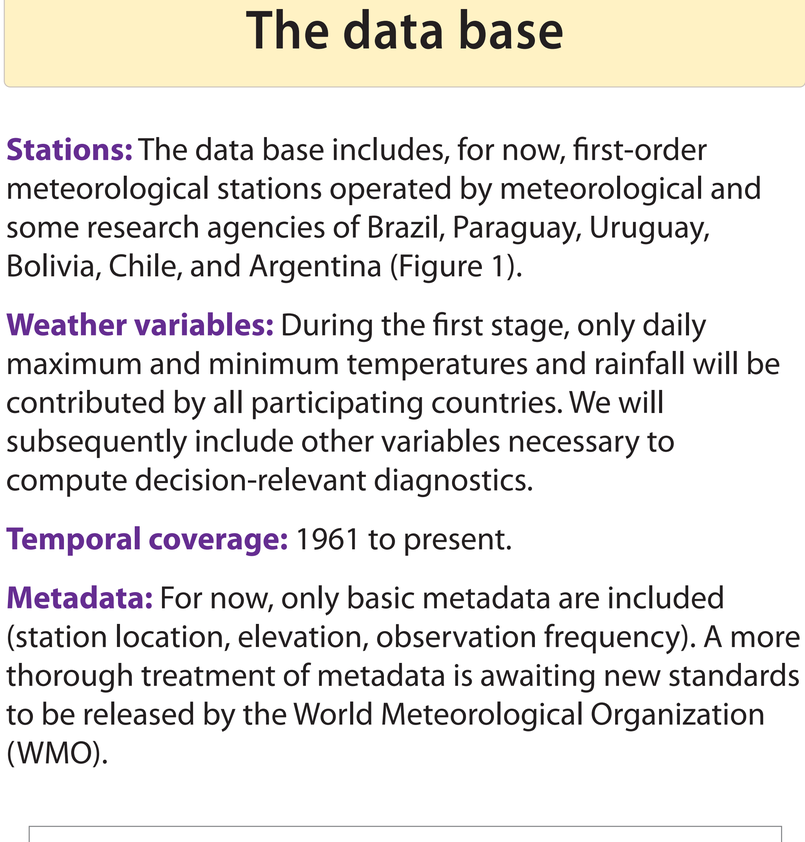

Figure 1: Meteorological stations to be included in the database of daily climate variables of the Regional Climate Center for southern South America.

Acknowledgements: This work was funded by

Inter-American Development Bank · Inter-American Institute for Global Change Research · U.S. National Science Foundation

Quality Control

We have built a superset of all quality control procedures used by agencies in the region, and added several published procedures. All QC methods are implemented in free, open-source R.

The quality control procedures can be grouped into six major "families" (Figure 2):

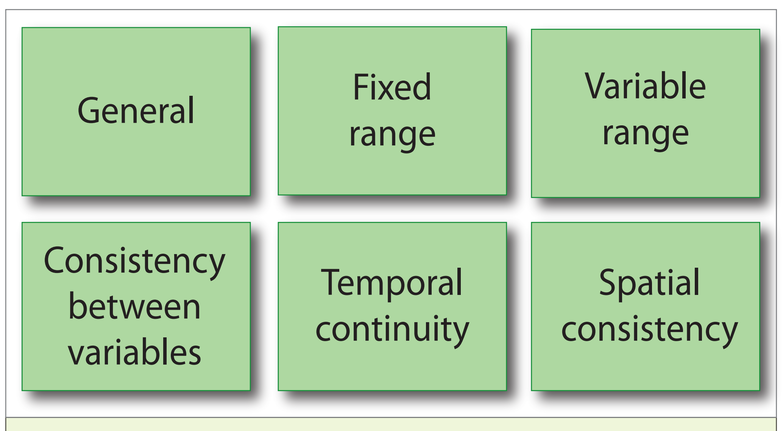

Figure 2: Families of quality control tests to be performed to identify suspect values in daily climate variables.

General: These tests check the overall integrity of the data, such as avoiding duplicated or out of sequence dates.

Fixed range: The tests flag physically impossible or extremely unlikely values. Thresholds used to identify suspect values are constant for all stations and times of the year.

Variable range: These tests identify suspect values based on thresholds or ranges that fluctuate throughout the year for each meteorological station.

