

Control de calidad de datos pluviométricos en tiempo real

Nota Técnica SMN 2021-98

**María Sol Rossi Lopardo¹, Hernán Veiga², María Paula
Hobouchian³, Gonzalo Diaz¹, Lorena Ferreira¹**

¹ Dirección de Servicios Sectoriales, Dirección Nacional de Pronósticos y Servicios para la Sociedad.

² Dirección Central de Monitoreo del Clima, Dirección Nacional de Ciencia e Innovación en Productos y Servicios.

³ Dirección de Productos de Modelación Ambiental y de Sensores Remotos, Dirección Nacional de Ciencia e Innovación en Productos y Servicios

Julio 2021

Información sobre Copyright

Este reporte ha sido producido por empleados del Servicio Meteorológico Nacional con el fin de documentar sus actividades de investigación y desarrollo. El presente trabajo ha tenido cierto nivel de revisión por otros miembros de la institución, pero ninguno de los resultados o juicios expresados aquí presuponen un aval implícito o explícito del Servicio Meteorológico Nacional.

La información aquí presentada puede ser reproducida a condición que la fuente sea adecuadamente citada.

Resumen

La detección de posibles errores en los datos de precipitación registrados por las estaciones meteorológicas resulta esencial para la provisión de servicios y el desarrollo de productos de calidad. Para ello, es necesario un procedimiento de control de los datos de precipitación, que opere de forma automática y en tiempo real. En este trabajo se presentan resultados preliminares de dos métodos operativos de control de calidad, a partir de diferentes estimaciones satelitales de precipitación, con el objetivo de filtrar los datos erróneos registrados por las estaciones meteorológicas convencionales y automáticas.

Abstract

Detecting possible errors in rain gauge data is essential for providing services and development quality products. In this context, it is necessary to have an automatic procedure on the precipitation data that operates in real time. In this work we present preliminary results from two operational quality control (QC) methodologies, developed from different satellite precipitation estimations, to filter wrong data recorded by conventional and automatic meteorological stations.

Palabras clave: Precipitación, control de calidad, QC, QPE-Merge

Citar como:

Rossi Lopardo M. S., Veiga H., Hobouchian M. P., Díaz G. y Ferreira L., 2021: Control de calidad de datos pluviométricos en tiempo real. Nota Técnica SMN 2021-98.

Prólogo

Esta Nota Técnica forma parte de una serie que recopila informes realizados en el marco del proyecto “Estimación de precipitación de mejor calidad a partir de la combinación de datos de satélite, radar y estaciones meteorológicas” en el marco del Programa 8.1.1 “Programa de explotación de los datos de sensores remotos” (año 2020). El objetivo final del proyecto es desarrollar e implementar un algoritmo que combine de forma óptima estimaciones de precipitación basadas en sensores remotos (satélite y radar meteorológicos) y observaciones in situ (pluviómetros) para lograr una estimación cuantitativa de precipitación que permita mejorar de manera significativa el monitoreo de la precipitación y otras aplicaciones desarrolladas dentro del Servicio Meteorológico Nacional. Con esta serie de Notas Técnicas se busca documentar el proceso de desarrollo de cada componente del proyecto: 1) observaciones pluviométricas, su integración a través de redes propias y de terceros y el desarrollo de metodologías de evaluación de confiabilidad de esta información y control de calidad en tiempo real, 2) desarrollo de una estimación de precipitación en base a radar meteorológicos, metodologías de control de calidad en tiempo diferido y generación de productos, y 3) desarrollo e implementación de una primera versión de una estimación satelital de precipitación corregida con pluviómetros, cálculo y análisis de distancias de correlación, metodologías de corrección de errores sistemáticos (climatología) y diarios, y validación.

1. INTRODUCCIÓN

La información meteorológica resulta de gran importancia para las diversas situaciones que afecten o modifiquen las actividades socio-económicas de una población. Principalmente la precipitación es una de las variables meteorológicas más esenciales, y por esta razón resulta importante que su medición cuente con una precisión adecuada.

En Argentina, la baja densidad de estaciones meteorológicas es una fuerte limitante cuando se pretende describir la distribución espacial y los gradientes de la precipitación. Esto implica grandes desafíos relacionados al desarrollo de nuevas y mejores técnicas de interpolación de los datos observados. En este contexto, el proyecto *“Estimación de precipitación de mejor calidad a partir de la combinación de datos de satélite, radar y estaciones meteorológicas”* llevado a cabo por el Servicio Meteorológico Nacional (SMN), está orientado al desarrollo de un algoritmo que combine de forma óptima los datos de satélite, radar y estaciones meteorológicas de superficie para obtener una estimación cuantitativa de la precipitación combinada (QPE-Merge, por sus siglas en inglés).

La calidad final del producto QPE-Merge, dependerá de la precisión de la información pluviométrica suministrada por las estaciones meteorológicas, provenientes de diversas redes de observación. La red del SMN está conformada principalmente por estaciones meteorológicas convencionales (EMCs) y en menor medida, estaciones meteorológicas automáticas (EMAs), distribuidas a lo largo del país (Fig. 1). Esta información, como se mencionó con anterioridad, no resulta suficiente, razón por la cual, motivó la utilización e integración de nuevas redes de terceros (Díaz y otros, 2021) y aquellas de países limítrofes, suministradas por el Sistema Mundial de Telecomunicaciones de la Organización Meteorológica Mundial (GTS, por sus siglas en inglés). Si bien los pluviómetros proporcionan una medida de la precipitación en un punto, que generalmente es más precisa que las observaciones provenientes de sensores remotos (Zhang y otros, 2014), existe la posibilidad que sean afectados por diferentes fuentes de error. Estos pueden estar asociados tanto a fallas humanas como a fallas en el instrumental, o a errores en la comunicación de los datos (Qi y otros, 2016).

Esta Nota Técnica describe un conjunto de procedimientos, a partir de información de sensores remotos, orientados a controlar la calidad de los datos de precipitación provenientes de estaciones meteorológicas convencionales y automáticas de las diversas redes mencionadas, con el objetivo de detectar y eliminar valores erróneos o sospechosos. Es importante destacar que los procedimientos aquí descritos relacionados con el segundo método propuesto fueron diseñados para los casos en que las estaciones cuentan con pocos datos para realizar una estadística robusta.

2. MÉTODOS DE CONTROL DE CALIDAD EN TIEMPO REAL

El desarrollo de controles de calidad que operen de forma eficiente y en tiempo real, es un tema complejo y de activa atención por parte tanto de la comunidad académica como de la operativa (Hubbard y otros, 2012). Por un lado, en líneas generales, los métodos tradicionales se basan mayormente en información climatológica de un lugar determinado, cuando se busca, por ejemplo, la detección de valores extremos (Hamada y otros, 2011). Por otro, existen técnicas, como las que lleva a cabo el SMN, orientadas a evaluar la consistencia interna de los datos, es decir, que sean acordes a las situaciones que se reportan.

