

Propuesta metodológica para completar series de tiempo mensuales de temperatura cuando no existen estaciones adyacentes

Methodological proposal to filling monthly temperature gaps in time series without adjacent stations

Luis Bernardo Bastidas,* Alberto Ismael Juan Vich** y María Cintia Piccolo***

Recibido: 12/11/2019. Aprobado: 12/06/2020. Publicado en línea: 21/10/2020.

Resumen. El uso de series temporales sin valores faltantes resulta fundamental para la veracidad de los análisis científico-geográficos. Dado que en el caso de la temperatura (variable objeto de estudio), es posible encontrar vacíos de información en sitios con escasas estaciones meteorológicas, el objetivo de investigación fue desarrollar una metodología para completar series de tiempo mensuales cuando no se dispone de estaciones meteorológicas adyacentes. El método planteado se vincula al uso de temperaturas anuales, medias mensuales y promedio seriales, de modo que su proporción permita calcular los datos carentes de información. Tomando como base una estación meteorológica localizada al centro-occidente árido argentino, la investigación se centró en cuatro fases vinculadas a la aplicación de la propuesta. Inicialmente, se aplicó la metodología planteada para calcular valores faltantes mensuales sobre la estación San Juan Aero, con $N=46$ años. Adicionalmente, se comparó la propuesta con otras metodologías que utilizan estaciones adyacentes y su aplicación sobre otra estación fuera del área de estudio. Todo lo anterior fue verificado mediante nueve índices

de precisión. Los resultados indican que la metodología propuesta presenta óptimos resultados en estaciones con $N \geq 30$ años, 10 % de datos faltantes y con precisión superior a otras metodologías de naturaleza similar. Se utilizó el método propuesto en series de tiempo pertenecientes a estaciones fuera del área de estudio y los resultados mostraron una alta fiabilidad en el cálculo de datos faltantes. Por lo tanto, se recomienda su uso para completar series temporales de temperaturas mensuales en estaciones sin pares adyacentes.

Palabras clave: datos faltantes, temperatura mensual, índices de precisión, estaciones adyacentes, series temporales.

Abstract. Complete time series with no missing values are essential for reliable scientific-geographic analyses. Temperature time series commonly show data gaps, particularly in meteorological stations located in regions with few scattered stations. Scarce meteorological stations exist in the arid

* Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales IANIGLA-Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Ruiz Leal s/n, Parque San Martín, Mendoza, República Argentina, 5500. Departamento de Geografía y Turismo DGyT – Universidad Nacional del Sur UNS. 12 de Octubre y San Juan, Bahía Blanca, República Argentina. 8000. <https://orcid.org/0000-0002-8570-9734>. Email: lbastidas@mendoza-conicet.gob.ar

** Instituto Argentino de Nivología, Glaciología y Ciencias Ambientales IANIGLA-Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Ruiz Leal s/n, Parque San Martín, Mendoza, República Argentina, 5500. Facultad de Filosofía y Letras FFyL – Universidad Nacional de Cuyo UNCUYO. Centro Universitario, Mendoza, República Argentina. M5502JMA. <https://orcid.org/0000-0002-0477-2476>. Email: aijvich@mendoza-conicet.gob.ar

*** Instituto Argentino de Oceanografía IADO - Consejo Nacional de Investigaciones Científicas y Técnicas CONICET. Florida 450, Bahía Blanca, República Argentina, 8000. Departamento de Geografía y Turismo DGyT-Universidad Nacional del Sur UNS. 12 de Octubre y San Juan, Bahía Blanca, República Argentina, 8000. <https://orcid.org/0000-0002-5184-9149>. Email: oipiccol@criba.edu.ar

central-western region of Argentina, where vast, sparsely populated, or unproductive areas far away from major urban centers and oases may have restrained the installation of sufficient stations. Thus, climate data records from existing stations, especially those in rural areas, often lack temporal continuity, and the data gaps have to be filled in based on data from adjacent stations.

However, this is not possible in the absence of nearby stations with reliable and sufficiently long records that can be used for estimating the missing data. This study aimed to develop an easy-to-apply, highly accurate operational method to fill data gaps in monthly temperature time series, which is particularly suitable for locations with no nearby meteorological stations.

The method developed herein is based on the use of annual and monthly means and the overall time series average. The method was tested on the 46-yr time series of monthly temperature data recorded at the meteorological station of San Juan Aero (base station), located in the Province of San Juan, central-western Argentina.

The base station is close to two other weather stations whose data were used to validate the results of one of the phases of the method. The study included: a) the application of the method proposed to the San Juan Aero station (base station) using subsets of varying lengths of the entire time series data set, and comparing the accuracy of the estimates thus obtained by means of ad hoc indices; b) the application of the same procedure used in a), but with varying percentages of missing data; c) the comparison of the missing values estimated by the method developed herein versus those estimated using conventional methods based on data from adjacent stations, and d) the application of the method developed herein to a meteorological station located outside the study area under different climate and environmental conditions.

The method proposed (Ti) estimates missing monthly temperature values as the product of the average of the mean annual temperatures of the years immediately before and after the year with missing monthly data multiplied by the

mean temperature of the target month, divided by twice the mean annual temperature of the data series.

The method was used to estimate missing monthly temperature values for the 46-yr time series recorded in the San Juan Aero station. Tests were run to determine the percentage of missing data (5%, 10%, and 15%) with which the method yields the best fit. The efficiency of the Ti method was compared versus three traditional methods (arithmetic mean, normal proportion, and inverse distance weighting) that impute the missing values from data recorded at nearby stations. Finally, based on the results from the previous stages, the Ti method was applied to a test station located some 150 km from the baseline station to determine whether it can also be applied to meteorological stations located outside the study region, under different physical environmental characteristics.

The results showed that the Ti method works better on meteorological stations having at least 30-yr records and no more than 10% missing data; under these conditions, its estimates are more accurate than those yielded by the three traditional methods tested and can be reliably applied to stations located outside the study region under different physical and environmental conditions.

The limitations of the Ti method are worth mentioning: it cannot be used when the baseline station records have a data gap longer than one full year, or when data for the same month are missing for two consecutive years. Given the results yielded by this method and taking into account the limitations mentioned above, compared to other methods that use data from nearby stations, we recommend using the Ti method to estimate missing monthly temperature values for meteorological stations lacking nearby stations.

Compared to the traditional methods tested, the Ti method seems highly valuable as a tool to fill missing data in temperature time series from isolated weather stations, which then could be used for climate analyses of remote zones.

Keywords: missing data; monthly temperature; accuracy indices; nearby weather stations; time series.

INTRODUCCIÓN

La temperatura es uno de los indicadores climáticos más representativos, indicativa de variados fenómenos geográficos en los territorios. La Organización Meteorológica Mundial, OMM (WMO, 2018), establece que los registros de las variables climáticas (como la temperatura), deben ser continuos y homogéneos en tiempo y espacio, de modo que su fiabilidad no sea disminuida por discontinuidades en su medición. Si bien la OMM y los reglamentos nacionales de cada país suelen ser claros a la hora de establecer pautas para la correcta recolección de datos climáticos (WMO,

2008), existen casos (sobre todo en países en vías de desarrollo) donde dichas normativas no se cumplen (UNFCCC, 2014). Ente los errores más comunes se destacan la falta de homogeneidad espacial en la distribución geográfica de las estaciones y la no continuidad temporal de los registros. Esto último genera vacíos de información, debido principalmente a omisión de lectura por parte del operario (para el caso de estaciones analógicas), fallas de sistema en instrumentos digitales o desmantelamiento temporal, parcial o total de la estación.

La OMM (WMO, 2017) establece que las series temporales climáticas deben tener un mínimo de 30 años de registros, los cuales inicialmente deben

garantizar su continuidad temporal sin vacíos de información. Dado lo anterior y en vista de los problemas que surgen en los vastos años de registros de una estación, es posible que las series de tiempo presenten datos faltantes, los cuales deben ser calculados o rellenados para los posteriores análisis que se realicen sobre las series consideradas.

