

## Homogenización de series climáticas y validación de datos digitales de Terraclimate en la Cuenca Media Balsas

### Homogenization of Climate Series and Validation of Terraclimate Digital Data in the Cuenca Medio Balsas (Balsas Medio Basin)

Jonathan Bautista Nájera<sup>1\*</sup>, Martín José Montero Martínez<sup>2</sup>, Miguel Ángel Gómez Albores<sup>3</sup> y Alfredo Méndez Bahena<sup>4</sup>

<sup>1</sup>Doctorado en Recursos Naturales y Ecología, Universidad Autónoma de Guerrero. Gran vía tropical No. 20, Col. Las Playas, C.P. 39390, Acapulco de Juárez, Guerrero, México.

<sup>2</sup>Instituto Mexicano de Tecnología del Agua (IMTA), Paseo Cuauhnáhuac No. 8532, Col. Progreso, C.P. 62550, Jiutepec, Morelos, México.

<sup>3</sup>Instituto Interamericano de Tecnología y Ciencias del Agua (IMTA), Carretera Toluca-Ixtlahuaca km 14.5, San Cayetano Morelos, C.P. Toluca, Estado de México, México.

<sup>4</sup>Laboratorio de Biología de la Conservación, Facultad de Ciencias Químico Biológicas, Universidad Autónoma de Guerrero (UAGro), Av. Lázaro Cárdenas s/n, Ciudad Universitaria, Chilpancingo, Guerrero, México.

#### ARTICLE INFO

##### Article history:

Recibido septiembre 29 2025

Revisado octubre 31 2025

Aceptado noviembre 13 2025

\* Corresponding author: [baunajo89@gmail.com](mailto:baunajo89@gmail.com)

Edited by Dr. Jorge Bello Martínez

##### Keywords:

*Climatol*, *Correlation*, *Precipitation*,  
*Temporal Series*, *Temperature*

*Este es un artículo en acceso abierto que se distribuye de acuerdo a los términos de la licencia*

*Creative Commons.Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)*

<https://doi.org/10.62384/fesgro.V11i1.14>

#### RESUMEN

La confiabilidad de los datos climáticos es fundamental para evaluaciones hidrológicas, pronóstico de eventos extremos y análisis de patrones. Este estudio evalúa la confiabilidad de datos de Terraclimate frente a observaciones de 83 estaciones climatológicas en la Cuenca Media Balsas. Dado que los registros in situ suelen presentar anomalías por fallas instrumentales, cambios de ubicación o ausencias, se aplicó un proceso de homogenización de datos con la herramienta Climatol para garantizar la calidad y comparabilidad de las series observadas. La evaluación se realizó mediante métricas estadísticas (Pearson, NSE, NRMSE y Pbias) en escalas espacial y temporal. Los resultados muestran una alta correlación en ambas variables. En temperatura, se observó un valor promedio de  $r=0.89$ ,  $NSE=0.74$ ,  $NRMSE$  de 11% y  $Pbias=-5\%$ . En precipitación, se registraron valores de  $r=0.93$ ,  $NSE=0.86$ ,  $NRMSE=41\%$  y  $Pbias=-5.5\%$ . Los resultados destacan la eficiencia de Climatol en la depuración de registros históricos y se confirma que TerraClimate reproduce adecuadamente la distribución climática de la cuenca, proporcionando una base confiable para evaluaciones hidrológicas digitales.

#### ABSTRACT

The reliability of climate data is critical for hydrological assessments, extreme event forecasts, and pattern analyses. This study evaluates the reliability of Terraclimate data against observations from 83 climatological stations in the Balsas Medio Basin. Since in situ records often contain anomalies due to instrumental failures, changes in station location, or missing data, a data homogenization process was applied with the Climatol tool to ensure the quality and comparability of the observed series. The evaluation was carried out using statistical metrics (Pearson, NSE, NRMSE and Pbias) on spatial and temporal scales. The results show a high correlation between the variables. For temperature, the average value was  $r = 0.89$ ,  $NSE = 0.74$ ,  $NRMSE = 11\%$ , and  $Pbias = -5\%$ . For precipitation, values of  $r = 0.93$ ,  $NSE = 0.86$ ,  $NRMSE = 41\%$ , and  $Pbias = -5.5\%$  were recorded. The results highlight Climatol's efficiency in cleaning historical records and confirm that TerraClimate adequately reproduces the basin's climatic distribution, providing a reliable basis for digital hydrological assessments.

