DOI: 10.24850/j-tyca-15-01-02

Articles

Geostatistics to integrate gauge measurements with downscaled satellite estimates suitable for the local scale

Geoestadística para integrar mediciones de campo con estimaciones satelitales adecuados para escala local

Felipe-Omar Tapia-Silva[1], ORCID: https://orcid.org/0000-0002-5618-7594

[1]Universidad Autónoma Metropolitana-Iztapalapa, Mexico City, Mexico, otapia@xanum.uam.mx

Corresponding author: Felipe-Omar Tapia-Silva, otapia@xanum.uam.mx

Abstract

In countries such as Mexico, there is a lack of rain measurement stations. Additionally, in the Bajo Grijalva Basin, data of only three or fewer stations are integrated into satellite products of missions such as Tropical Rainfall Monitoring Mission (TRMM) and Global Precipitation Mission (GPM). Although Satellite missions enable obtaining rainfall at constant spacing (*e.g.*, 11 km for GPM), this resolution is not suitable for local

management. Integrating a larger quantity of gauge data with downscaled satellite values allows for obtaining local-scale precipitation data. In this work, Ordinary kriging (OK) was applied to downscale yearly aggregated precipitation satellite data (GPM-IMERG and TRMM: TMPA/3B43) and regression kriging (RK) to integrate them with the gauge measurements available in the basin of study. The resulting data were compared with the interpolation results of gauge measurements using OK and universal kriging (UK). Leave-one-out cross-validation (Lou-CV), principal components analysis, a correlation matrix, and a heat map with cluster analysis helped to evaluate the performance and to define similarity. An Inverse Distance Weighting (IDW) interpolation was included as a low-performance criterion in the comparison. OK performed well to downscale GPM satellite estimates. The RK integration of gauge data with downscaled GPM data got the best validation values compared to the interpolation of gauge measurements. Geostatistical methods are promising for downscaling satellite estimates and integrating them with all the available gauge data. The results indicate that the evaluation using performance metrics should be complemented with methods to define similarity among the values of the obtained spatial layers. This approach allows obtaining precipitation data useful for modeling and water management at the local level.

Keywords: Bajo Grijalva, geostatistical data downscaling, regression kriging, satellite precipitation, tropical basin.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

Resumen

En países como México hacen falta más estaciones de medición de lluvia. Además, en la cuenca Grijalva, datos de solo tres o menos estaciones se integran en productos satelitales de misiones como Tropical Rainfall Monitoring Mission (TRMM) o Global Precipitation Mission (GPM). Aunque las misiones satelitales permiten obtener estimaciones de lluvia a un espaciamiento constante (p. ej., 11 km para GPM), esta resolución no es adecuada para gestión local. La integración de una mayor cantidad de datos de pluviómetros con valores de satélite aumentados de escala puede ser útil para obtener datos de precipitación de escala local. En este trabajo se aplicó kriging ordinario (OK) a los datos satelitales de precipitación (GPM y TRMM) agregados anualmente y regresión kriging (RK) para integrar los datos resultantes con datos de todos los pluviómetros disponibles. Los resultados de esta integración se compararon con los resultados de la interpolación de datos de pluviómetros utilizando OK y kriging universal (UK). Una interpolación del inverso de la distancia al cuadrado (IDW) se consideró como criterio de bajo desempeño. Los métodos de evaluación y de definición de similaridad fueron validación cruzada (Lou-CV), análisis de componentes principales, matriz de correlación y mapa de calor con análisis de conglomerados. OK funcionó bien para desescalar las estimaciones satelitales de GPM. La integración RK de datos de pluviómetros con datos de GPM desescalados con OK obtuvo los mejores parámetros de validación en comparación con las interpolaciones de mediciones de pluviométros. Los métodos geoestadísticos son prometedores para desescalar las estimaciones satelitales e integrarlas con todos los datos disponibles de pluviómetros.