Existen varias metodologías empleadas para hallar dichos vacíos, los cuales generalmente se apoyan en estaciones vecinas, como los métodos de proporción o razón normal, interpolación lineal, promedio aritmético, distancia inversa, entre otros (Guevara, 2003; Yozgatligil *et al.*, 2013). Adicionalmente, otros autores hacen uso de otros tipos de metodologías o modelos, las cuales de una u otra forma necesitan ser construidas o validadas a través de estaciones adyacentes. Dichos métodos utilizan información satelital (Li *et al.*, 2017; Xu *et al.*, 2018), análisis de redes neuronales (Yozgatligil *et al.*, 2013), interpolación de Kriging (Shtiliyaynova *et al.*, 2017), uso de datos de reanálisis (Xu *et al.*, 2018b), modelos climáticos (Esquivel *et al.*, 2015) o algoritmos especializados, como Niching (Jiménez *et al.*, 2014).

Otros autores, como Alfaro y Soley (2009), aplican un modelo autoregresivo a la serie temporal, para hallar datos faltantes cuando no existen estaciones cercanas. Sin embargo, los autores no presentan la precisión del método y solo exponen en forma general las características del mismo. Adicionalmente, Duque (1980) presenta una metodología denominada “razón normal sin estaciones adyacentes”, la cual permite calcular datos faltantes de una estación, basada en la propia serie temporal a través de la resolución de ecuaciones para cada dato. Esta metodología fue originalmente diseñada para hallar datos de precipitación, por lo que su naturaleza no permite estimar vacíos de información de variables no acumulativas (como la temperatura).

En cierto modo, las metodologías que usan información de estaciones vecinas, pueden compensar los datos faltantes de la estación problema, basándose en los principios de continuidad y autocorrelación espacial de las variables climáticas (Tobler, 2004). Lo anterior tiene validez si las normas de la OMM fuesen cumplidas eficientemente, de modo que el territorio esté homogéneamente

ocupado por estaciones meteorológicas y que además éstas posean los registros completos.

En referencia a lo expuesto, pueden presentarse casos en los cuales se tenga una estación con datos faltantes que no posea estaciones vecinas, o en caso de tenerlas, estas no presenten registros en el momento cronológico del vacío de información a completar. Por tanto, el usuario debe recurrir a alguna metodología (basada generalmente en estaciones vecinas más alejadas), que le permitan recuperar los datos faltantes con la mayor fiabilidad posible. Por lo antes expuesto, el objetivo de la investigación se basa en proponer un método de trabajo, cuyo fin sea completar series de tiempo mensuales de temperatura en aquellas localidades donde no existan estaciones cercanas, de modo que resulte de fácil aplicación y elevada precisión en sus cálculos.

METODOLOGÍA

La estación base para la aplicación de la metodología se localiza en la Provincia de San Juan, centro-occidente de la República Argentina. Dicha estación, denominada San Juan Aero (serial 87311), pertenece al Servicio Meteorológico Nacional (SMN, 2019) y posee 46 años de registros mensuales, los cuales fueron analizados con diversos procedimientos para calcular y validar la metodología propuesta.

La estación base está cercana a dos estaciones, las cuales sirvieron para comprobar una de las fases metodológicas subsiguientes. Dichas estaciones adyacentes son: Km. 47,3 (serial 1208) e INTA-San Juan (serial NH0445), las cuales son administradas por la Secretaría de Infraestructura y Política Hídrica de la Nación (SIPHN, 2019) y por el Instituto Nacional de Tecnología Agropecuaria (INTA, 2019), respectivamente. Estas estaciones se ubican al occidente de la República Argentina, específicamente al sur de la provincia de San Juan, que junto a Mendoza y San Luis, conforman la región argentina de Cuyo. Esta zona, al este (piedemonte) de la cordillera de Los Andes, se caracteriza por presentar un clima árido, con precipitaciones medias inferiores a los 200 mm

anuales, temperaturas medias entre los 17 y 19 °C (Capitanelli, 1967; Albiol, 2014; Albiol, 2018) y altitudes entre 500 y 700 msnm.

Las estaciones San Juan Aero e INTA-San Juan, se ubican sobre la planicie donde se asienta la ciudad y oasis de San Juan (principal polo residencial y económico de la provincia). La tercera estación, Km. 47,3, se localiza hacia el piedemonte occidental, a las márgenes del río San Juan. En la Figura 1 se muestra la amplia región, donde la temperatura se registra en tres estaciones meteorológicas, cuyas distancias entre ellas superan incluso los 50 km en algunos tramos (Tabla 1).

La metodología se centró en cuatro fases fundamentales: a) aplicación de la propuesta metodológica (y su contraste de precisión mediante índices)

para la estación San Juan Aero (estación base), con diferente número (N) de datos respecto al total original; b) similar procedimiento a lo expresado en (a) pero considerando porcentajes diferentes de datos faltantes sobre San Juan Aero; c) contraste entre la propuesta dada y el cálculo de datos faltantes basados en otras metodologías en función de las estaciones vecinas Km. 47,3 e INTA-San Juan y d) aplicación de la propuesta sobre una estación fuera del área de estudio y con condiciones climáticas y de entorno diferentes.

La propuesta metodológica (T_i), hace alusión a la estimación de datos faltantes de temperatura mensual, la cual está en función del cociente entre el producto del promedio de temperaturas anuales anterior y posterior al año donde se encuentra el mes faltante (TA_{i-1} ; TA_{i+1}), y la temperatura mensual media correspondiente al mes problema (TMM_i), sobre dos veces la temperatura anual media de la serie de datos (TAM) (ecuación 1).

Tabla 1. Distancias entre las estaciones meteorológicas consideradas.

Estaciones	San Juan Aero	Km. 47,3	INTA-San Juan
San Juan Aero	-	58,15 km	24,89 km
Km. 47,3	58,15 km	-	70,97 km
INTA-San Juan	24,89 km	70,97 km	-

Fuente: elaboración propia.

$$T_i = \frac{(TA_{i-1} + TA_{i+1}) * TMM_i}{2 * TAM} \quad (1)$$

La limitación de esta metodología radica en que el vacío de información no puede ser superior

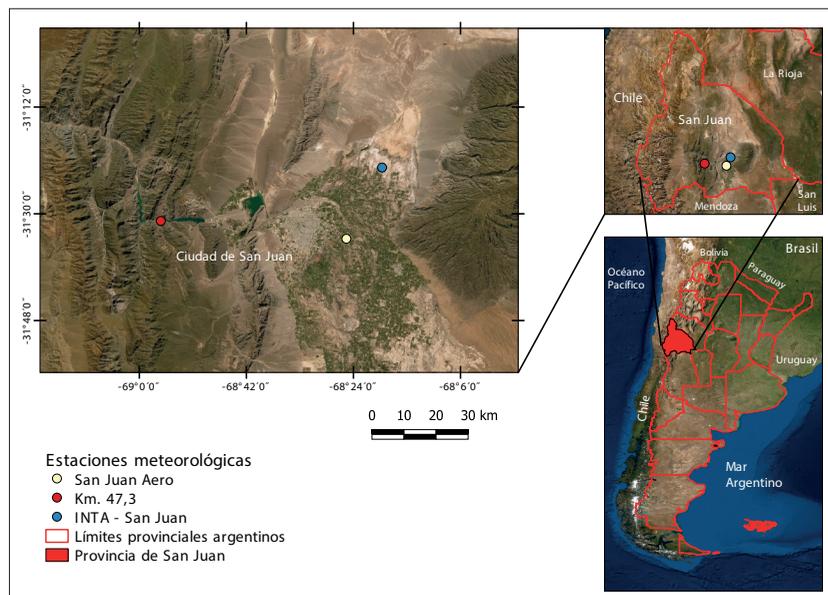


Figura 1. Ubicación y distribución geográfica de las estaciones meteorológicas. Fuente: elaboración propia

a un año continuo o igual o mayor a dos intervalos vacíos de información del mismo mes en años consecutivos, ya que la variabilidad de los mismos (y por ende su fiabilidad), se vería comprometida.