## Introducción

La información climática constituye un elemento fundamental en estudios hidrológicos, agrícolas, energía y en la gestión de riesgos hidrometeorológicos, entre otros campos (Sham et al., 2025). La disponibilidad de datos precisos y confiables permite analizar patrones, modelar escenarios y tomar decisiones basadas en evidencia, esencial para la gestión sustentable de recursos hídricos y adaptación a la variabilidad climática (Adeyeri et al., 2022).

La información climática puede ser obtenida a través de estaciones climatológicas in situ, o mediante fuentes digitales como TerraClimate, ERA5 o CHIRPS, las cuales ofrecen información en distintas escalas espaciales y temporales. No obstante, la confiabilidad de los datos digitales debe evaluarse con información in situ para determinar su veracidad y detectar posibles sesgos o inconsistencias (Cimadevilla, 2025; Dorji et al., 2025). Una de las limitantes en este tipo de evaluaciones es la disponibilidad de series temporales climatológicas, así como las incertidumbres derivadas de fallas instrumentales, reubicación de estaciones, cambios en el uso/cobertura de suelo o periodos con ausencia de datos (Venema et al., 2012).

Existen distintas herramientas de homogenización de datos para superar estas limitantes (RHtest, HOMER, MASH, SPLIDHOM, Climatol) estas permiten rellenar valores faltantes y reducir inconsistencias en las series de datos climáticos mediante métodos de interpolación y estadística robusta (Alejo-Sanchez et al., 2025; Guijarro 2018; Mestre et al., 2011; Mastre et al. 2013; Szentimrey 2003; Wang et al., 2010; Wang y Feng 2013). Entre las cuales destaca Climatol desarrollada por Guijarro (2018) la cual integra distintos métodos estadísticos para detectar anomalías, ajustar series temporales y espaciales y generar series climáticas de mayor confiabilidad (Guijarro et al., 2023). A diferencia de otros métodos más centrados en análisis univariados o que requieren mayor intervención manual, Climatol ofrece un enfoque multivariante y automatizado que permite el tratamiento simultáneo de grandes redes de estaciones, incorporando la información espacial de estaciones vecinas para detectar y corregir inhomogeneidades mediante métodos iterativos y gráficos diagnósticos.

Esta característica es especialmente valiosa en regiones con datos incompletos o con alta variabilidad espacial, como México. Además, al estar implementado en R, Climatol facilita la reproducibilidad, transparencia y trazabilidad de los procesos de homogeneización, garantizando resultados comparables y replicables, atributos fundamentales para la investigación científica.

La evaluación de datos climáticos permite determinar la confiabilidad de datos digitales, lo cual es de gran relevancia para estudios regionales con relieves heterogéneos y recursos hídricos limitados como la Cuenca Media Balsas. Esta región es de gran importancia por su almacenamiento y distribución de recursos hídricos, así como en la generación de energía eléctrica. Además, su heterogeneidad climática puede impactar en la disponibilidad de estos recursos.

En este contexto, el objetivo de esta investigación consiste en homogenizar series climáticas de precipitación y temperaturas máxima y mínima en la Cuenca Media Balsas y contrastarlas con la información digital proporcionada por TerraClimate para el periodo 1963-2022. Los resultados de este estudio permitirán generar información confiable y robusta, que será de gran utilidad para estudios hidrológicos posteriores y para la toma de decisiones en la gestión sostenible de recursos hídricos frente a la variabilidad climática de la región.

## Materiales y Métodos

### Área de estudio

La Cuenca Media Balsas (CMB) se localiza en la porción centro de la región hidrológica del Balsas No. 18, entre los estados de Guerrero, México y Michoacán. Cubre una superficie de 32,000 km<sup>2</sup> aproximadamente y se sitúa sobre un relieve heterogéneo con elevaciones que van desde los 150 hasta los 1600 msnm. Además, presenta una alta diversidad climática desde climas semiáridos a climas fríos, con temperaturas entre 8 °C y 28 °C y precipitaciones entre 600 y 1800 mm anuales (Cuervo-Robayo et al., 2014; García 1980).

### Recopilación y pre procesamiento de datos

Se recopilaron datos diarios de precipitación y temperaturas máxima y mínima de 130 estaciones climatológicas operadas por la Comisión Nacional del Agua (CONAGUA) del periodo 1963-2022. La ubicación de estas corresponde al interior de la Cuenca Media Balsas y en de un buffer de 20 km de distancia. Los datos climáticos diarios se procesaron mediante un script desarrollado en RStudio que permitió convertir los registros originales en formatos: .est y .dat requeridos para análisis posteriores.