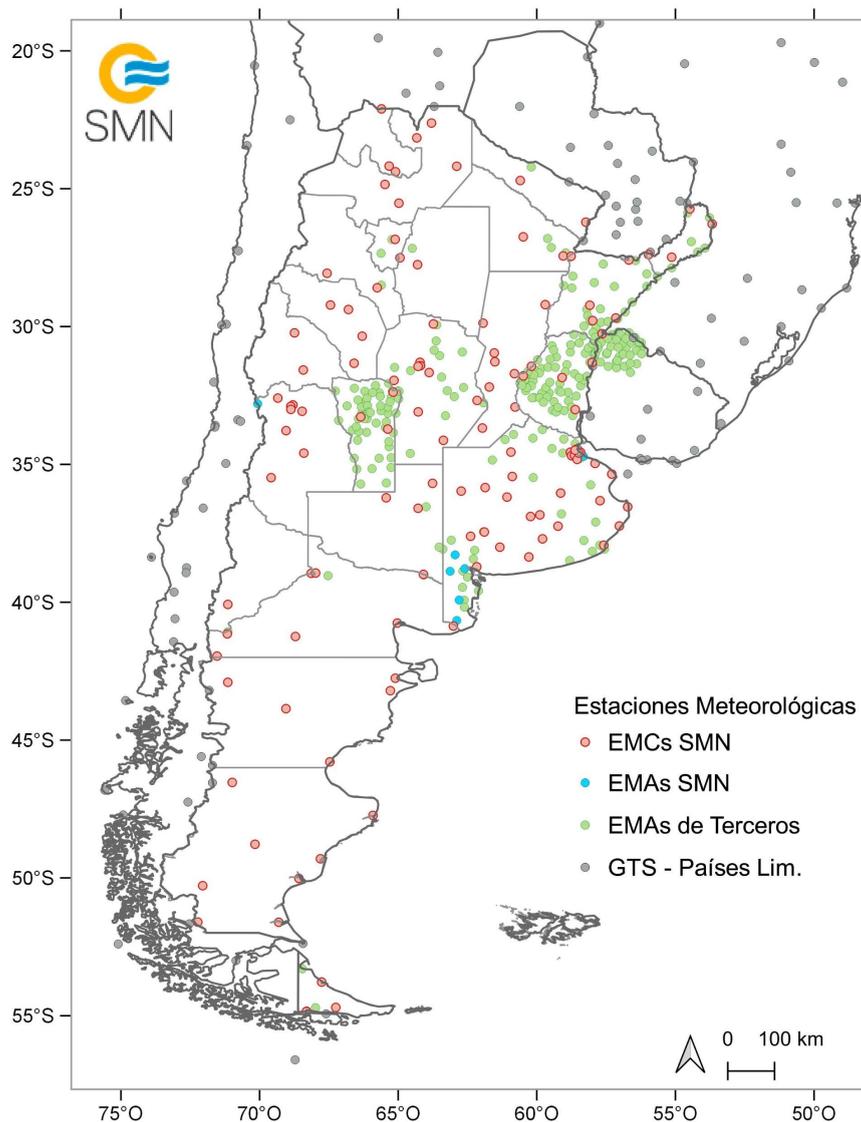


Fig. 1: Ubicación espacial de las estaciones meteorológicas utilizadas para los controles de calidad: Red Nacional de estaciones convencionales propias del SMN (puntos rosas), estaciones meteorológicas automáticas SMN (puntos celestes), estaciones meteorológicas automáticas de terceros (puntos verdes) y estaciones meteorológicas convencionales de la Red Global de Telecomunicaciones -Países Limítrofes (puntos grises), ubicadas en el dominio 56°S - 78°O, 19°S - 48°O.

El SMN realiza una serie de controles de calidad sobre los datos meteorológicos recibidos en tiempo real, es decir, en el momento de la recepción de la información. Estos test únicamente controlan los datos de las estaciones convencionales, pero no son evaluados los registros de las estaciones automáticas.

En el momento que se recibe la información, se realiza un análisis en las estaciones que informan precipitaciones, verificando que en las observaciones horarias precedentes se haya informado, al menos una vez, fenómenos de tiempo presente que indiquen la ocurrencia de precipitación (y viceversa, si los datos horarios informan precipitaciones, los totales acumulados deben reflejarla). Estos controles se aplican a valores acumulados de 6, 12 y 24 horas. Asimismo, se verifica que el registro diario (24 horas) coincida con la suma de los parciales correspondientes, de 6 o 12 horas. Los resultados de estos controles son analizados en tiempo real, y de ser necesario, se dialoga con la estación para validar el dato sospechoso. Para complementar lo anterior, a principios del mes siguiente se vuelven a verificar los valores de precipitación, analizando los datos diarios y mensuales informados en distintos mensajes meteorológicos.

En los últimos años, los métodos de control de calidad han avanzado, incorporando nuevas técnicas que utilizan la información obtenida a través de sensores remotos. Zhang y otros (2014), propusieron controles de calidad simples aplicados a datos horarios de estaciones automáticas, a partir de la información proporcionada por radares meteorológicos. De manera similar, Qi y otros (2016) también utilizaron información a partir de múltiples radares y múltiples sensores (MRMS, por sus siglas en inglés) para ejecutar análisis de calidad de las estaciones.

En línea con los avances mencionados, se desarrollaron e implementaron dos métodos que pretenden examinar y filtrar los valores de precipitación que sean considerados poco confiables, y de esta forma, reducir la incertidumbre del producto final QPE-Merge. Las estaciones meteorológicas que se utilizaron para llevar a cabo los controles son las EMCs propias, EMAs pertenecientes al SMN y de terceros, como también estaciones de países limítrofes, que se encuentran dentro del dominio: 56°S – 78°O, 19°S – 48°O (Fig. 1).

Si bien el SMN lleva a cabo un control de calidad eficiente, como se mencionó con anterioridad, la propuesta de los métodos detallados en esta Nota Técnica podrían ser un complemento de ellos, y un avance necesario para las EMAs que no poseen controles actualmente.

Ambos métodos fueron desarrollados a partir de información proveniente de estimaciones satelitales de precipitación. El primero, orientado a la detección de valores falsos de no-precipitación, mientras que el segundo, pretende identificar valores excesivamente elevados respecto de los valores históricos. Cabe destacar que en este trabajo se realiza una primera aproximación al desarrollo de estas nuevas técnicas de controles de calidad, las cuales no pretenden modificar los valores pluviométricos. El objetivo es identificar y eliminar datos sospechosos o potencialmente erróneos.

Las técnicas utilizadas en los controles de calidad, fueron programadas computacionalmente en lenguaje R (R Core Team, 2021)¹ y se encuentran detalladas en las siguientes secciones.

¹ R es un entorno de programación diseñado para realizar análisis estadísticos y visualizar datos. Corresponde a un software abierto y sin costo, y está disponible para varias plataformas (Windows, Mac OS, Linux) bajo los términos de la Licencia Pública General GNU (GNU-GPL, por sus siglas en inglés; ver <http://www.r-project.org/Licenses/GPL-3>).

2.1 MÉTODO 1: Detección de valores de precipitación “Falso Cero”

El primer control de calidad se diseñó con el objetivo específico de hallar aquellas estaciones meteorológicas que no hayan reportado precipitación (valores de 0 mm), y sin embargo, a partir de información satelital, se pueda inferir la ocurrencia de un evento de precipitación en un momento y lugar determinados. Esta situación nos estaría indicando la existencia de posibles datos erróneos, y que en adelante denominaremos ‘Falso Cero’. Zhang y otros (2014), utilizaron métodos similares para los datos horarios provenientes de las estaciones meteorológicas automáticas, en función a la información de radares meteorológicos, generando una máscara de precipitación/no precipitación. Si la información de precipitación del radar evidenciaba la existencia de un evento precipitante, y la estación reportaba información de no-precipitación (0 mm), se eliminaban los datos proporcionados por esa estación.

Para llevar a cabo este objetivo, se utilizó el producto satelital ‘Mapeo de la Precipitación Global por Satélite’ (GSMaP, por sus siglas en inglés), el cual proporciona la tasa de precipitación global horaria con una resolución espacial de 0.1°. Este producto utiliza un modelo de transferencia radiativa, a partir de la información proveniente de la misión de ‘Medición Global de la Precipitación’ (GPM, por sus siglas en inglés; Okamoto y otros, 2005).

El producto GSMaP consta de varias versiones y actualizaciones realizadas entre la agencia espacial estadounidense y la japonesa (NASA y JAXA, por sus siglas en inglés respectivamente). La validación local de este producto fue realizada para la versión GSMaP-NRT, a partir de la cual se observó que, a pesar de presentar más errores en la cuantificación de la precipitación que otros productos satelitales, es más eficiente para la detección de los eventos de lluvia (Hobouchian y otros, 2018).

En este trabajo se utilizó una versión más reciente, la cual incorpora la corrección de la estimación a partir de los valores observados de las estaciones meteorológicas a nivel global (GSMaP-Gauge, Mega y otros, 2018). Considerando los resultados encontrados por Hobouchian y otros (2018), se puede esperar que el producto continúe manteniendo la eficacia en la detección de los eventos de lluvia, y que al incorporar una corrección con información pluviométrica resulte en una mejor estimación.

La información GSMaP-Gauge fue integrada para transformar los datos horarios en información diaria, para que sea compatible con la información pluviométrica diaria disponible en el SMN. Un día pluviométrico considera la precipitación acumulada entre las 12:00 UTC del día en curso y las 12:00 UTC del día posterior (09:00 HOA a 09:00 HOA).

A partir de la información de la precipitación estimada diaria del producto GSMaP-Gauge, se evaluaron dos alternativas de selección de los píxeles vecinos, que serán utilizados para la detección de los ‘Falsos Cero’ de precipitación. La primera alternativa, define los píxeles vecinos en un área de influencia fija, correspondiente a un radio de 15 km, en cada punto de las estaciones meteorológicas (Fig. 2a). Este criterio sólo considerará los píxeles inmediatos, dado que la resolución espacial de las estimaciones es de 10 km, asumiendo que la información de estos debería ser semejante al píxel de la estación, según el principio de autocorrelación espacial.