Las metodologías de contraste consideradas para la fase c) mencionada con anterioridad, son: promedio aritmético (PA), proporción normal (PN) y ponderación de distancia inversa (PDI). El promedio aritmético (Shabalala *et al.*, 2019) se basa en calcular un mes faltante (PA_x) en función de la media térmica (T_x) entre estaciones vecinas para el mismo período considerado (ecuación 2).

$$PA_x = \frac{\sum_{i=1}^n T_x}{N} \quad (2)$$

De acuerdo a Shabalala *et al.*, (2019), mediante el método (ecuación 3) de proporción normal (PN), se estiman datos faltantes a través del cociente ponderado entre la media anual de la estación problema (TM_x) y de la estación piloto (TM_i) respectivamente. Dicho cociente, luego es multiplicado en la estación vecina por el valor que cronológicamente concuerde con el faltante a hallar (T_x).

$$PN_x = \frac{1}{N} \sum_{i=1}^n T_x \frac{TM_x}{TM_i} \quad (3)$$

Por último (ecuación 4), el método de ponderación de distancia inversa (PDI_x) asume el peso de las distancias (d) entre las estaciones vecinas versus el valor que coincida con el faltante a calcular (T_x) en las estaciones vecinas. Shabalala *et al.*, (2019) expresa esta ecuación como:

$$PDI_x = \frac{\sum_{i=1}^n T_x * (d^2)^{-1}}{\sum_{i=1}^n (d^2)^{-1}} \quad (4)$$

Las pruebas realizadas para verificar la fiabilidad entre los datos calculados y los medidos, son: índice de similitud estructural modificado mSSIM (Mo *et al.*, 2013; Mo *et al.*, 2014), correlación de Pearson r (Tetzner *et al.*, 2019), coeficiente de determinación r^2 (Čadro *et al.*, 2017), error cuadrático medio

RMSE (Bright y Gueymard, 2019) y su versión normalizada NRMSE (Castañeda-Gonzalez *et al.*, 2019), error medio absoluto MAE (El Kasri *et al.*, 2019), eficiencia de Nash-Sutcliffe NSE (Rivera *et al.*, 2018), sesgo ponderado PB (Rivera *et al.*, 2018) e índice refinado de concordancia RIA (Willmott *et al.*, 2012). Las siglas anteriores provienen de sus versiones originales en idioma inglés (modified structural similarity index, root mean squared error, normalized root mean squared error, mean absolute error, Nash-Sutcliffe efficiency, percent bias y refined index of agreement, respectivamente).

El método modificado de similitud estructural o mSSIM (ecuación 5), representa una medida de concordancia entre dos vectores de datos (en este caso, x: observado e y: calculado), el cual utiliza el producto de tres subíndices como la media, varianza y correlación de Pearson. De esta forma, resulta un índice robusto con la potencialidad para indicar similitud en cuanto a patrones o estructura y a cercanía entre dos series de datos. Su rango de resultados se ubica entre -1 y 1, siendo 1 el valor óptimo de concordancia absoluta (Mo *et al.*, 2013; Mo *et al.*, 2014).

$$mSSIM = \left(\frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \right) \left(\frac{2\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2} \right) \left(\frac{\sigma_{xy}}{\sigma_x\sigma_y} \right) \quad (5)$$

Dónde: $\bar{x} = \bar{x} - \Psi_{xy}$; $\bar{y} = \bar{y} - \Psi_{xy}$; $\Psi_{xy} = \min(x_i, y_i \ i = 1, 2, \dots, N)$.

Su significancia se evaluó mediante un test t de Student al 95 % de confianza. El resto de estimadores de eficiencia (ecuaciones 6 a 13), se concentran en la Tabla 2.

Fase a)

Usando N= 46 años de datos mensuales originales de la estación San Juan Aero, se aplicó la metodología propuesta en la ecuación 1. Para ello, se dispuso que hipotéticamente cada valor original de la estación se tomara individualmente como faltante, calculando entonces para cada dato, una temperatura mensual media (TMM_i) y una temperatura media anual (TAM). Luego de obtener las dos series de datos, observados y calculados, se

Tabla 2. Estimadores de eficiencia de metodologías entre datos calculados (C) y observados (O).

Estimador	Ecuación del estimador	Ecuación
r	$r = \frac{\sum(O - \bar{O})(C - \bar{C})}{\sqrt{(\sum(O - \bar{O})^2)\sqrt{(\sum(C - \bar{C})^2)}}}$	(6)
r ²	$r^2 = (r)^2$	(7)
RMSE	$RMSE = \sqrt{\frac{1}{N} \sum (C - O)^2}$	(8)
NRMSE (%)*	$NRMSE = \frac{RMSE}{\bar{O}} \times 100$	(9)
MAE	$MAE = \frac{1}{N} \sum C - O $	(10)
NSE	$NSE = 1 - \frac{\sum(C - O)^2}{\sum O}$	(11)
PB	$PB = \frac{\sum(C - O)}{\sum O}$	(12)
RIA	$RIA = 1 - \frac{\sum C - O }{2 \sum O - \bar{O} }, \text{ cuando } \sum C - O \leq 2 \sum O - \bar{O} $ $RIA = 1 - \frac{2 \sum O - \bar{O} }{\sum C - O }, \text{ cuando } \sum C - O > 2 \sum O - \bar{O} $	(13)

* Como modificación a la ecuación presentada por Castañeda-Gonzalez et al., (2019), y para lograr mayor contraste numérico, el valor del NRMSE se multiplica por 100 para obtener su equivalente en porcentaje (%).

Fuente de ecuaciones, por orden de aparición: Tetzner et al., (2019); Čadro et al., (2017); Bright y Gueymard, (2019); Castañeda-Gonzalez et al., (2019); El Kasri et al., (2019); Rivera et al., (2018); Rivera et al., (2018); Willmott et al., (2012).

contrastó su fiabilidad de acuerdo a las ecuaciones 5 a 13, verificando luego la relación existente entre ellos (ajuste por r²). Posteriormente, se repite el mismo procedimiento principal de la fase a), pero con diferentes n originales para San Juan Aero, los cuales se definieron como: N= 30, N= 20 y N=

10. Para reducir la longitud original de datos, se comenzó desde el final de la serie (año 2013) hacia el pasado 30, 20 y 10 años respectivamente.

Fase b)

Teniendo N= 46 años de datos mensuales para San Juan Aero, se realizaron tres pruebas, eliminando al azar para cada una, el 5, 10 y 15 % de la totalidad de 552 meses originales. Para cada prueba, dichos datos suprimidos fueron estimados mediante la propuesta (ecuación 1), prosiguiendo a cada una el mismo procedimiento de fiabilidad metodológica de la fase a). En este sentido, para cada prueba (con 5, 10 y 15 % de datos faltantes), se contrastaron los datos observados versus los calculados, siendo posteriormente analizados mediante los índices de precisión mencionados anteriormente.

Fase c)

En esta tercera etapa, se contrastó la eficiencia del método propuesto T_i (ecuación 1) versus tres metodologías adicionales (PA, PM y PDI, ecuaciones 2, 3 y 4 respectivamente), que se basan en el cálculo de datos faltantes en función de estaciones adyacentes. Si bien podrían ser necesarias al menos tres estaciones vecinas, y que la estación problema se encuentre equidistante entre las demás (Guevara, 2003), la realidad demuestra que aunque esta condición no siempre se cumpla, pueden utilizarse las estaciones disponibles y con la máxima longitud de datos posible. Para esta investigación, las estaciones cercanas Km. 47,3 e INTA-San Juan poseen originalmente una longitud de 16 y 45 años respectivamente. Sin embargo, dado que la serie de Km. 47,3 se encontraba con varios datos faltantes, en función de esta estación se optó por definir el período cronológico para esta fase, la cual se extendió para el resto de estaciones (incluido San Juan Aero) entre los años 1996 y 2001 (72 meses continuos).

Similar a lo expresado en la fase a) anterior, se procedió al cálculo de cada dato (simulando que fuesen faltantes) de los 72 meses considerados (1996–2001) para la estación base San Juan Aero. Inicialmente se hallaron de acuerdo al método propuesto (ecuación 1), y posteriormente respecto a PA, PN y PDI (ecuaciones 2, 3 y 4, respectivamente), aplicando respecto a cada una,

los estimadores de eficiencia antes mencionados (ecuaciones 5 a 13).

Posteriormente y bajo la premisa de que la autocorrelación espacial disminuye con el aumento de la distancia (Tobler, 2004), se probó hasta qué distancia hipotética, los métodos que usan estaciones adyacentes se vuelven menos confiables respecto a la propuesta. Para ello, se recurrió a calcular el gradiente horizontal de temperatura ($^{\circ}\text{C}/\text{distancia}$) en función de los datos reales (1996 – 2001) entre las estaciones Km. 47,3 y San Juan Aero, e INTA-San Juan y San Juan Aero.