### Homogenización de datos climáticos

La serie de datos climáticos obtenidos de CONAGUA fueron homogeneizados y completados (relleno de datos faltantes) mediante la herramienta Climatol en RStudio. Esta herramienta consiste en un paquete especializado para el análisis de datos climatológicos que detecta y corrige valores atípicos y ajusta discontinuidades en las series temporales, lo cual permite garantizar la coherencia entre estaciones.

Además de identificar anomalías causadas por por fallas en las técnicas de medición, prácticas operativas, reubicación de estaciones o cambios en el uso y cobertura de suelo, Climatol emplea métodos de interpolación basados en la correlación de información entre estaciones cercanas para rellenar ausencia de datos. Dentro de los métodos empleados en este proceso destaca Paulhus y Kohler (1952) adaptado para rellenar precipitaciones diarias a partir de medias de valores medidos circundantes. Los datos normalizados se interpolan mediante:

$$\hat{y} = \frac{\sum_{j=1}^n w_j x_j}{\sum_{j=1}^n w_j}$$

Donde  $\hat{y}$  es un dato estimado,  $x_j$  son los datos vecinos disponibles y  $w_j$  sus pesos.

### Correlación y evaluación de datos climáticos

Los datos climáticos homogeneizados mediante Climatol ( $O_i$ ) fueron comparados con datos digitales de TerraClimate ( $S_i$ ) para determinar su verosimilitud, para el análisis solo fueron consideradas las 83 estaciones situadas al interior de la CMB. El análisis se realizó a través del Coeficiente de correlación de Pearson (Eq 1), Porcentaje de sesgo (Pbias) (Eq 2) y el coeficiente de eficiencia de Nash Sutcliffe (Eq 3) a nivel de estación, a nivel mensual y general.

$$r = \frac{\sum (O_i - \bar{O})(S_i - \bar{S})}{\sqrt{\sum (O_i - \bar{O})^2 \sum (S_i - \bar{S})^2}} \quad (\text{Eq 1})$$

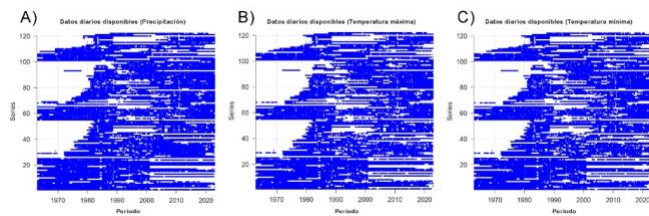
$$PBIAS = \frac{\sum_{i=1}^n (O_i - S_i)}{\sum_{i=1}^n O_i} * 100 \quad (\text{Eq 2})$$

$$NSE = 1 - \frac{\sum_{i=1}^n (O_i - S_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (\text{Eq 3})$$

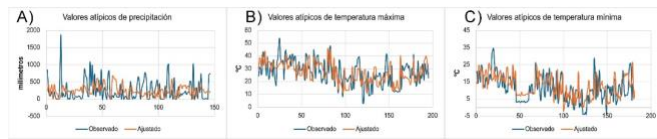
Donde  $O$ , son los valores observados,  $S$ , los simulados,  $\bar{O}$  y  $\bar{S}$ , son las medias correspondientes y  $n$  el número de datos observados.

**Resultados y Discusión**

El análisis preliminar de Climatol mostró la disponibilidad de datos diarios a través de gráficos que evidencian los periodos con información disponible y aquellos con datos faltantes en el periodo 1963-2022 (Figura 1). La identificación de los valores atípicos (outliers) en las series climáticas de cada variable analizada permitió reducir inconsistencias y realizar correcciones a nivel mensual. La Figura 2 muestra el ajuste de estos valores anómalos observados para mejorar la consistencia de las series temporales y minimizar posibles sesgos en los análisis posteriores.

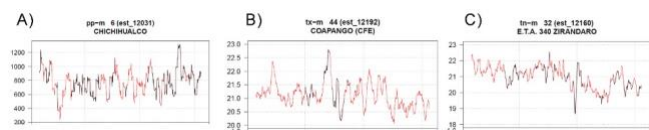


**Figura 1.** Disponibilidad de datos diarios de precipitación (A), temperatura máxima (B) y mínima (C) del periodo 1963-2022.



**Figura 2.** Corrección de datos atípicos de las series temporales de precipitación (A), temperaturas: máxima (B) y mínima (C).