Los resultados indican que la evaluación usando parámetros para evaluar la efectividad de la interpolación usando datos medidos debe complementarse con métodos para definir similaridad entre las capas espaciales obtenidas. Este enfoque permite obtener datos de precipitación útiles para modelado y manejo del agua a nivel local.

Palabras clave: Bajo Grijalva, cuenca tropical, desescalamiento geoestadístico de datos, precipitación satelital, regresión kriging.

Received: 28/02/2022

Accepted: 06/06/2022

Published online: 15/07/2022

Introduction

Knowledge of where, when, and how much rain falls is essential for scientific research and societal applications (Skofronick-Jackson *et al*., 2018). A better understanding of the spatial and temporal precipitation (PP) patterns is still necessary to quantify the risks and design suitable mitigation measures in the context of climate change (Agou, Varouchakis, & Hristopulos, 2019). As Smalley and L'Ecuyer (2015) point out, practical decision-making for hydrologists, infrastructure, and land use, under forecasts of increasing or decreasing PP volume, can only be made with

fine-scale, detailed knowledge of the volume, and spatial distribution of PP.

According to New, Todd, Hulme, and Jones (2001), gauges that measure the PP at a single point remain the most common approach to ground-based measurement and are the ultimate reference and the only measurement method available in many regions of the world. However, the direction and magnitude of climatic trends cannot be reliably inferred from single-site records, even over relatively homogeneous terrain (Pielke *et al*., 2000). Achieving good locations for stations for data collection is difficult (WMO, 2008). Synoptic observations should be representative of an area up to 100 km around the station, but for small-scale or local applications, the considered area may have dimensions of 10 km or less (WMO, 2008).

In countries like Mexico, particularly in tropical regions as the study area, there is a lack of rain measurement stations. Tapia-Silva, Silván-Cárdenas, and Rosales-Arriaga (2013) reported that the National Meteorological Service (SMN) stations included approximately 3 300 observation sites. Assuming that each site was representative of an area of 100 km^2, as defined by WMO (2008) for flat areas, 330 000 km^2 would be covered, which was only 17 % of the territorial extension of Mexico.

PP estimates by satellite media have been investigated intensely since the 1970s (New *et al*., 2001). One of the emblematic missions has been the Tropical Rainfall Monitoring Mission (TRMM) (Kummerow, Barnes, Kozu, Shiue, & Simpson, 1998; Kummerow *et al*., 2000; Huffman *et al*., 2007). TRMM had a satellite-borne active microwave system, as well as a passive microwave radiometer. Global Precipitation Mission

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

(GPM) (Kidd *et al*., 2020) is a joint US and Japan mission launched in 2014 that follows, extends, and enhances the legacy of the TRMM (Kidd *et al*., 2020). The Integrated multi-satellite retrievals for GPM (IMERG), released in 2014, uses inter-calibrated estimates from the international constellation of PP-relevant satellites and other data, including monthly surface PP gauge data, to obtain half-hour, 0.1° x 0.1° gridded datasets (Kidd *et al*., 2020, p. 343).