Inicialmente, se interpoló entre San Juan Aero e INTA-San Juan y se halló el valor de gradiente horizontal ($^{\circ}\text{C}/\text{distancia}$) por cada mes promedio, haciendo lo mismo después entre San Juan Aero y Km. 47,3. Habiendo calculado los gradientes horizontales, se simuló un aumento de distancia de las estaciones Km. 47,3 e INTA-San Juan respecto a la estación base San Juan Aero. Mediante

dichas distancias hipotéticas, cada mes de estas estaciones fue modificado de acuerdo al gradiente, obteniendo así nuevas series de datos. Dado que la distancia entre INTA-San Juan y San Juan Aero es originalmente 24,89 km, y entre Km. 47,3 y San Juan Aero es 58,15 km, se consideró extrapolar las estaciones cada diez kilómetros, comenzando desde 30 km y 70 km respectivamente, y de allí, cuatro estancias a 10 kilómetros de distancia cada uno (véase Figura 2).

Después de aplicar el gradiente horizontal, en función de las distancias hipotéticas a cada mes de las estaciones Km. 47,3 e INTA-San Juan, se calcularon las pruebas de confiabilidad igual que las realizadas en las fases anteriores a) y b) (ecuaciones 5 a 13).

Fase d)

Luego de las verificaciones dadas por las anteriores fases, se aplicó la metodología (T_i) sobre la estación

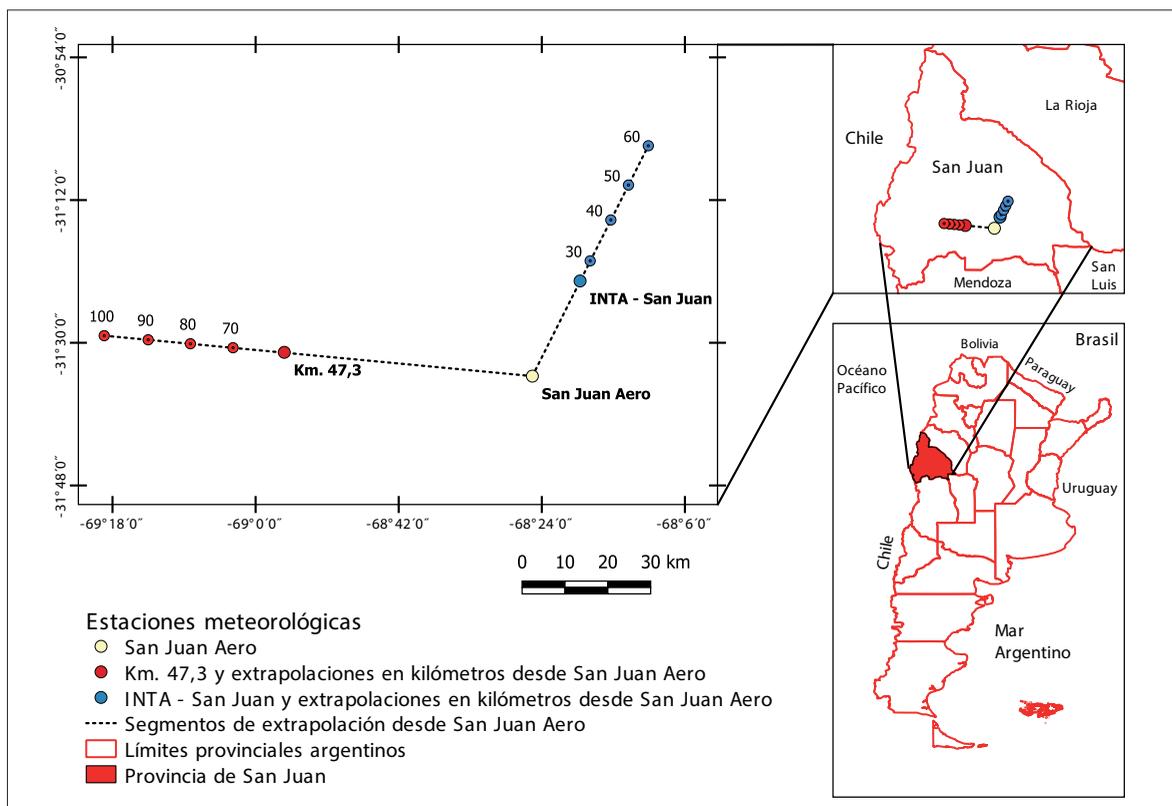


Figura 2. Distancias hipotéticas extrapoladas desde las estaciones Km. 47,3 e INTA-San Juan respecto a San Juan Aero. Fuente: elaboración propia.

Tupungato-Punta de Vacas (SIPHN, 2019), la cual se encuentra ubicada a unos 150 km al suroeste de la estación base San Juan Aero. La estación Tupungato-Punta de Vacas se asienta sobre el valle intramontano del río Mendoza, en la cordillera de los Andes, a unos 2 450 msnm, presentando un clima seco, con precipitaciones entre los 200 y 300 mm anuales y temperaturas medias en torno a los 11,1 °C para el período 2002 – 2016. La justificación de elección de esta estación como base para la fase d), radica en que la misma se encuentra bajo un régimen climático y de entorno distinto a la estación base San Juan Aero. Dado lo anterior, el objetivo radicó en verificar si la metodología puede ser aplicable a estaciones fuera de la región de estudio y bajo características físicas de entorno diferentes.

Para la estación Tupungato-Punta de Vacas, se tomó el período 2002 – 2016 como representativo por su carencia de datos faltantes. Para efectos de la prueba, se eliminaron al azar el 10 % (18 meses) del total de la serie de tiempo disponible (180 meses o 15 años de registros). Al aplicar la metodología propuesta sobre dichos faltantes hipotéticos, se contrastaron a pares sólo los valores observados y calculados, aplicándoles los mismos índices de fiabilidad presentados anteriormente en la fase a) (ecuaciones 5 a 13).

Por último, teniendo tablas resúmenes por cada prueba de las fases mencionadas con anterioridad,

se analizaron las potencialidades de la propuesta, además de sus ventajas o desventajas frente a otros métodos de cálculo, con y sin estaciones vecinas.

RESULTADOS Y DISCUSIÓN

La aplicación de las metodologías, contrastes de índices de fiabilidad y comparación de resultados hallados, permite establecer la confianza sobre los cálculos de datos faltantes hallados mediante el método propuesto. En primera instancia y dentro de la fase a), se aplicó el método T_i para obtener series calculadas para longitudes de 46, 30, 20 y 10 años (estación San Juan Aero), las cuales luego fueron contrastadas mediante los índices para evaluar su precisión (Tabla 3).

Los resultados gráficos de los mismos, teniendo en los ejes X: años, y en Y: valores de índices correspondientes, se muestran en la Figura 3.

No existen diferencias sustanciales entre los datos observados y calculados para los $N= 46$ y $N= 30$ (Figura 3). Sin embargo, se observa que a partir de $N= 30$, estos comienzan a disminuir su precisión. Si bien dichas disminuciones son comunes a todos los índices utilizados y con una proporción no tan elevada, el óptimo se encontraría en los 30 años, el cual concuerda con el período normal de

Tabla 3. Resumen de índices para evaluar precisión del método propuesto, de acuerdo a distintas longitudes de registros en la estación San Juan Aero.

Método	N (años)			
	N=46	N=30	N=20	N=10
mSSIM	0,984*	0,984*	0,983*	0,982*
RMSE	1,163	1,174	1,226	1,295
NRMSE (%)	6,370	6,380	6,630	6,960
r	0,984	0,984	0,983	0,982
r ²	0,969	0,969	0,966	0,964
MAE	0,937	0,942	0,987	1,021
NSE	0,969	0,969	0,966	0,964
PB	5,129	5,115	5,334	5,487
RIA	0,960	0,960	0,958	0,958

Los datos con asterisco (*) indican resultados significativos con $\alpha = 0,05$. Fuente: elaboración propia.