Temporal continuity: These tests check for the presence of unusual jumps or spikes in daily values of a weather variable.

Consistency among variables: The checks examine inconsistencies among values of different variables for the same day of observation.

Spatial consistency: These tests check the values of a weather variable for a given station by comparing them with simultaneous observations in other stations that are geographically close.

Examples: QC Maximum temperature Pehuajó

Variable range: Deviations from a fitted seasonal cycle

A Generalized Additive Model (GAM) is used to fit a flexible seasonal cycle (i.e., no explicit functional form is specified). The tests identifies as suspect those values that show extreme deviations from the fitted cycle. The flagging thresholds are expressed in terms of percentiles of computed deviations. See an example of this test in Figure 3.

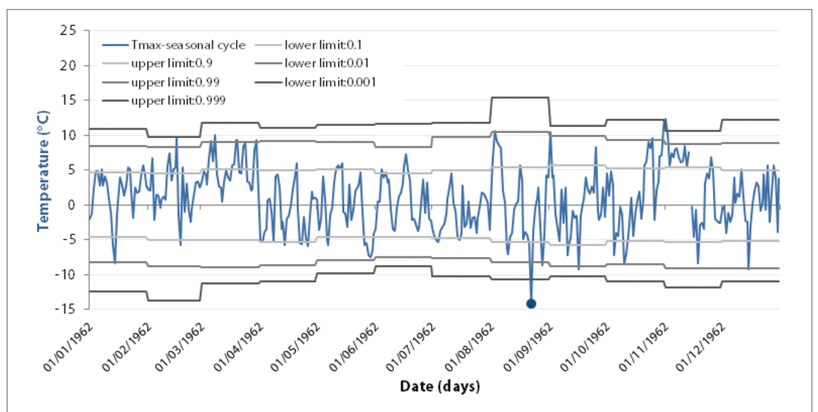

Figure 3. Differences between observed maximum temperature and the fitted seasonal cycle. Different thresholds for rejection (corresponding to different percentiles of differences) are indicated by the horizontal lines. Tmax on 22 August 1962 is highlighted with a blue point, indicating it has been flagged as "suspect."

Temporal continuity: Jumps between consecutive days

The test computes observed differences between the values of a variable on two consecutive days. A flagging threshold is defined in terms of extreme percentiles of the empirical distribution of differences. A large difference would be flagged as suspicious, and values for the two days in question must be verified. See an example of this test in Figure 4.

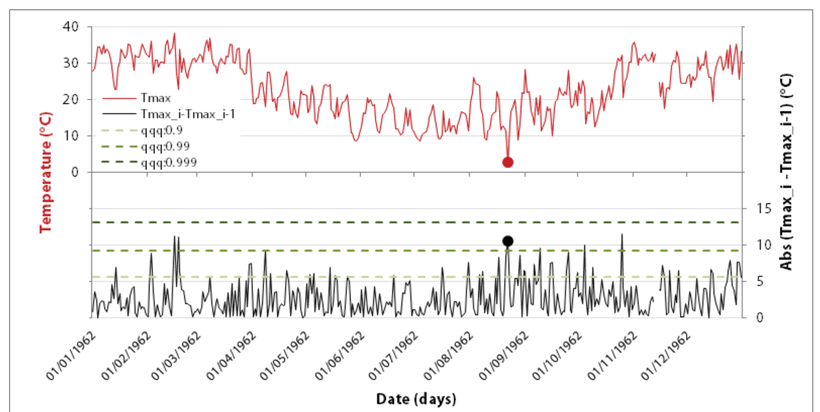

Figure 4. Top panel: Maximum temperature (Tmax, red line). Tmax on 22 August 1962 is flagged as "suspect." Bottom panel: Absolute difference of Tmax values on consecutive days. The horizontal lines indicate different rejection thresholds based on percentiles of observed differences.

Spatial consistency: weighted spatial regression

This test compares a station's data against the data from neighboring stations. First, a series of univariate linear regressions are fitted to obtain estimates of the value at a central station predicted from each neighbor. The various estimates are weighted using an estimate of the standard error of each regression. Large differences between predicted and observed values are flagged as suspect. (Hubbard et al., 2012; Hubbard et al., 2005; Hubbard et al., 2007; Kunkel et al., 2005). See an example of this test in Figure 5.