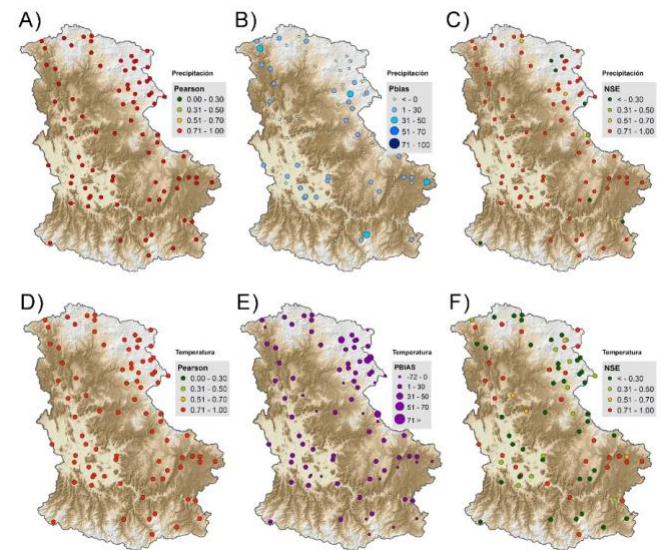
Las series mensuales homogenizadas y ajustadas se presentan en la Figura 3, estas corresponden a estaciones representativas dentro de la CMB y reflejan el proceso realizado por Climatol para garantizar series prolongadas y permitir análisis posteriores y confiables. En rojo se muestra el relleno de datos mediante los distintos métodos que incorpora Climatol y en negro los datos observados ajustados.



**Figura 3.** Homogenización de series climáticas por estación (precipitación, temperaturas máxima y mínima).

La correlación obtenida a nivel de estación mostró valores de Pearson entre 0.75 y 1.00 para precipitación (Figura 4A), lo que sugiere una consistencia favorable en la variabilidad temporal. El porcentaje de sesgo (Pbias) varió entre -474.7% y 43.2% (Figura 4B), sin embargo, el 84% de las estaciones (70 de 83) presentó sesgos entre -29% y 30%, lo que evidencia un desempeño adecuado en los datos de TerraClimate, con excepción de algunas estaciones ubicadas al noreste y sureste de la cuenca, donde se detectaron valores extremos posiblemente asociados a problemas en los datos originales. La precisión de TerraClimate evaluada mediante el coeficiente de eficiencia de Nash-Sutcliffe (NSE) varió entre -52.94 y 0.99 (Figura 4C), con el 89% de las estaciones (74 de las 83) dentro de un rango de alta confianza (0.57 al 1.00), lo que respalda la confiabilidad de TerraClimate en esta variable.

La temperatura mostró valores de correlación de Pearson entre 0.75 y 1.00 (Figura 4D) que reflejan una alta consistencia temporal. El Pbias varió entre -71.7% y 45.1% (Figura 4E), de los cuales, el 92% de las estaciones (75 de 83) presentó sesgos moderados entre -30% y 27.6% que indica un desempeño aceptable en la representación de TerraClimate, con excepción de algunas estaciones ubicadas al noreste que presentaron datos extremos posiblemente asociadas a la calidad de datos medidos. La precisión evaluada mediante NSE varió entre -66.2 y 0.98 (Figura 4F), sin embargo, solo el 48% de las estaciones (40 de 83) alcanzó rangos confiables (0.35 y 0.98). Lo que sugiere que menos del 50% de las estaciones simula de manera adecuada la temperatura.



**Figura 4.** Distribución espacial de métricas estadísticas obtenidas por estación dentro de la Cuenca Media Balsas.

La correlación de la precipitación a nivel mensual arrojó valores más altos entre enero y mayo y en noviembre, con rangos entre 0.54 y 0.77, mientras que, en los otros meses, los valores de Pearson se ubican entre 0.24 y 0.36, lo que evidencia una debilidad de TerraClimate para reproducir la variabilidad temporal en esos meses. El Pbias mensual varió entre -23.2% y 11.4%, que indica un desempeño generalmente adecuado del modelo. En cuanto al NSE, TerraClimate presentó mejor desempeño entre febrero y abril (0.48-0.52) mientras que en el resto mostró eficiencia limitada (-0.21-0.19). A diferencia de la precipitación, la temperatura presentó mejor desempeño mensual, con valores de Pearson entre 0.86 y 0.89, Pbias entre 4% y 6% y NSE entre 0.70 y 0.74, lo que indica que TerraClimate reproduce de manera más consistente la variabilidad temporal de temperatura que de precipitación.