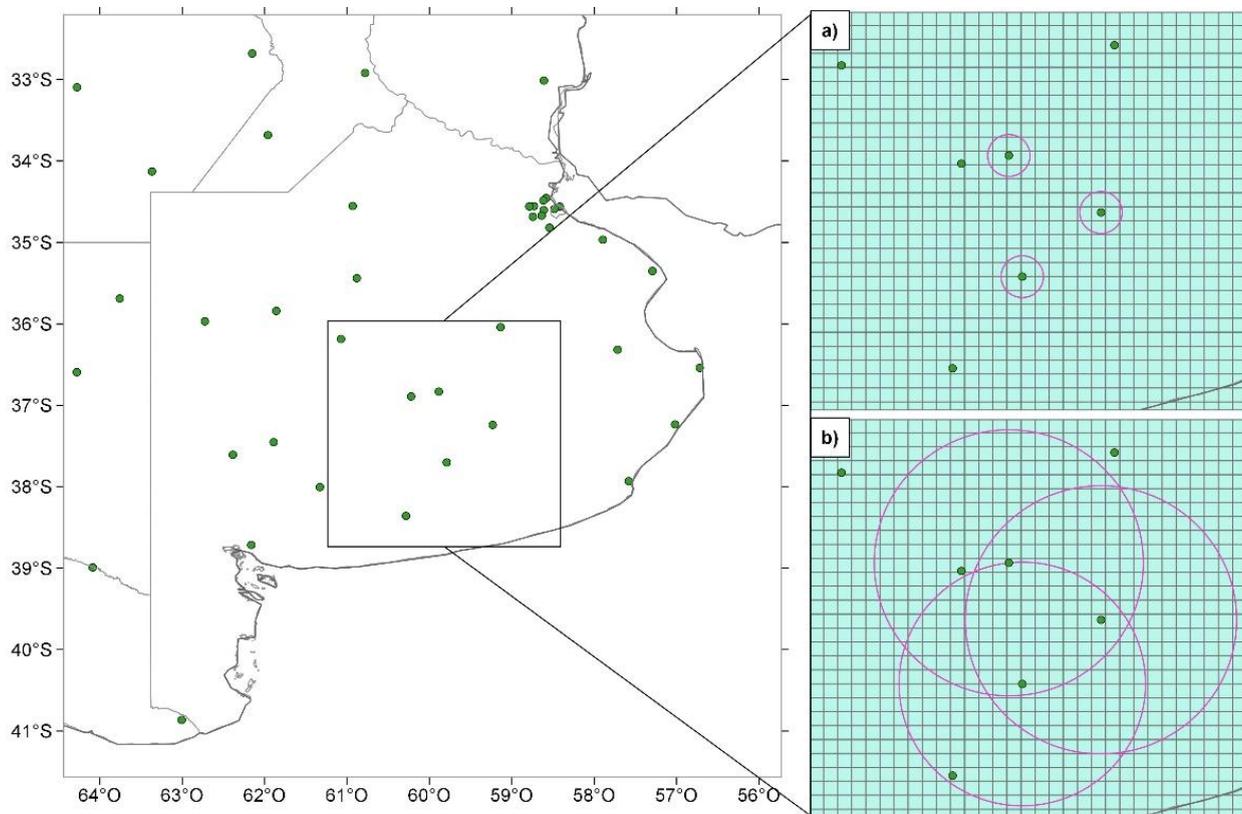


Fig. 2: Representación espacial del área de influencia (círculos rosas), ejemplificado para tres puntos de estaciones meteorológicas (puntos verdes) de la región centro-este de Argentina: (a) con la alternativa de un radio de 15 km y (b) áreas variables a partir de la distancia de correlación del trimestre MAM. El área celeste en (a) y (b) corresponde a la retícula GSMaP de 0.1°.

Para la segunda alternativa, se delimitaron las áreas de forma dinámica (Fig. 2b) utilizando las distancias de correlación espacial calculadas por Hobouchian y otros, (2021). Las distancias de correlación representan la distancia hasta la cual existe influencia del dato de precipitación de una estación meteorológica en particular, y están definidas para cada trimestre del año (diciembre-enero-febrero (DEF), marzo-abril-mayo (MAM), junio-julio-agosto (JJA), septiembre-octubre-noviembre (SON)). Por lo tanto, el radio de influencia, corresponderá al valor de la distancia de correlación que posea cada punto de estación meteorológica, y será variable en función del trimestre correspondiente y en cada estación (Fig. 2b). Cabe destacar que en el caso de las zonas donde no se pudo obtener el dato de distancias de correlación con la metodología mencionada previamente (por ejemplo, en zonas de relieve), se considera el área de influencia de 15 km.

Para determinar la ocurrencia de un evento “Falso Cero”, fueron utilizadas dos variables de entrada: (i) la información pluviométrica diaria proveniente de las estaciones meteorológicas, y (ii) la información diaria de GSMaP-Gauge de los píxeles vecinos de cada estación, delimitados por las dos alternativas de áreas de influencia mencionadas previamente, las fijas y las dinámicas.

La evaluación se realizó a partir de un criterio condicional establecido a cada estación que registra un valor de precipitación de 0 mm (no precipitación). En estos casos, la condición que debe cumplirse para filtrar la estación, es que al menos el 50 % de los píxeles vecinos de la estimación satelital, dentro de las áreas de

influencia, fijas y dinámicas, deberían poseer valores mayores a 1 mm. Esto resultaría en que la información sospechosa de esas estaciones, no sería incorporada al proceso del algoritmo de corrección QPE (Fig. 3).

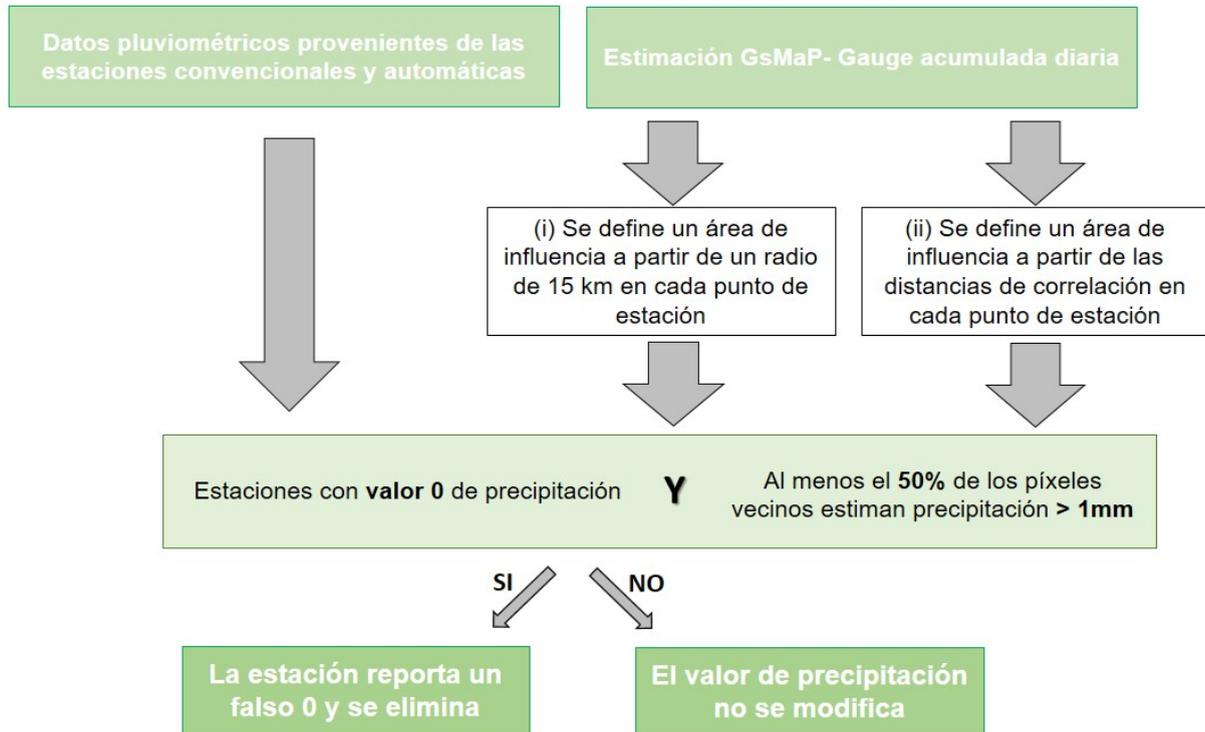


Fig. 3: Diagrama de flujo del método de detección de "Falsos Ceros".

2.2 MÉTODO 2: Detección de valores extremos de precipitación

El segundo método de control de calidad fue diseñado para identificar datos de precipitación extremadamente altos y posiblemente erróneos. El objetivo, entonces, es filtrar los valores que se presentan muy por encima de los observados en el pasado. Para ello se estableció un umbral, obtenido a partir del cálculo de un valor estadístico para cada mes del año. El cálculo del umbral se basa en lo propuesto por varios autores (González-Rouco y otros, 2001; Aguilar y Prohom, 2011) y el Centro Regional del Clima para el sur de América del Sur en su esquema de control de calidad (Veiga y otros, 2014).