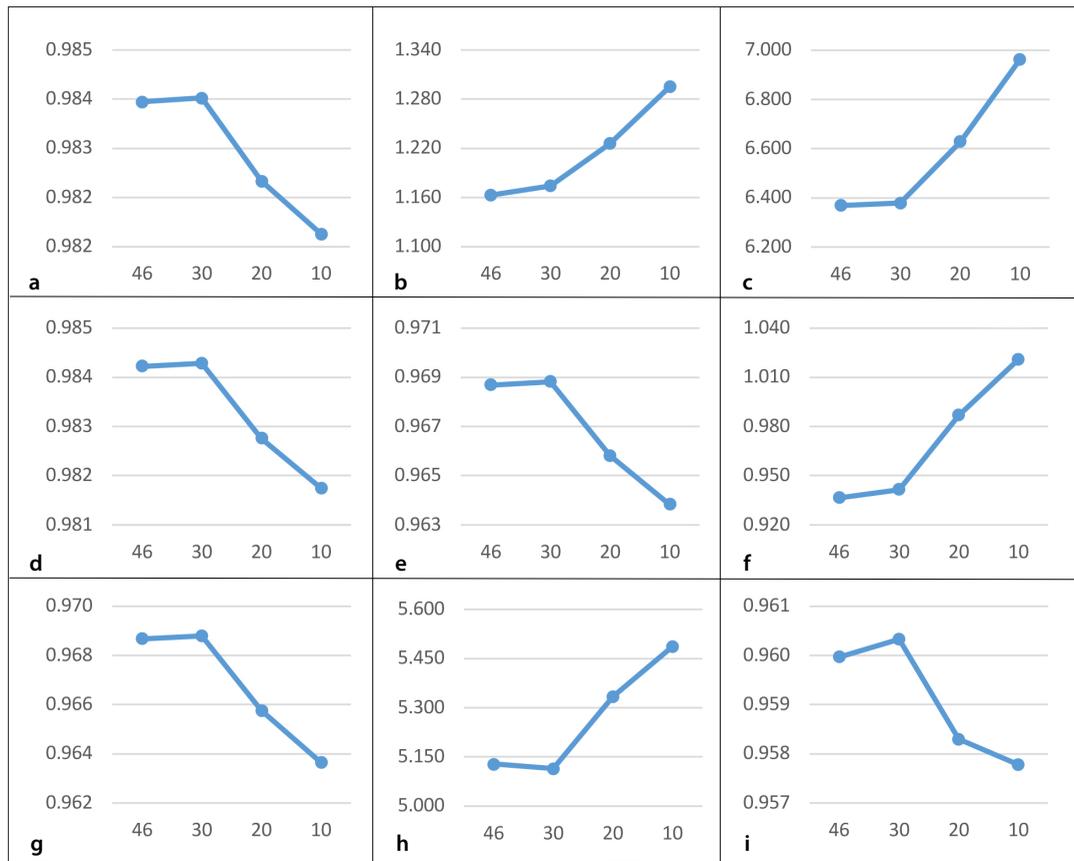


Figura 3. Índices de precisión aplicados a los datos observados versus los calculados (con magnitudes respectivas a cada índice en ejes Y) para diferentes longitudes de registros a completar (ejes X). Leyenda alfabética: a: mSSIM; b: RMSE; c: NRMSE (%); d: r; e: r²; f: MAE; g: NSE; h: PB; i: RIA. Fuente: elaboración propia.

referencia, dado por la Organización Meteorológica Mundial (WMO, 2017). En todo caso, es notable la confiabilidad del método presentado, el cual se encuentra por encima de 0,96 en los índices cuyo óptimo es 1, y con errores igualmente pequeños, respecto a RMSE, NRMSE (%), MAE y PB.

Adicionalmente, es importante acotar el valor de similitud dado por el mSSIM, que arrojó para todos los N considerados, valores significativos (Tabla 3) superiores a 0,98, mostrando la similitud entre los valores calculados con los observados (Figura 4).

En cuanto a la fase b), la eliminación del 5, 10 y 15 % de datos, para posteriormente calcularlos mediante el método propuesto, permitió observar, en términos generales, que la precisión del método

disminuye conforme aumentan los datos faltantes (eliminados). Si bien dicha disminución es proporcional con el aumento de datos calculados, la fiabilidad del método, aún con 15 % de estimaciones, es aceptable, ubicándose incluso por encima de 0,99 en aquellos cuyo óptimo es 1 (véanse Tabla 4 y Figura 5).

En la Figura 5 se observa que la confiabilidad del método propuesto, dado por los diferentes índices de precisión, es inversamente proporcional al aumento de datos faltantes. Cabe destacar que para los índices RMSE y NRMSE (%), la pendiente de la curva cambia abruptamente a partir del 10 % de datos calculados. Si bien no podría considerarse este valor como un óptimo aceptable de datos faltantes, el valor concuerda con la convencionalidad

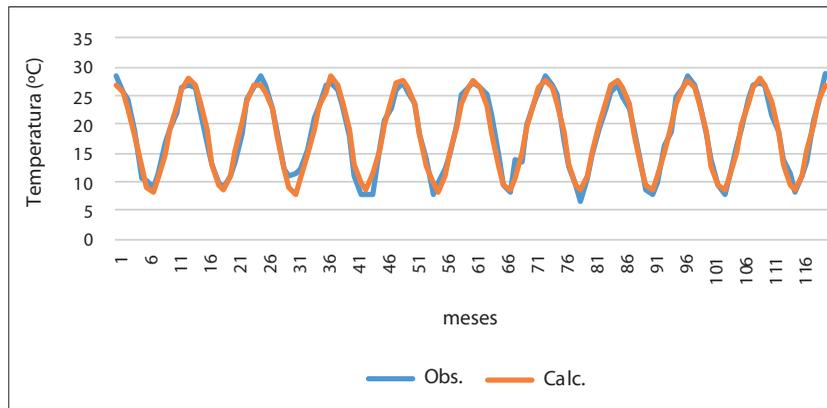


Figura 4. Valores observados (Obs.) versus calculados (Calc.) para la prueba con N= 10 años (120 meses). Fuente: elaboración propia.

Tabla 4. Resumen de índices para evaluar precisión del método propuesto, de acuerdo a distintos porcentajes de datos faltantes mensuales en los 46 años originales de registros de la estación San Juan Aero.

	Datos faltantes (%)		
	5	10	15
mSSIM	0,999	0,998	0,998
RMSE	0,372	0,385	0,440
NRMSE (%)	2,038	2,106	2,411
r	0,999	0,998	0,998
r ²	0,998	0,997	0,996
MAE	0,049	0,095	0,134
NSE	0,998	0,997	0,996
PB	0,267	0,522	0,736
RIA	0,998	0,996	0,994

Fuente: elaboración propia.

de algunos estudiosos de variables hidroclimáticas, los cuales consideran que una estación con más del 10 % de datos faltantes, se debería desestimar (Guevara, 2003).

Los resultados de la aplicación de la fase c), en la cual se contrastan los índices versus las distintas distancias tomadas en consideración, se muestran en la Figura 6.

El eje Y corresponde con los rangos de precisión de cada índice aplicado, mientras que el eje X se refiere a los grupos de metodologías contrastados (T_i , PDI, PA y PN) de forma original y mediante

su modificación por el gradiente horizontal térmico. En la Figura 6, los conjuntos agrupados de barras adicionales a la aplicación original de los métodos, hacen referencia a las distancias (dadas por el gradiente mencionado), que se utilizaron para las estaciones Km. 47,3 e INTA-San Juan. En este sentido, para estas estaciones, las distancias se toman respectivamente como 70 y 30 km, 80 y 40 km, 90 y 50 km y por último 100 y 60 km. Las barras que varían en cada gráfico, corresponden a los métodos PDI, PA y PN modificados por el gradiente horizontal (distancias) (Figura 6). En este sentido, las barras de T_i para cada estadio son invariables, dado que este método no toma en cuenta estaciones adyacentes.

En primer lugar, se aprecia cómo el T_i se encuentra en segunda posición, respecto a los demás métodos para la aplicación original (en índices con óptimo 1) y en tercer lugar en aquellos que miden errores (RMSE, NRMSE %, MAE y PB). A pesar de lo anterior, para el primer grupo de índices, el T_i registra valores superiores a 0,95 en todos los casos. Si bien representan números elevados en cuanto a precisión, las demás metodologías que utilizan estaciones adyacentes presentan una fiabilidad mayor entre sus datos calculados y los observados (aunque sea por diferencias pequeñas respecto a T_i).