En términos globales, la correlación general de precipitación mostró un coeficiente de Pearson de 0.93, Pbias de 5.5% y NSE de 0.86, mientras que la temperatura obtuvo 0.88 en Pearson, 11.16% en Pbias y 0.74 en NSE. Ambas variables mostraron altas consistencias en la simulación climática de toda la Cuenca durante el periodo analizado.

La homogenización de datos climáticos (precipitación, temperatura, velocidad de viento) ha demostrado mejorar la precisión de series temporales, que eleva la calidad de datos para análisis posteriores (Domonkos, 2024; Minola et al., 2024). Los resultados obtenidos en la presente investigación son consistentes con estudios previos sobre TerraClimate realizados a nivel global. Por ejemplo, (Romero-Hernández et al., 2024) compararon distintos productos incluyendo TerraClimate con datos observados entre 1981-2018, encontraron que TerraClimate reproduce de manera consistente la precipitación observada. Estudios globales también reportaron alta

eficiencia de Terraclimate en variables como temperatura, precipitación y escorrentía (Araghi et al., 2023; Wiwoho & Astuti, 2022).

A pesar de la buena resolución espacial y temporal de Terraclimate y de la alta correlación obtenida para la Cuenca Media Balsas, su desempeño puede variar según la región y variable climática analizada. Por lo que se recomienda el uso de TerraClimate como una herramienta complementaria para regiones con información observada limitada.

## Conclusiones

El presente estudio demostró la eficiencia de utilizar Climatol para el relleno, homogenización y detección de anomalías en los datos climáticos observados, lo que permitió obtener una base confiable para comparar productos climáticos digitales como Terraclimate. Este último demostró gran efectividad en el modelado de precipitación y temperatura a nivel espacial y general con ligeras limitaciones a nivel mensual, aunque su desempeño puede variar a nivel mensual y según la región o variable climática analizada.

El análisis estadístico realizado respalda su uso en estudios hidrológicos, agrícolas y gestión de recursos naturales a nivel regional. Por tal motivo, investigaciones futuras podrían incluir métodos de corrección de sesgo para mejorar la precisión de datos observados limitados. La evidencia robusta proporcionada por este estudio de la calidad de Terraclimate como fuente confiable para análisis climáticos posteriores dentro de la Cuenca Media Balsas. Gro. (Rodríguez et al., 2016).

## Expresiones de Gratitud

El autor agradece a la subcoordinación de Eventos Extremos y Cambio Climático del Instituto Mexicano de Tecnología del Agua por brindarle la oportunidad de realizar una estancia de investigación en sus instalaciones cuyo resultado resalta en esta investigación.

## Fuente de financiamiento

El presente trabajo fue realizado gracias al apoyo de beca de la Secretaría de Ciencia Humanidades, Tecnología e Innovación (SECIHTI) correspondiente a becas nacionales de estudios de posgrado (No. 823576).