Para calcular el valor umbral se utilizaron los datos de la estimación de precipitación IMERG (por sus siglas en inglés; Huffman y otros, 2015) que integra la información de múltiples satélites de la misión GPM en el periodo 2001-2019. Se eligió esta fuente de datos porque demuestra buena habilidad en sus estimaciones en Argentina (Hobouchian y otros, 2018) y cuenta con casi 20 años de información. Utilizar IMERG, en lugar de datos observados, permite calcular el umbral para todas las estaciones, incluso en las que tienen pocos registros.

La estimación de precipitación IMERG utiliza información de sensores de microondas (activas y pasivas) a bordo de múltiples satélites. El algoritmo combina las estimaciones de precipitación recopiladas durante la

misión TRMM (2000-2015) con estimaciones de precipitación obtenidas durante la operación GPM (2014 presente). Para el cálculo de los umbrales se utiliza la versión 'Final Run' en el periodo 2001-2019. Este producto tiene una resolución espacial de 0.1°, cobertura global, latencia de aproximadamente 3 meses, e incorpora un ajuste pluviométrico en su paso final para mejorar su calidad. Los datos originales se integran a paso diario utilizando el criterio de 'día pluviométrico'.

Dada la variabilidad subestacional en las precipitaciones diarias, el cálculo del umbral necesario para el control de calidad de los datos se realiza con ventanas temporales de un mes, logrando una mayor sensibilidad que si fueran estacionales o anuales. A su vez, se utiliza el rango intercuartil (definido como la diferencia entre el percentil 75 y el percentil 25) que es resistente a los valores extremos de las series temporales, por lo tanto es un buen estadístico para utilizar en los controles de calidad que buscan identificar valores extremos. Para cada estación meteorológica se toma el umbral más alto de los calculados en los píxeles comprendidos en un radio de influencia. El radio de influencia se construye de la misma forma que en el Método 1 (Fig. 2): (a) radio de 15 km y (b) basado en el mapa de distancias de correlación espacial de la precipitación estimada en Argentina (Hobouchian y otros, 2021).

El umbral PS_i para cada mes se calcula de la siguiente forma:

$$PS_i = perc90_i + (n \times ri_i)$$

donde $perc90_i$ es el percentil 90 de los valores diarios de precipitación mayor o igual a 1 mm para el mes i (estimado usando todos los años disponibles para ese mes en el registro histórico, por ejemplo todos los febreros de 2001 a 2019), ri_i es el rango intercuartil para ese mes (estimado de la misma manera que el percentil 90) y n es un factor que multiplica a ri_i para definir cuántos mm de lluvia por encima del percentil 90 se consideran para eliminar datos sospechosos (en este caso n es igual a 10).

El cálculo del umbral fue ajustado para detectar valores ampliamente superiores a los registrados en el pasado. Se modificó el percentil utilizado, optando por el percentil 90 en lugar de 75, mientras que el factor n se fijó en 10. Notar que otros autores utilizan $n=5$. Los valores diarios de precipitación que resulten mayores que el valor umbral serán eliminados por el control y no serán incorporados en el proceso del algoritmo de corrección QPE (Fig. 4).

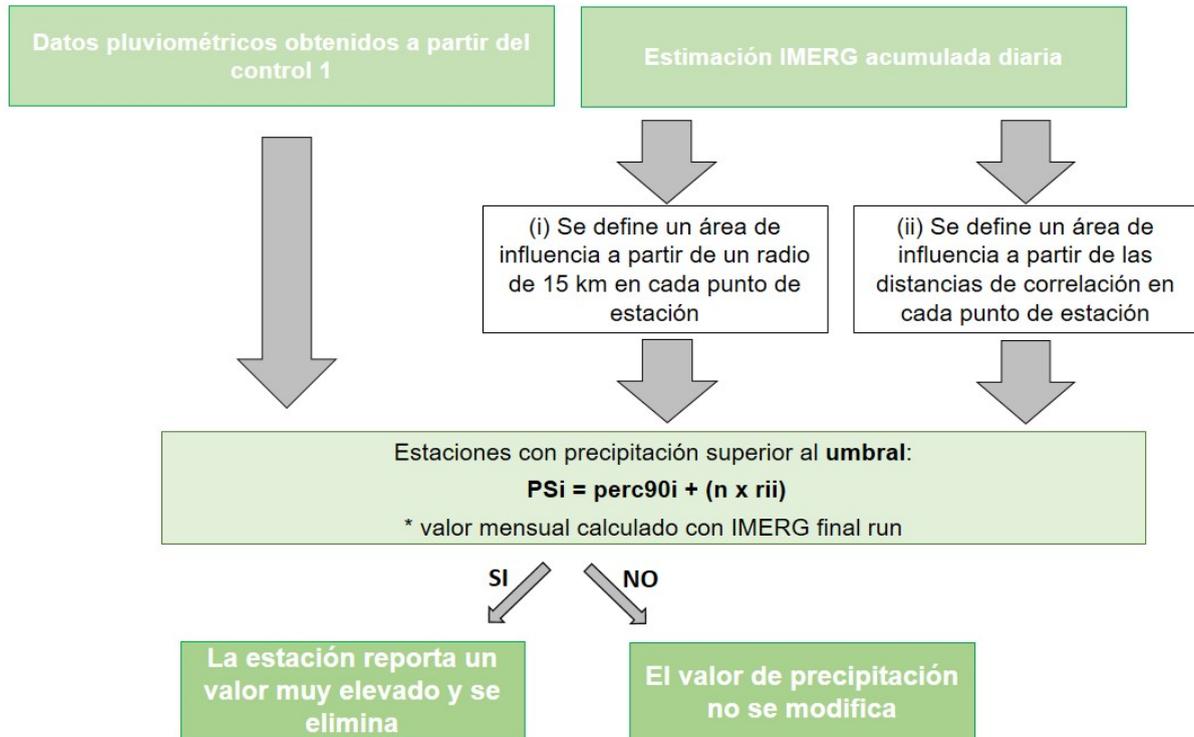


Fig. 4: Diagrama de flujo del método de detección de valores extremos de precipitación.

3. DESCRIPCIÓN DE LOS RESULTADOS

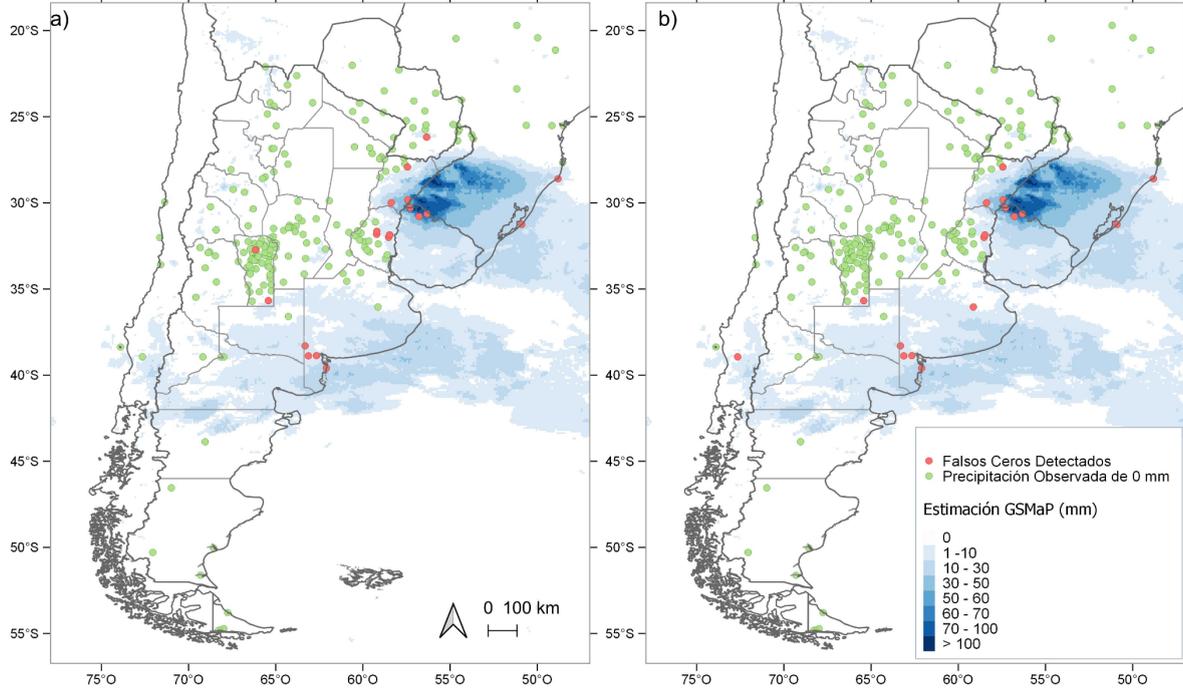
3.1 Resultados del MÉTODO 1

Para evaluar el primer control de calidad, se realizó el test de 'Falsos Ceros' para tres fechas diferentes: 04-05-2020, 19-10-2020 y 18-11-2020.