Dado que el objetivo de esta fase es determinar si el T_i puede tener mejor precisión respecto a otras metodologías que utilicen estaciones adyacentes, también se observa que a medida que las distancias entre San Juan Aero y las dos estaciones adyacentes INTA-San Juan y Km. 47,3 aumenta, la precisión

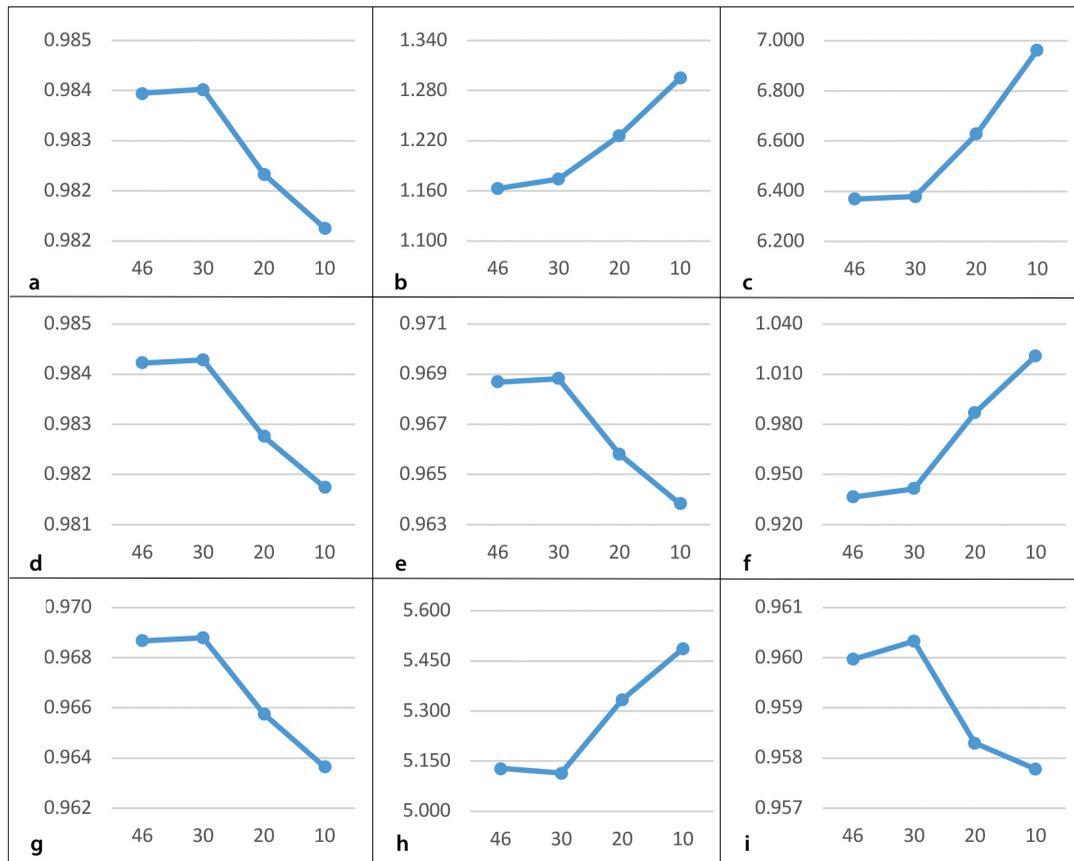


Figura 5. Índices de precisión (según magnitudes respectivas en ejes Y) aplicados entre los datos observados versus calculados (T_i), para diferentes longitudes de registros (ejes X). Leyenda alfabética: a: mSSIM; b: RMSE; c: NRMSE (%); d: r; e: r^2 ; f: MAE; g: NSE; h: PB; i: RIA. Fuente: elaboración propia.

de las metodologías disminuye respecto al método propuesto T_i (ley de Tobler). Por otra parte, T_i arroja precisiones más altas a las metodologías a partir de la distancia 70/30 (70 km desde Km. 47,3 y 30 km desde INTA-San Juan respecto a San Juan Aero) en todos los índices, a excepción de r y r^2 , en los cuales sigue por debajo de PA y PN hasta casi igualarse a estos a una distancia de 100/60 km respectivamente.

Lo anterior indica que metodologías como PA y PN (salvo PDI, que no arrojó resultados precisos), proporcionan mayor exactitud que T_i a las distancias originales consideradas; es decir, que si una estación con datos faltantes posee estaciones adyacentes, resulta más fiable utilizar métodos como PA o PN para obtener mejores resultados.

Esto tendría validez siempre y cuando se verifique que las estaciones vecinas no estén a grandes distancias (generalmente < 80 km) y que se localicen dentro del mismo régimen climático que la estación problema (Guevara, 2003).

Por el contrario, si existe una estación con datos faltantes mensuales de temperatura, y particularmente no posee estaciones vecinas, una alternativa significativa fiable sería utilizar el método propuesto T_i , el cual presenta elevada similitud respecto a los datos observados (mSSIM = 0,983), errores significativamente bajos (RMSE= 1,201 y NRMSE %= 6,557), una elevada correlación o relación de estructura ($r= 0,983$), alto coeficiente de determinación ($r^2= 0,966$), un error absoluto medio MAE = 0,979, robusta eficiencia de Nash-Sutcliffe

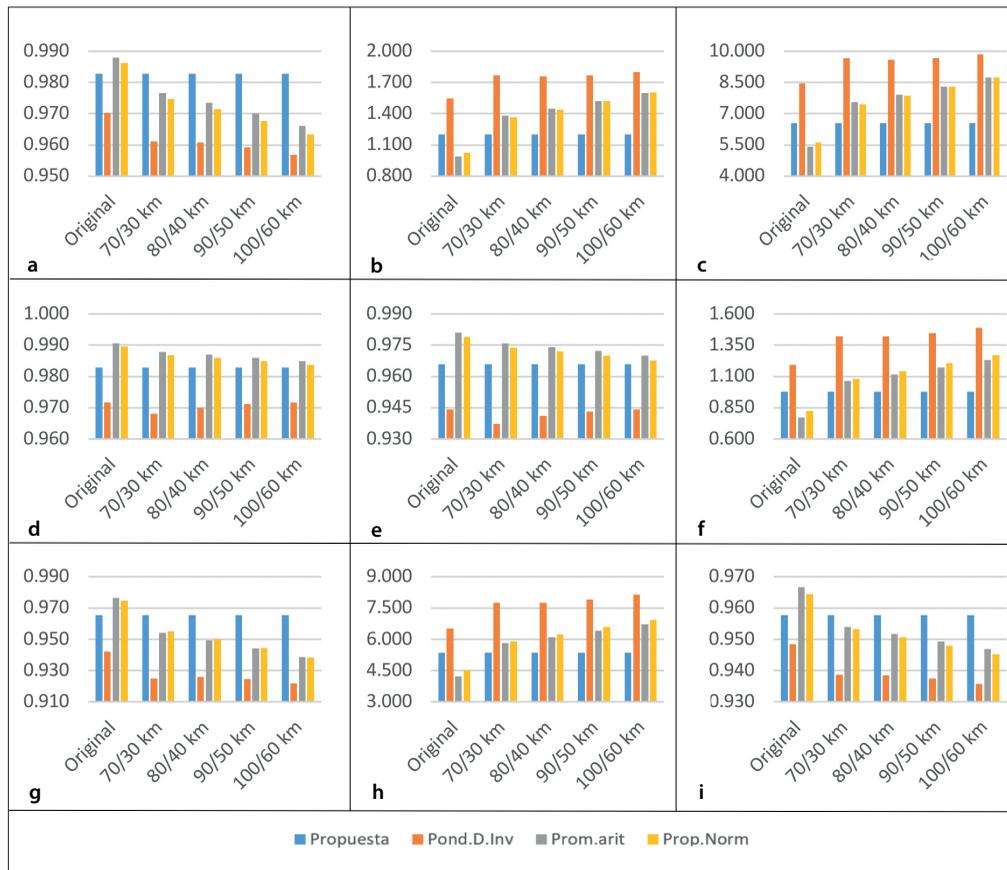


Figura 6. Eficiencia del método propuesto versus metodologías (según magnitudes respectivas en ejes Y) que usan estaciones adyacentes a diversas distancias (ejes X), respecto a la estación base San Juan Aero. Leyenda alfabética: a: mSSIM; b: RMSE; c: NRMSE (%); d: r; e: r²; f: MAE; g: NSE; h: PB; i: RIA. Fuente: elaboración propia.