## Declaración de conflicto de intereses

No existen conflictos de intereses

## Referencias

- Adeyeri, O. E., Laux, P., Ishola, K. A., Zhou, W., Balogun, I. A., Adeyewa, Z. D., & Kunstmann, H. (2022). Homogenising meteorological variables: Impact on trends and associated climate indices. *Journal of Hydrology*, 607, 127585. <https://doi.org/10.1016/j.jhydrol.2022.127585>
- Alejo-Sanchez, L. E., Márquez-Grajales, A., Salas-Martínez, F., Franco-Arcega, A., López-Morales, V., Acevedo-Sandoval, O. A., González-Ramírez, C. A., & Villegas-Vega, R. (2025). Missing data imputation of climate time series: A review. *MethodsX*, 15, 103455. <https://doi.org/10.1016/j.mex.2025.103455>
- Araghi, A., Martinez, C. J., & Adamowski, J. F. (2023). Evaluation of TerraClimate gridded data across diverse climates in Iran. *Earth Science Informatics*, 16(2), 1347–1358. <https://doi.org/10.1007/s12145-023-00967-z>
- Cimadevilla, E. (2025). Why the relational data model matters for climate data management. *Computers & Geosciences*, 201, 105931. <https://doi.org/10.1016/j.cageo.2025.105931>
- Cuervo-Robayo, A. P., Téllez-Valdés, O., Gómez-Albores, M. A., Venegas-Barrera, C. S., Manjarrez, J., & Martínez-Meyer, E. (2014). An update of high-resolution monthly climate surfaces for Mexico. *International Journal of Climatology*, 34(7), 2427–2437. <https://doi.org/10.1002/joc.3848>
- Domonkos, P. (2024). Relative Homogenization of Climatic Time Series. *Atmosphere*, 15(8), 957. <https://doi.org/10.3390/atmos15080957>
- Dorji, N., Awange, J. L., & Zerihun, A. (2025). Reliability of satellite, reanalysis and observation-based gridded temperature datasets for climate change impact studies in Bhutan. *Science of Remote Sensing*, 12, 100275. <https://doi.org/10.1016/j.srs.2025.100275>
- García, E. (1998). *Climas: Clasificación de Köppen, modificado por García (Escala 1:1 000 000)*. México: Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO)
- Guijarro, J. A. (2018). *Homogenization of climatic series with Climatol [Technical report]*. State Meteorological Agency (AEMET), Balearic Islands Office, Spain. <https://doi.org/10.13140/RG.2.2.27020.41604>
- Guijarro, J. A., López, J. A., Aguilar, E., Domonkos, P., Venema, V. K. C., Sigró, J., & Brunet, M. (2023). Homogenization of monthly series of temperature and precipitation: Benchmarking results of the MULTITEST project. *International Journal of Climatology*, 43(9), 3994–4012. <https://doi.org/10.1002/joc.8069>
- Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., ... & Stepanek, P. (2013). HOMER: A homogenization software—Methods and applications. *Idójarás*, 117(1), 47–67.
- Mestre, O., Gruber, C., Prieur, C., Caussinus, H., & Jourdain, S. (2011). SPLIDHOM: A Method for Homogenization of Daily Temperature Observations. *Journal of Applied Meteorology and Climatology*, 50(11), 2343–2358. <https://doi.org/10.1175/2011JAMC2641.1>
- Minola, L., Zhang, G., Ou, T., Kukulies, J., Curio, J., Guijarro, J. A., Deng, K., Azorin-Molina, C., Shen, C., Pezzoli, A., & Chen, D. (2024). Climatology of near-surface wind speed from observational, reanalysis and high-resolution regional climate model data over the Tibetan Plateau. *Climate Dynamics*, 62(2), 933–953. <https://doi.org/10.1007/s00382-023-06931-3>
- Romero-Hernández, C. M., Avila-Diaz, A., Quesada, B., Medeiros, F., Cerón, W. L., Guzman-Escalante, J., Ocampo-Marulanda, C., Rodrigues Torres, R., & Zuluaga, C. F. (2024). Bias-corrected high-resolution precipitation datasets assessment over a tropical mountainous region in Colombia: A case of study in Upper Cauca River Basin. *Journal of South American Earth Sciences*, 140, 104898. <https://doi.org/10.1016/j.jsames.2024.104898>
- Sham, F. A. F., El-Shafie, A., Jaafar, W. Z. W., S. A., Sherif, M., & Ahmed, A. N. (2025). Advances in AI-based rainfall forecasting: a comprehensive review of past, present, and future directions with intelligent data fusion and climate change models. *Results in Engineering*, 27, 105774. <https://doi.org/10.1016/j.rineng.2025.105774>
- Szentimrey, T. (2003, October). Multiple analysis of series for homogenization (MASH): Verification procedure for homogenized time series. In *Fourth Seminar for Homogenization and Quality Control in Climatological Databases (WMO-TD No. 1236, pp. 193–201)*. Budapest, Hungary: WMO
- Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnické, P., Viarar, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., ... Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8(1), 89–115.

<https://doi.org/10.5194/cp-8-89-2012>

- Wang, X. L., & Feng, Y. (2013). RHtestsV3: User manual [Manual]. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada. Retrieved from <http://etccdi.pacificclimate.org/software.shtml>
- Wang, X. L., Chen, H., Wu, Y., Feng, Y., & Pu, Q. (2010). New Techniques for the Detection and Adjustment of Shifts in Daily Precipitation Data Series. *Journal of Applied Meteorology and Climatology*, 49(12), 2416–2436. <https://doi.org/10.1175/2010JAMC2376.1>
- Wiwoho, B. S., & Astuti, I. S. (2022). Runoff observation in a tropical Brantas watershed as observed from long-term globally available TerraClimate data 2001–2020. *Geoenvironmental Disasters*, 9(1), 12. <https://doi.org/10.1186/s40677-022-00214-5>