En la Fig. 5 se presentan los mapas correspondientes al dominio analizado, donde se puede observar la estimación de precipitación GSMaP-Gauge correspondiente a cada fecha (1) 04-05-2020, (2) 19-10-2020 y (3) 28-11-2021. Adicionalmente, se visualizan las estaciones meteorológicas que reportaron un valor de precipitación de 0 mm (puntos verdes), y las estaciones en las cuales se detectó un 'Falso Cero' (puntos rosas) a partir de las áreas de influencia (a) fijas y (b) variables.

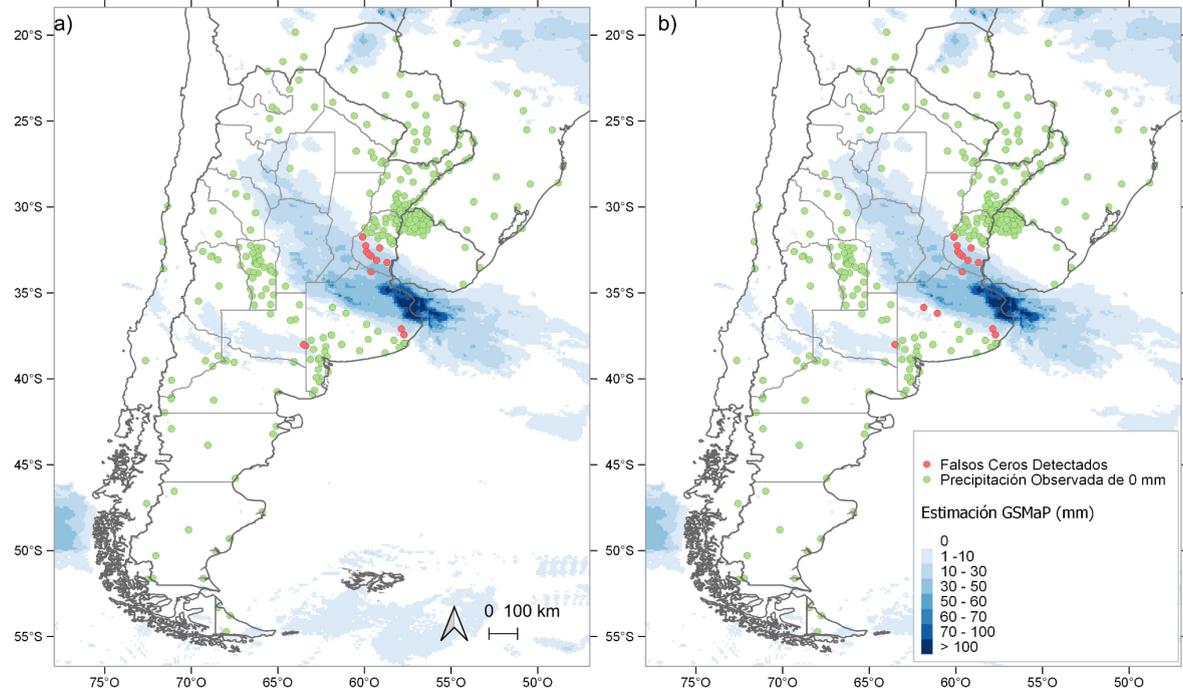
1. Servicio Meteorológico Nacional

Primer Control: Detección de Falso Cero
04 de Mayo del 2020



2. Servicio Meteorológico Nacional

Primer Control: Detección de Falso Cero
19 de Octubre del 2020



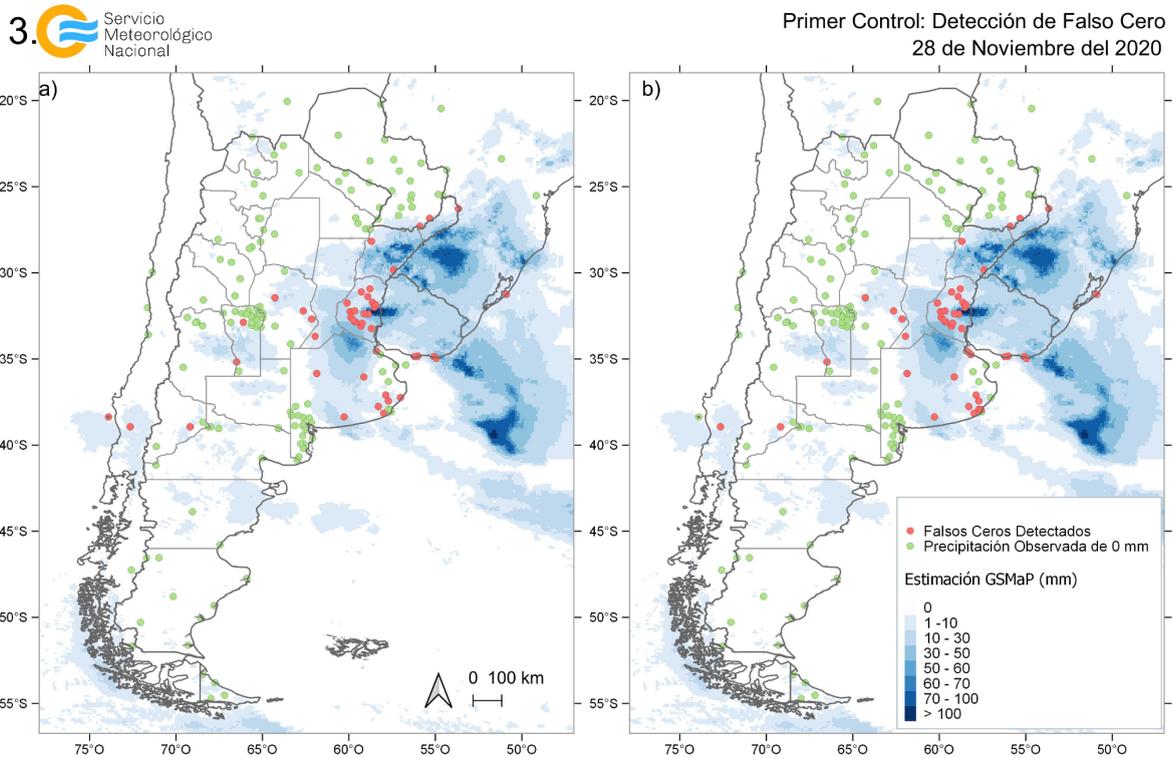


Fig. 5: Resultados del primer control de 'Falso Cero' para las tres fechas de prueba (1) 04-05-2020, (2) 19-10-2020 y (3) 28-11-20, en el dominio 56°S – 78°O, 19°S – 48°O. La estimación GSMaP-Gauge se encuentra representada con una gama de celestes-azules, dónde celestes más claros indican menores valores de precipitación y tonos azules indican mayores valores de precipitación. Los puntos verdes representan las estaciones meteorológicas con valor de precipitación 0 mm, y los puntos rojos corresponden a las estaciones filtradas a partir de aplicar las (a) áreas de influencia fijas establecida en un radio de 15 km para cada estación y (b) áreas de influencia delimitadas a partir de los valores de distancia de correlación.

Los resultados obtenidos a partir de aplicar el primer control se resumen en la Tabla 1. Las estaciones meteorológicas filtradas por 'Falsos Ceros' a partir de aplicar áreas de influencia fijas fueron, en términos generales, menores que con las áreas de influencia variables. Sin embargo, estas diferencias resultaron ser muy pequeñas. El test aplicado durante el día 28-11-2020, filtró alrededor del 23% del total de estaciones meteorológicas con valores de 0 mm de precipitación, con ambos métodos de selección de píxeles vecinos.

