

# **CONTROL DE CALIDAD DE LA BASE DE DATOS DEL CENTRO REGIONAL DEL CLIMA PARA EL SUR DE AMÉRICA DEL SUR: EFICIENCIA DEL CONTROL EN PEHUAJÓ, ARGENTINA**

**Natalia Herrera<sup>1</sup>, Hernán Veiga<sup>1</sup>, María de los Milagros Skansi<sup>1</sup>, Guillermo Podestá<sup>2</sup>**

[nherrera@smn.gov.ar](mailto:nherrera@smn.gov.ar)

<sup>1</sup>**Servicio Meteorológico Nacional**

<sup>2</sup>**Escuela Rosenstiel de Ciencias Marinas y Atmosféricas. Universidad de Miami**

## **RESUMEN**

Para desarrollar productos y servicios climáticos es indispensable contar con una base de datos climática robusta y confiable, en la que todos sus registros sean de buena calidad. Para ello es necesario contar con un esquema de control de calidad a través del cual se puedan identificar posibles datos erróneos. El Centro Regional del Clima para el sur de América del Sur (CRC-SAS), en colaboración con un proyecto de investigación sobre servicios climáticos financiado por el Instituto Interamericano para la Investigación del Cambio Global (IAI) ha compilado una base de datos diarios que incluye 13 variables meteorológicas, abarcando el período 1961 al presente. Los datos provienen de estaciones convencionales pertenecientes a los servicios meteorológicos e hidrológicos (SMHNs) de Argentina, Bolivia, Brasil, Chile, Paraguay y Uruguay y al Instituto Nacional de Tecnología Agropecuaria (INTA) argentino.

El control de calidad incorpora una serie de procedimientos utilizados por los SMHNs y el INTA, además de otros publicados en la literatura científica. Estos controles se organizaron en seis “familias” que agrupan tests de características similares: (1) Tests generales, (2) Tests de rango fijo, (3) Tests de rango variable (4) Tests de continuidad temporal, (5) Tests de consistencia entre variables y (6) Tests de consistencia espacial.

Un componente crítico del proceso de control de calidad de datos climáticos es la evaluación del desempeño de los diferentes tests. Un test es más eficiente cuando detecta la mayor cantidad de errores verdaderos y tiene una baja tasa de falsos errores (“error de tipo 1”) (Durre et al, 2008). También es importante conocer qué tipos de errores no fueron detectados (“error de tipo 2”).

El objetivo de este trabajo es evaluar la eficiencia de los tests para detectar errores en la temperatura máxima con el fin de poder elegir una configuración de los mismos para utilizar operativamente en el control de calidad de la base de datos del CRC-SAS. Para

ello se usó la serie diaria de temperatura máxima de Pehuajó (Provincia de Buenos Aires, Argentina). Se realizó una inspección de todos los datos para 1961-2012 (18763 datos), encontrándose 387 datos verdaderamente erróneos. Se utilizaron distintas configuraciones para los tests implementados y se examinó separadamente el desempeño de cada test. Las familias de tests de consistencia espacial y de continuidad temporal fueron las que mostraron mayor eficiencia en la detección de errores, mientras que los tests de rango variable fueron los que tuvieron menor eficiencia. Por ejemplo, tomando las configuraciones correspondientes para que todos los tests arrojen aproximadamente 1800 datos sospechosos, uno de los tests de consistencia espacial encontró 159 datos erróneos (41%); uno de los de continuidad temporal encontró 179 (49%); y uno de los tests de rango variable encontró solamente 70 (18%) datos erróneos.

El análisis de la configuración operativa del control de calidad de la base de datos del CRC-SAS arroja que los datos erróneos que no fueron detectados presentaron pequeñas diferencias absolutas con respecto al valor correcto, cercanas a 0°C; en cambio, los errores que fueron detectados presentaron diferencias absolutas de hasta 10.0 °C.

## **ABSTRACT**

Development of climate products and services requires reliable climate data records with good quality. A Quality Control scheme (QC) is necessary to identify possibly erroneous data. In response to these societal needs, the Regional Climate Center for southern South America (RCC-SSA) and a research project on climate services funded by the Inter-American Institute for Global Change Research (IAI) has compiled a database of daily weather variables for first-order meteorological stations operated by meteorological and other research agencies of Brazil, Paraguay, Uruguay, Bolivia, Chile, and Argentina for the period 1961-present. Each daily record includes up to 13 meteorological variables.

The RCC-SSA built a superset of all quality control procedures used by agencies in the region, and added several other published procedures. The quality control procedures can be grouped into six major “families”: (1) general tests, (2) fixed range tests, (3) variable range tests, (4) temporal continuity tests, (5) consistency among variables tests, and (6) spatial consistency tests.

A critical component of a QC process is the evaluation of the performance of different tests. A test is considered to be more efficient when it detects a large proportion of true errors and has a low rate of false alarms (type-I errors) (Durre et al, 2008). It is important also to assess the extent to which true errors remain undetected (type-II errors).

The aim of this study is to evaluate the efficiency of QC tests in detecting true errors in

daily maximum temperature records for Pehuajó (Buenos Aires, Argentina). These results are being used to define appropriate thresholds and parameters for operational QC procedures. An inspection of 1961-2012 data (18,763 values) was performed, and 387 truly erroneous data were found. Different test thresholds and parameters were tested and the performance of each test was examined. The “spatial consistency” and “temporal continuity” tests were more efficient in detecting true errors, whereas “variable range” tests showed lower efficiency. For example, with 1800 “suspicious” values, one of the spatial consistency tests found 159 (41% of the 387 true errors) truly erroneous data; one of the temporal continuity tests found 179 (49%); and one of the variable range tests found only 70 (18%) erroneous data. The truly erroneous values that were not identified by the QC procedures generally showed small absolute differences (close to 0 °C) with respect to the correct value; however, errors that were correctly identified by the QC have larger absolute differences (up to 10.0 °C) with respect to the correct values.

**Palabras clave:** control de calidad, base de datos, servicios climáticos.