(NSE= 0,965), un sesgo porcentual relativamente bajo (PB= 5,345, lo que indica un sesgo de sobreestimación de los valores calculados respecto a los observados) y un elevado valor de RIA= 0,958. A consecuencia de lo anterior, dichos índices dan una idea de la buena aproximación de los valores calculados mediante T_i , respecto a las observaciones sobre la estación San Juan Aero.

Posteriormente, como parte de los resultados de la fase d), en la Tabla 5 se presentan los valores de fiabilidad (ecuaciones 5 a 13), aplicados sobre los

cálculos realizados en la estación Tupungato-Punta de Vacas, la cual, como se mencionó con anterioridad, presenta condiciones climáticas y de entorno diferentes al sitio de emplazamiento de San Juan Aero. Adicionalmente, en la Figura 7 se observa la similitud entre los valores observados y calculados respectivamente. Cabe destacar que en esta figura, el eje X no corresponde a una sucesión cronológica de meses, sino que emula el contraste entre pares de valores observados y calculados, considerado dentro del 10 % de datos faltantes.

Tabla 5. Índices para evaluar precisión del método T_i propuesto a la estación Tupungato-Punta de Vacas.

Índice	mSSIM	r	r ²	RMSE	NRMSE%	MAE	NSE	PB	RIA
Valor	0,951	0,954	0,910	1,880	16,310	1,550	0,890	13,460	0,928

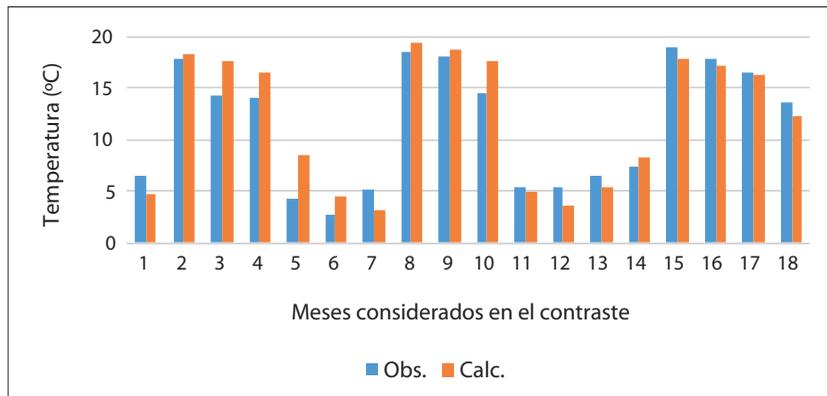


Figura 7. Valores observados (Obs.) versus calculados (Calc.) en la estación Tupungato-Punta de Vacas. Fuente: elaboración propia.

CONCLUSIONES

Es importante trabajar con series completas de datos, para lo cual es fundamental disponer de métodos de estimación eficientes para completar dichas series. Algunas metodologías, como el promedio aritmético (PA), ponderación de distancia inversa (PDI) y proporción normal (PN), representan métodos muy usados para completar series de datos, basándose en estaciones adyacentes. Sin embargo, dado que no todas las estaciones con vacíos de información pueden tener estaciones vecinas cercanas, correlacionadas o con datos en los períodos que la estación problema carece, se propuso el método T_i . Este calcula los datos faltantes mensuales de acuerdo a los registros medios anuales, medios mensuales y anuales de la propia estación.

Las pruebas realizadas para validar dicha metodología se aplicaron sobre la estación San Juan Aero, ubicada en la provincia argentina de San Juan (centro occidente del país). Dentro de las fases metodológicas empleadas, se observó que la fiabilidad del método T_i respecto a los nueve indicadores utilizados, comienza a disminuir si se considera utilizar una serie temporal < 30 años (360 meses). Por tanto, las diferencias de precisión no son tan sustanciales para series de tiempo más largas, con $N > 30$. En todo caso, aún con un N más pequeño (10 años), la precisión respecto a los índices cuyo óptimo es 1, resulta ser mayor a 0,95 en todos los casos. Dentro de esta misma fase (a), es evidente cómo el índice modificado de similitud estructural

(mSSIM), arroja resultados elevados entre los valores calculados y los observados para los diversos N considerados, estando en todos los casos por encima de 0,98. Esto indica que la similitud entre los datos calculados y observados es muy elevada, ya que la metodología los normaliza respecto a su propia serie temporal.

Por otra parte, dentro de la fase b), se demuestra cómo las precisiones disminuyen, conforme el número de datos faltantes aumenta. Tomando en cuenta el 5, 10 y 15 % de ellos respecto a los 46 años originales de la estación San Juan Aero, se recomienda desestimar aquellas estaciones que posean más del 10 % de datos faltantes respecto a su N original. Dicho valor se toma en cuenta en concordancia con otros autores como Guevara (2003) y con los resultados de esta investigación.

En relación a la fase c), se concluye que si se tiene una estación con serie de datos faltantes, y ésta tiene estaciones vecinas, resulta más confiable utilizar dichas estaciones adyacentes para hallar sus vacíos de información (bajo previa verificación de correlación entre ellas). Sin embargo, si la estación problema no posee pares vecinos, o éstos se encuentran muy alejados, bajo regímenes climáticos diferentes o no cuentan con los registros que la estación problema carece, se puede utilizar el método T_i . Su uso garantiza que los datos calculados tendrán una aceptable precisión, respecto a los datos que deberían haber sido registrados en su lugar, manteniendo valores de similitud de más de 0,98 y errores relativamente bajos en relación a otras metodologías.

Adicionalmente, durante la fase d), fue aplicada la metodología propuesta sobre el 10 % de datos faltantes hipotéticos de una estación fuera del área de estudio, la cual posee características climáticas y de entorno distintas a su par base. Lo anterior denotó una elevada similitud entre datos observados y calculados ($mSSIM > 0,95$), además de elevada precisión dada por los restantes ocho índices considerados. Esto indica que la versatilidad del método puede extrapolarse a otras estaciones con condiciones geográficas diferentes.

Es notable resaltar las limitantes del método propuesto, las cuales radican en que el mismo no puede ser aplicable si la estación base cuenta con un vacío de información superior a un año continuo, o a dos faltantes del mismo mes en años consecutivos.

Dados los resultados expuestos, y tomando en cuenta las limitantes del método y su relación frente a otras metodologías que usan estaciones adyacentes, se recomienda su utilización para calcular valores mensuales de temperatura. Junto a metodologías como la presentada por Duque (1980), en el cálculo de datos faltantes mensuales de precipitación sin estaciones adyacentes, la propuesta T_i adquiere relevancia, ya que estaciones aisladas y con datos faltantes de temperatura, podrían ser completadas y ser usadas como parte de análisis de series temporales o hasta en estudios de variabilidad y cambio climático en zonas remotas, lo que representa un alto valor para la ciencia actual.

Dentro del campo de acción de la geografía o climatología, es importante contar con herramientas (que en este caso), puedan ser válidas y veraces para estimar valores faltantes. Más allá de lo idóneo de poder analizar series temporales completas, la estimación de los vacíos en estaciones con presencia de ellos, puede ser crucial para la eficiente aplicación de análisis posteriores, cuya certidumbre tiene relación directa con los datos en los que se basan.

AGRADECIMIENTOS

Los autores agradecen al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET – Argentina) por el soporte y financiamiento de

esta investigación, la cual forma parte de la tesis doctoral del autor Bastidas.