Tabla 1: Cantidad total de estaciones meteorológicas filtradas a partir del Método 1 para las diferentes fechas de análisis.

| Fecha | 04.05.2020 | 19.10.2020 | 28.11.2020 |
|---|------------|------------|-------------|
| <i>Estaciones meteorológicas totales con reporte</i> | 446 | 460 | 459 |
| <i>Estaciones con valor de precipitación igual a 0 mm</i> | 216 | 310 | 198 |
| <i>Estaciones filtradas con áreas de influencia fijas</i> | 19 (8,7%) | 13 (4,19%) | 46 (23,23%) |
| <i>Estaciones filtradas con áreas de influencia variables</i> | 17 (7,8%) | 14 (4,5%) | 47 (23,73%) |

Adicionalmente, se observó que, las estaciones removidas en función del tipo y propietario de estación, fueron diferentes en ambos test aplicados, aunque en su mayoría pertenecían a EMAs de terceros (Fig. 6). Sin embargo, se encontró que el test de áreas de influencia dinámica, filtró en todas las fechas de análisis, de 6% a 17% de las estaciones convencionales pertenecientes al SMN (Fig. 6b). De manera contraria, el test de áreas fijas, sólo filtró el 18% el día 28-11-2020.

Con el objetivo de identificar las redes de EMAs de terceros que fueron detectadas con valores 'Falsos Ceros', se evaluó el porcentaje de estaciones filtradas respecto al total de cada red en particular, considerando los datos de las tres fechas de análisis en conjunto. Debido a las pocas fechas en las que se probó el test, sólo se identificaron nueve redes que presentaron valores sospechosos con el método aplicado: Bolsa de Cereales de Córdoba, Agua y Saneamientos Argentinos (AySA), Ministerio de Producción de Corrientes, Instituto Nacional de Tecnología Agropecuaria (INTA), Bolsa de Cereales de Entre Ríos, Sistema Nacional de Radares Meteorológicos (SINARAME), Represa de Salto Grande, Sudoeste de Buenos Aires (SOBA) y Universidad de la Punta de San Luis (Fig. 7).

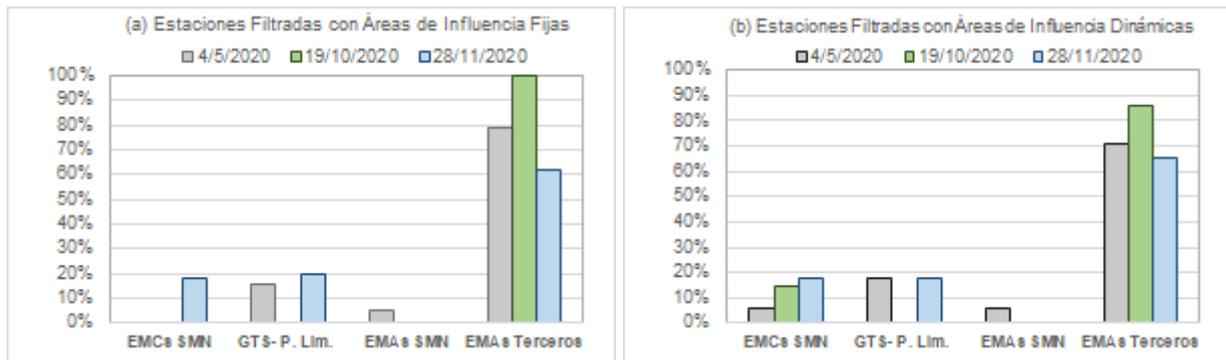


Fig. 6: Porcentaje de estaciones filtradas según el tipo y propietario de estación: estaciones convencionales SMN (EMCs SMN), estaciones convencionales GTS - Países Limítrofes (GTS-P.Lim), estaciones automáticas SMN (EMAs SMN) y estaciones automáticas de terceros (EMAs Terceros), respecto al total de las estaciones filtradas para cada fecha de análisis: 04-05-2020 (barras grises) – 19-10-2020 (barras verdes) – 28-11-2020 (barras celestes), según (a) áreas de influencia fijas y (b) áreas de influencia variables.

Particularmente, se observó que el 100% de las estaciones pertenecientes a la red AySA fueron eliminadas, cuando se utilizaron áreas de influencia dinámicas, y el 32% aplicando áreas de influencia fijas. Por otro lado, el test identificó el 32% de las estaciones de la red SOBA con valores 'Falso Cero', con ambas áreas de influencia. La Red de Bolsa de Cereales de Entre Ríos y del Ministerio de Producción de Corrientes, presentaron valores similares con la implementación de ambas áreas de influencia, considerando aproximadamente al 20% de las estaciones de dichas redes con valores 'Falso 0'. Respecto a las redes restantes, el test identificó menos del 10% de las estaciones con valores sospechosos.

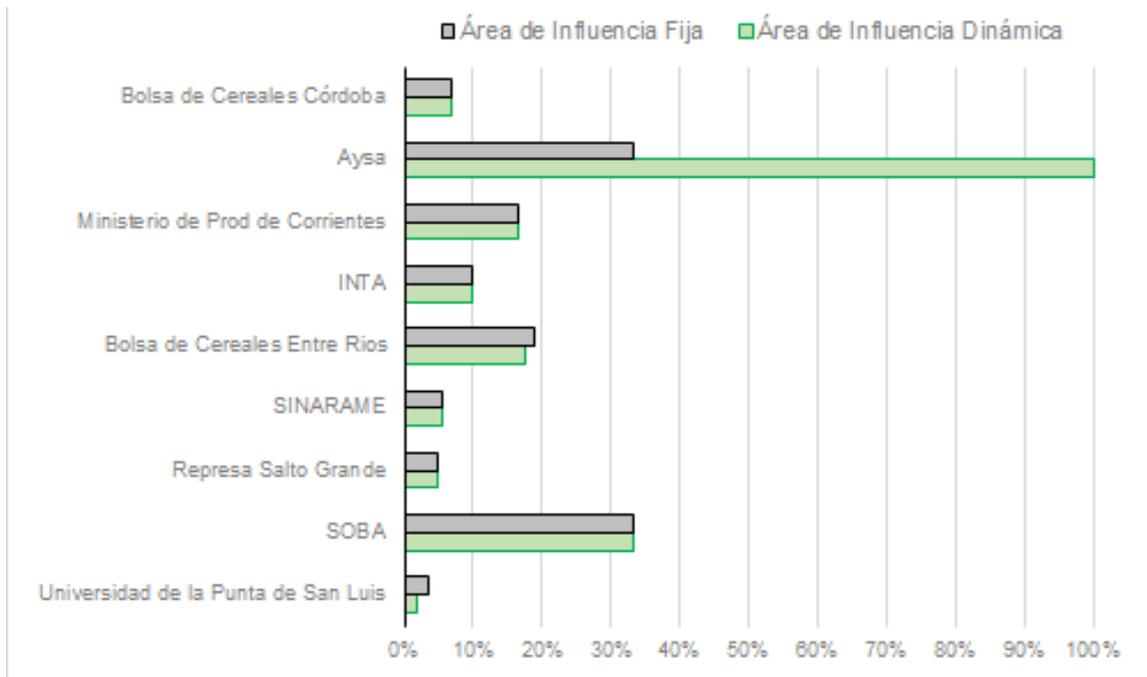


Fig. 7: Porcentaje de estaciones filtradas, respecto al total de datos de cada red, según la selección de áreas de influencia fija (barras grises) y áreas de influencia dinámica (barras verdes).

3.2 Resultados del MÉTODO 2

La evaluación del segundo método de control de calidad de datos de precipitación se realizó en tres estaciones meteorológicas, dos convencionales (Dolores y Resistencia) y una automática (San Germán). El objetivo de este test es detectar valores de precipitación extremadamente altos, con umbrales específicos para cada estación meteorológica y mes del año. A modo de ejemplo, en los tres casos se utilizó un área de influencia fija de 15 km de radio para el cálculo de los umbrales. Los ejemplos están acompañados por gráficos donde se presenta la precipitación diaria en un periodo de 4/5 meses, junto con los umbrales de detección de valores sospechosos.

En el primer ejemplo, San Germán (provincia de Buenos Aires, Fig. 8), se puede ver como las precipitaciones informadas los días: 8/11/2020 y 21/4/2021 superan al umbral mensual utilizado para detectar valores sospechosos. Los datos de estos días son eliminados por el test de control de calidad.