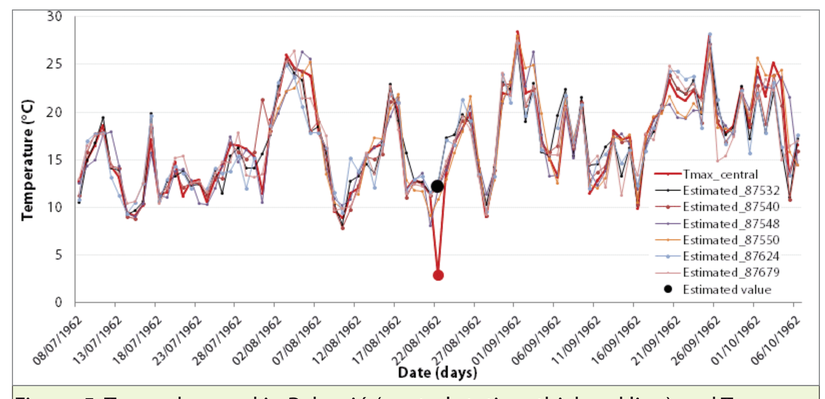

Figure 5. Tmax observed in Pehuajó (central station, thick red line) and Tmax values predicted from various neighboring stations. The weighted estimate of Tmax in Pehuajó for 22 August 1962 (12.1°C) is indicated by the black dot. The suspect recorded Tmax value for that day (2.9°C) is shown as a red dot.

Verification of suspect records

The records flagged as "suspect" by the various QC tests will be subsequently verified against official climate records from participating meteorological services and agencies in the CRC-SAS. Figure 6 illustrates the interface being developed to allow interactive data verification.

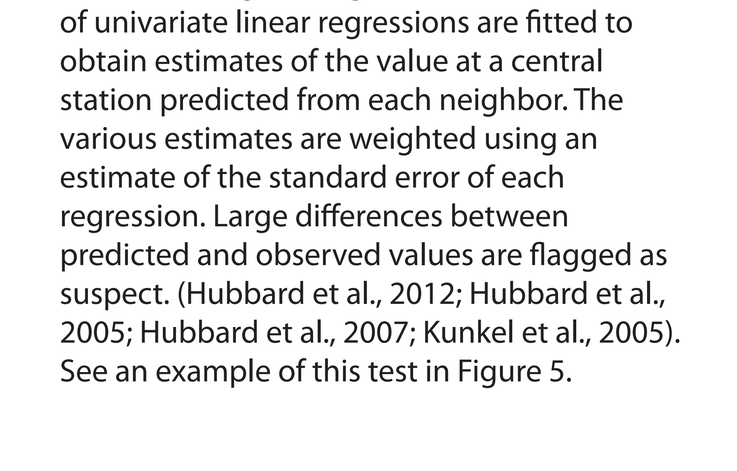

Figure 6. Sample screen of the interface developed for interactive verification of values in the CRC-SAS were identified as "suspect" by the various quality control tests.

Homogeneity

After climate records are quality-controlled and verified, they will be subjected to further tests to assess their homogeneity, i.e., the possible presence of effects not due to climate (relocation of a station, replacement of an instrument). The approach to be used is the state-of-the-art HOMER methodology (Mestre et al., 2013).

References

Hubbard, K., You, J. y Shulski, M., 2012. Toward a Better Quality Control of Weather Data. In: M.S.F. Nezhad (Editor), Practical Concepts of Quality Control. InTech.

Hubbard, K.G., Guttman, N.B., You, J. y Chen, Z., 2007. An Improved QC Process for Temperature in the Daily Cooperative Weather Observations. Journal of Atmospheric and Oceanic Technology, 24(2): 206-213

Hubbard, K.G. y You, J., 2005. Sensitivity Analysis of Quality Assurance Using the Spatial Regression Approach—A Case Study of the Maximum/Minimum Air Temperature. Journal of Atmospheric and Oceanic Technology, 22(10): 1520-1530

Kunkel, K.E. et al., 2005. Quality Control of Pre-1948 Cooperative Observer Network Data. Journal of Atmospheric and Oceanic Technology, 22(11): 1691-1705.

Mestre, O. et al., 2013. HOMER : HOMogenisation softwarE in R- methods and applications. IDÖJÁRÁS - Quarterly Journal of the Hungarian Meteorological Service, 117(1): 47-67.