BIBLIOGRAFÍA

- Albiol, C. (2014). Análisis de las precipitaciones de Mendoza entre 1970 y 2010: variabilidad, tendencia y comportamiento cíclico. En *X Jornadas Nacionales de Geografía Física, Malargüe* (10° ed., pp. 1–10). Recuperado de <https://redargentinadegeografiasica.files.wordpress.com/2014/04/trabajo-albiol.pdf>
- Albiol, C. (2018). Aplicación de criterios para la identificación de precipitaciones extremas en los oasis mendocinos, 1970 - 2010. *Vientos Del Norte*, 2, 10-35. Recuperado de [http://editorial.unca.edu.ar/Publicacione on line/CD VIENTOS 2017/PDF/N6 Vol II/Vientos del Norte DICIEMBRE 2018.pdf](http://editorial.unca.edu.ar/Publicacione%20on%20line/CD%20VIENTOS%202017/PDF/N6%20Vol%20II/Vientos%20del%20Norte%20DICIEMBRE%202018.pdf)
- Alfaro, E. y Soley, F. (2009). Descripción de dos métodos de rellenado de datos ausentes en series de tiempo meteorológicas. *Revista de Matemática: Teoría y Aplicaciones*, 16(1), 60. <https://doi.org/10.15517/rmta.v16i1.1419>
- Bright, J. y Gueymard, C. (2019). Climate-specific and global validation of MODIS Aqua and Terra aerosol optical depth at 452 AERONET stations. *Solar Energy*, 183(3), 594–605. <https://doi.org/10.1016/j.solener.2019.03.043>
- Čadro, S., Uzunović, M., Žurovec, J. y Žurovec, O. (2017). Validation and calibration of various reference evapotranspiration alternative methods under the climate conditions of Bosnia and Herzegovina. *International Soil and Water Conservation Research*, 5(4), 309–324. <https://doi.org/10.1016/j.iswcr.2017.07.002>
- Capitanelli, R. (1967). Boletín de Estudios Geográficos. *Climatología de Mendoza*, 14(54–57), 409. Recuperado de http://bdigital.uncu.edu.ar/objetos_digitales/11181/01-capitanelli.pdf
- Castañeda-Gonzalez, M., Poulin, A., Romero-Lopez, R., Arsenault, R., Brisette, F. y Turcotte, R. (2019). Sensitivity of seasonal flood simulations to regional climate model spatial resolution. *Climate Dynamics*, 53(7–8), 4337–4354. <https://doi.org/10.1007/s00382-019-04789-y>
- Duque, R. (1980). *Precipitación: formación, medición y análisis de datos* (CIDIAT, Ed.). Mérida, Venezuela: Universidad de Los Andes.
- El Kasri, J., Lahmili, A., Latifa, O., Bahi, L. y Soussi, H. (2019). Comparison of the relevance and the performance of filling in gaps methods in climate datasets. *Advances in Intelligent Systems and Computing*, 913, 13–21. https://doi.org/10.1007/978-3-030-11881-5_2

- Esquivel, G., Cerano, J. Sánchez, I. López, A. y Gutiérrez, O. (2015). Validación del modelo ClimGen en la estimación de variables de clima ante escenarios de datos faltantes con fines de modelación de procesos. *Tecnología y Ciencias Del Agua*, VI(4).
- Guevara, J. (2003). *Métodos de estimación y ajuste de datos climáticos* (2nd ed.; Consejo de Desarrollo Científico y Humanístico, Ed.). Caracas, Venezuela: Publicidad Gráficas León S.R.L.
- INTA. (2019). Sistema de información y gestión agrometeorológica – SIGA. Instituto Nacional de Tecnología Agropecuaria. Recuperado de <http://sig2.inta.gov.ar/#/>
- Jiménez, V., Will, A. y Lamelas, C. (2014). Imputación de datos climáticos utilizando algoritmos genéticos niching. *Acta de La XXXVII Reunión de Trabajo de La Asociación Argentina de Energías Renovables y Medio Ambiente Vol.*, 2(2006), 139–148.
- Li, H., Deng, X. y Smith, E. (2017). Missing data imputation for paired stream and air temperature sensor data. *Environmetrics*, 28(1), 1–12. <https://doi.org/10.1002/env.2426>
- Mo, R., Ye, C. y Whitfield, P. (2013). *Some Similarity Indices with Potential Meteorological Applications*. Recuperado de http://www3.telus.net/ruping/lib/reports/2013-002-CMML-TR_Mo_et al.pdf
- Mo, R., Ye, C. y Whitfield, P. (2014). Application potential of four nontraditional similarity metrics in hydrometeorology. *Journal of Hydrometeorology*, 15(5), 1862–1880. <https://doi.org/10.1175/JHM-D-13-0140.1>
- Rivera, J., Marianetti, G. y Hinrichs, S. (2018). Validation of CHIRPS precipitation dataset along the Central Andes of Argentina. *Atmospheric Research*, 213(January), 437–449. <https://doi.org/10.1016/j.atmosres.2018.06.023>
- SIPHN. (2019). Base de datos hidrológica integrada. Secretaría de Infraestructura y Política Hídrica de la Nación. Recuperado de <http://bdhi.hidricosargentina.gob.ar/>
- SMN. (2019). Estadísticas climáticas normales. Servicio Meteorológico Nacional. Recuperado de <http://www.smn.gov.ar/serviciosclimaticos>
- Shabalala, Z., Moeletsi, M. Tongwane, M. y Mazibuko, S. (2019). Evaluation of infilling methods for time series of daily temperature data: Case study of Limpopo Province, South Africa. *Climate*, 7(7). <https://doi.org/10.3390/cli7070086>
- Shtiliyanova, A., Bellocchi, G. Borrás, D. Eza, U. Martín, R. y Carrère, P. (2017). Kriging-based approach to predict missing air temperature data. *Computers and Electronics in Agriculture*, 142(March), 440–449. <https://doi.org/10.1016/j.compag.2017.09.033>
- Tetzner, D., Thomas, E. y Allen, C. (2019). A Validation of ERA5 Reanalysis Data in the Southern Antarctic Peninsula—Ellsworth Land Region, and Its Implications for Ice Core Studies. *Geosciences*, 9(7), 289. <https://doi.org/10.3390/geosciences9070289>
- Tobler, W. (2004). On the First Law of Geography : A Reply. *Annals of the Association of American Geographers*, 94(2), 304–310. Recuperado de https://www.geog.ucsb.edu/~tobler/publications/pdf_docs/The-1st-Law-A-Reply.pdf
- UNFCCC. (2014). *Measurement, Reporting And Verification For Developing Country Parties*. United Nations Framework Convention on Climate Change. (U. N. C. C. Secretariat, Ed.). Recuperado de https://unfccc.int/files/national_reports/annex_i_natcom/application/pdf/non-annex_i_mrv_handbook.pdf
- Willmott, C., Robeson, S. y Matsuura, K. (2012). Short Communication A refined index of model performance. *International Journal of Climatology*, 32(13), 2088–2094. <https://doi.org/10.1002/joc.2419>
- WMO. (2008). *Guide to meteorological instruments and methods of observation*. World Meteorological Organization. (WMO), Ed. Recuperado de <https://www.weather.gov/media/epz/mesonet/CWOP-WMO8.pdf>
- WMO. (2017). WMO Guidelines on the Calculation of Climate Normals. World Meteorological Organization. *WMO (World Meteorology Organization)*, (1203), 2929. Recuperado de https://library.wmo.int/doc_num.php?explnum_id=4166
- WMO. (2018). *Guide to Climatological Practices*. World Meteorological Organization. (2018th ed.; P. B. Chair, Ed.). Recuperado de https://library.wmo.int/doc_num.php?explnum_id=5541
- Xu, Y., Knudby, A. Shen, Y. y Liu, Y. (2018). Mapping Monthly Air Temperature in the Tibetan Plateau from MODIS Data Based on Machine Learning Methods. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(2), 345–354. <https://doi.org/10.1109/JSTARS.2017.2787191>
- Xu, W., Li, Q. Jones, P. Wang, X. Trewin, B. Yang, S. Zhu, C. Zhai, P. Wang, J. Vincent, L. Dai, A. Gao, Y. y Ding, Y. (2018b). A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dynamics*, 50(7–8), 2513–2536. <https://doi.org/10.1007/s00382-017-3755-1>
- Yozgatligil, C., Aslan, S. Iyigun, C. y Batmaz, I. (2013). Comparison of missing value imputation methods in time series: The case of Turkish meteorological data. *Theoretical and Applied Climatology*, 112(1–2), 143–167. <https://doi.org/10.1007/s00704-012-0723-x>