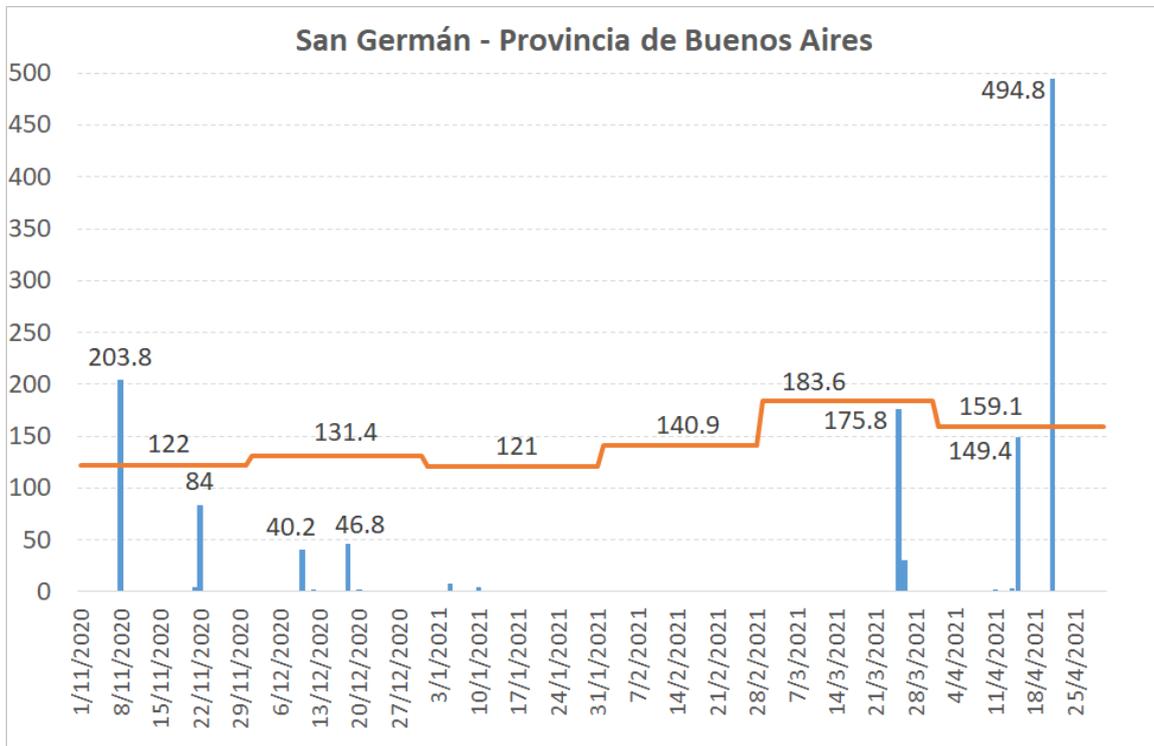


Fig. 8: Precipitaciones diarias en San Germán (barras azules), y umbrales mensuales (líneas rojas).

En el caso de Resistencia (provincia de Chaco, Fig. 9), ningún valor diario de precipitación en el periodo seleccionado superó el umbral de control. No obstante, el día 8/1/2019 se registraron 224 mm (máximo valor entre 1964 y 2020), y el umbral para detectar valores sospechosos del primer mes del año es 361.3 mm. En principio, los umbrales utilizados en Resistencia parecen ser elevados.

En el último ejemplo, Dolores (provincia de Buenos Aires, Fig. 10), el día 5/1/2021 la estación meteorológica informó 276 mm (valor récord para la serie 1931-2020). Este valor resultó superior al umbral para el mes de enero (172.4 mm) y fue eliminado por el sistema. Sin embargo, el valor fue corroborado, es un valor correcto, extremadamente superior a los observados en el pasado. El segundo valor más alto de la serie de precipitación diaria en Dolores para el periodo 1931-2021 es 142.3 mm, registrado el 20/04/1970 y en los meses de enero el segundo valor más alto es 135 mm (28/01/1990). Por lo que el umbral establecido parece adecuado, si bien en este caso llevó a eliminar un valor correcto.

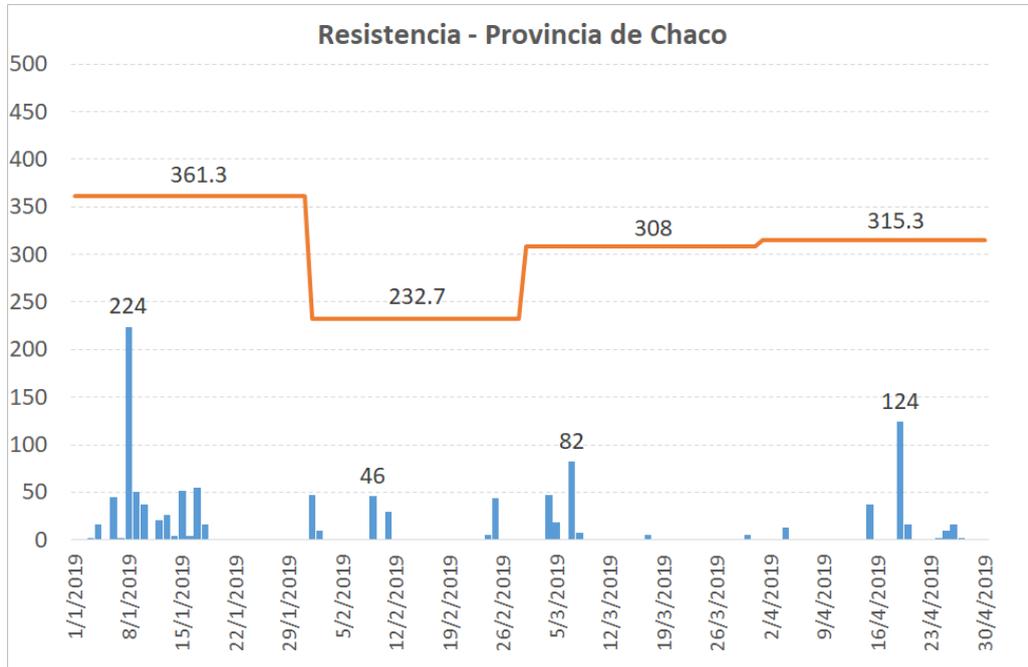


Fig. 9: Precipitaciones diarias en Resistencia (barras azules), y umbrales mensuales (líneas rojas).

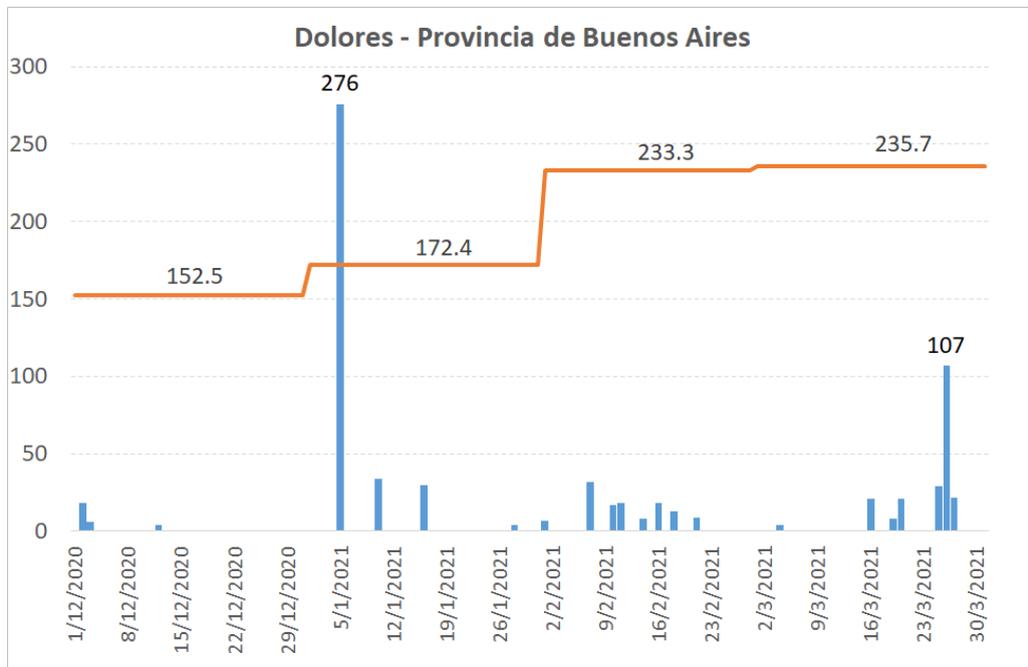


Fig. 10: Precipitaciones diarias en Dolores (barras azules), y umbrales mensuales (líneas rojas).

4. DISCUSIÓN Y CONCLUSIÓN

Existen diversas técnicas para consistir datos meteorológicos a través de un proceso de control de calidad. Se pueden utilizar métodos estadísticos para analizar las series temporales de los distintos parámetros. También los test pueden incluir registros de estaciones cercanas para realizar la evaluación, o como en este caso, utilizar datos de sensores remotos. La ventaja de esta última opción es la posibilidad de contar con información disponible más allá del tiempo de funcionamiento de cada estación meteorológica o de la densidad de estaciones en la zona. La desventaja es que esta fuente de datos puede tener distintos problemas que pueden llevar a errores en la evaluación realizada por los controles de calidad, asociados a errores en los valores de cada estimación satelital.

Para evaluar los datos que se utilizan en la primera versión del producto QPE-Merge, se analizaron técnicas alternativas sencillas, y se priorizaron procesos de rápida implementación. Los resultados preliminares de las pruebas realizadas hasta este momento indican que el primer control de calidad, orientado a la detección de 'Falsos Ceros', no ha mostrado grandes diferencias entre las dos alternativas de selección de áreas de influencia en el total de estaciones filtradas. Sin embargo, se encontraron diferencias del tipo de redes filtradas, en particular sobre la cantidad de EMCs removidas. De las tres fechas analizadas se observó que el día 28-11-2020, el método identificó un gran número de estaciones con posibles valores de 'Falsos Ceros'. Para esa fecha particular, se desestimaron los datos de varias estaciones convencionales, representando un 18% y 17% del total de las estaciones filtradas con las técnicas de selección de píxeles vecinos de área fija y dinámica, respectivamente. Este resultado podría conducir a la pérdida de información de campo valiosa, ya que como se mencionó con anterioridad, el SMN realiza controles de calidad en tiempo real sobre las estaciones convencionales, por lo que en general esta información se considera precisa y confiable. En este sentido, el porcentaje de estaciones convencionales eliminadas debería ser bajo. Si bien el objetivo de esta Nota Técnica no es generar una evaluación de las redes automáticas de terceros, resultó de interés evaluar la cantidad de estaciones pertenecientes a cada red con valores sospechosos en la aplicación del método 1. Continuar con este análisis podría contribuir a la evaluación de calidad de las redes automáticas de terceros.

Asimismo, el segundo control necesita realizar más pruebas y ajustes. Por ejemplo, en Dolores se eliminó un valor correcto, si bien fue un valor extremadamente superior a los registrados en el pasado. El caso opuesto es Resistencia, donde los valores umbrales son elevados. Sin embargo, el método demostró ser eficiente en San Germán. En general, más allá de los ejemplos presentados, los umbrales calculados parecen funcionar correctamente. A su vez, es importante validar los resultados obtenidos utilizando como fuente de información estimaciones satelitales de precipitación. Para ello, la propuesta es replicar este control utilizando datos observados en las estaciones meteorológicas que cuenten con series suficientemente largas. El objetivo de esta tarea es comparar los valores umbrales obtenidos con ambos métodos, y analizar la robustez de los umbrales calculados con datos de sensores remotos.

Luego de este primer paso, resulta de interés continuar con diferentes evaluaciones y ajustes. Se considera principalmente importante evaluar el porcentaje de píxeles vecinos (en el método 1), la aplicación de diferentes umbrales, ya sea del valor de precipitación (en el método 1), el cálculo del umbral mensual (en el método 2), como del radio de influencia a utilizar (en ambos métodos), existiendo la posibilidad de que el área de influencia no necesariamente debiera tener una forma circular perfecta. Esto se debe a que la continuidad espacial de la precipitación puede cambiar marcadamente según la dirección que se esté evaluando (anisotropía). Por lo tanto, los pasos próximos deberían incorporar diferentes pruebas de los valores umbrales y del radio de influencia, considerando el tipo de proceso atmosférico que estuviese

generando el evento precipitante. El objetivo principal, será optimizar la tasa de falsos errores (datos correctos que son eliminados erróneamente).

En esta primera instancia, los métodos desarrollados no pretenden corregir los valores de precipitación, sino descartar del QPE aquellos datos potencialmente erróneos o sospechosos. Este proceso de control de calidad de datos meteorológicos es un primer paso, que busca que los valores de precipitación sean más confiables y mejorar la calidad del producto QPE. Para lograr este objetivo, es necesario ajustar ambos métodos y lograr una efectiva tasa de detección de valores erróneos o al menos sospechosos. En este sentido, es importante destacar que la evaluación del funcionamiento de los controles, requiere un análisis más completo que permita estudiar el impacto final en el producto QPE-Merge, para tomar una decisión adecuada acerca de su utilidad. A su vez, continuar estudiando nuevos controles de calidad permitiría identificar errores que los métodos desarrollados no detecten.

5. REFERENCIAS

- Aguilar, E. y Prohom, M., 2011: RClimDex-extraQC (EXTRAQC Quality Control Software). User Manual. Centre for Climate Change, C3, Universitat Rovirai Virgili, Tarragona, España. Disponible en <http://www.c3.urv.cat/softdata.php>
- Díaz G., Vita M., Hobouchian M. P. , Ferreira L. y Giordano L., 2021: Expansión de la red de referencia empleando los datos de precipitación de las estaciones meteorológicas automáticas de terceros. Nota Técnica SMN 2021-90.
- González-Rouco J.F., Jiménez J.L., Quesada V., Valero F., 2001: Quality Control and Homogeneity of Precipitation Data in the Southwest of Europe. *Journal of Climate*, 14(5): 964-978.
- Hamada A., Arakawa O., Yatagai A., 2011: An automated quality control method for daily rain gauge data. *Global Environ. Res.*, 15, 165–172.
- Hobouchian, M. P., R. de Elía, Y. García Skabar, L. Vidal, J. Ruiz y M. Maas, 2021: Distancias de correlación espacial de la precipitación en Argentina. Nota Técnica SMN 2021-96.
- Hobouchian M. P., García Skabar Y., Salio P., Viale M., Matsudo C., 2018: Evaluación de estimaciones de precipitación por satélite en el sur de Sudamérica. XIII Congreso Argentino de Meteorología, Rosario, Argentina.
- Huffman G. J., Bolvin D. T., Nelkin E. J., 2015: Day 1 IMERG Final Run Release Notes. http://pmm.nasa.gov/sites/default/files/document_files/IMERG_FinalRun_Day1_release_notes.pdf
- Mega T., Ushio T., Matsuda T., Kubota T., Kachi M., Oki R., 2018: Gauge-Adjusted Global Satellite Mapping of Precipitation. *IEEE Transactions on Geoscience and Remote Sensing*, 57, 1928 – 1935.
- Okamoto K. i., Ushio T., Iguchi T., Takahashi N., Iwanami K., 2005: The global satellite mapping of precipitation (GSMaP) project. *Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, IGARSS '05.*
- Qi Y., Martinaitis S., Zhang J., & Cocks S., 2016: A Real-Time Automated Quality Control of Hourly Rain Gauge Data Based on Multiple Sensors in MRMS System. *Journal of Hydrometeorology*, 17(6), 1675–1691.
- R Core Team, 2021: R: A language and environment for statistical computing. R Foundation for Statistical Computing. <http://www.R-project.org>.
- Veiga H., Herrera N., Skansi M. M., Podestá G.P., 2014: Reporte Técnico CRC-SAS-2014-001: Descripción de controles de calidad de datos climáticos diarios implementados por el Centro Regional del Clima para el Sur de América del Sur. https://www.crc-sas.org/es/guias_crc.php.
- Zhang J., Qi Y., Langston C., Kaney B., & Howard K., 2014: A Real-Time Algorithm for Merging Radar QPEs with Rain Gauge Observations and Orographic Precipitation Climatology. *Journal of Hydrometeorology*, 15(5), 1794–1809.

Instrucciones para publicar Notas Técnicas

En el SMN existieron y existen una importante cantidad de publicaciones periódicas dedicadas a informar a usuarios distintos aspectos de las actividades del servicio, en general asociados con observaciones o pronósticos meteorológicos.

Existe no obstante abundante material escrito de carácter técnico que no tiene un vehículo de comunicación adecuado ya que no se acomoda a las publicaciones arriba mencionadas ni es apropiado para revistas científicas. Este material, sin embargo, es fundamental para plasmar las actividades y desarrollos de la institución y que esta dé cuenta de su producción técnica. Es importante que las actividades de la institución puedan ser comprendidas con solo acercarse a sus diferentes publicaciones y la longitud de los documentos no debe ser un limitante.

Los interesados en transformar sus trabajos en Notas Técnicas pueden comunicarse con Ramón de Elía (rdelia@smn.gov.ar), Luciano Vidal (lvidal@smn.gov.ar) o Martin Rugna (mrugna@smn.gov.ar) de la Dirección Nacional de Ciencia e Innovación en Productos y Servicios, para obtener la plantilla WORD que sirve de modelo para la escritura de la Nota Técnica. Una vez armado el documento deben enviarlo en formato PDF a los correos antes mencionados. Antes del envío final los autores deben informarse del número de serie que le corresponde a su trabajo e incluirlo en la portada.

La versión digital de la Nota Técnica quedará publicada en el Repositorio Digital del Servicio Meteorológico Nacional. Cualquier consulta o duda al respecto, comunicarse con Melisa Acevedo (macevedo@smn.gov.ar).