According to GPCC (2012), data from very few gauge stations (three ay the most) have been integrated into TRMM and GPM for the basin's region. TRMM included rain products with a resolution of approximately 30 km^2 and IMERG (GPM) of around 11 km^2. As suggested by Smalley and L'Ecuyer (2015), knowledge about the volume and the spatial distribution of PP at the fine scale cannot be represented using this resolution. Additionally, rainfall satellite products are not error-free (Huffman *et al*., 2020, p. 350; Anagnostou *et al*., 2010; AghaKouchak, Behrangi, Sorooshian, Hsu, & Amitai, 2011; Zulkafli *et al*., 2014, p. 515), and extended validation processes have been developed (Kidd *et al*., 2020, pp. 11-12; Anagnostou *et al*., 2020). The errors are due to the sensor frequencies and channels, the type of PP, its heterogeneity within the sensor's footprint, and the algorithm used to calculate the PP rate (Massari & Maggioni, 2020, p. 515). Regarding the estimation methods implemented in Greene and Morryssey (2000), uncertainty was associated with satellite PP estimates, stemming from unknown variations in the space and time of the physical and statistical relationships between PP and satellite-sensed radiance.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Field measurements have been frequently used to validate or compare the PP information obtained by satellites (Kidd *et al*., 2020, pp. 11-12; Laurent, Jobard, & Toma, 1998; Bowman, 2005; Bell, 2003). As described in the last paragraphs, the available gauge measurements and satellite estimates cannot capture the PP spatial variability in zones like the studied basin. That is why some integration schemes have been developed. For example, New *et al*. (2021) presented a procedure that weights the individual input components by the inverse of the random error to produce a final merged product. Wu, Zhang, Sun, Lin and He (2018) merged data from TRMM Multi-Satellite P Analysis (TMPA) 3B42 with rain gauges.

Another integration possibility is the application of geostatistics, particularly kriging (Matheron, 1963). According to Curran and Atkinson (1998), a powerful synergy between geostatistics and remote sensing has been realized since the 1980s. However, Van der Meer (2012) indicates that, although the research regarding the use of geostatistics in remote sensing is growing, this has not yet been established as a standard practice. A geostatistical integration can be implemented, given that satellite products allow obtaining rainfall estimates for a particular region at constant spacing and that point measurement stations allow obtaining the value considered accurate. Regression kriging (RK) (Hengl, Gerard, Heuvelink, & Rossiter, 2007) can predict unknown values based on field measurements using estimations from satellite images as auxiliary variables.

Kriging has been defined as the best unbiased linear estimator (Cressie, 1990; Hengl, 2009) to predict values at unmeasured locations.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

These techniques have been successfully explored to generate more representative PP spatial layers from point measurements (Smalley & L'Ecuyer, 2015; Holawe & Dutter, 1999; Goovaerts, 2000; Keblouti, Ouerdachi, & Boutaghane, 2012). However, in countries like Mexico, the kriging application presents problems in capturing the local spatial variability due to the mentioned reduced number of gauge stations. Additionally, their location is not favorable since many of them were installed in easily accessible areas without consideration of a geographic sampling design, and others are no longer functioning.

TRMM Satellite estimates have been previously downscaled and integrated with field measurements using kriging (Abdollahipour, Ahmadi, & Aminnejad, 2022). Park, Kyriakidis, and Hong (2017) used kriging to downscale monthly TRMM-3B43 data at ~25km over South Korea. After that, the authors applied methods such as kriging with external drift (KED) and simple kriging with local means to integrate gauge data with the downscaled precipitation estimates. Chen, Gao, Yiguo, and Li (2020) downscaled TRMM daily PP data covering China's Henan Province for the period 1 January 2015 to 31 December 2016 through a "temporal upscaling – spatial downscaling – temporal downscaling" strategy. After that, the authors merged the results with gauge observations in a multivariate geostatistical framework. Chen, Zhang, She and Chen (2019) used kriging and other spatial analysis methods as geographically weighted regression to downscale PP data from the TRMM-3B43V7 product and to integrate the results with rain gauge data from 2001 to 2014 on reaches of the Yangtze River in China. Cersosimo *et al*. (2018) downscaled other satellite data over south Italy as Operative Precipitation

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Estimation at Microwave Frequencies (OPEMW) and Microwave Humidity Sounder (MHS) observation using KED. Other geostatistical integration schemes of PP field measurements and satellite estimates have been reported in the literature (Wang & Lin, 2015; Lin & Wang, 2011; Verdin, Rajagopalan, Kleiber, & Funk 2015; Sivasubramaniam, Sharma, & Alfredsen, 2019; Wu *et al*., 2018; Nerini *et al*., 2015). Also, work has been done to integrate ground radar and rain gauge data using geostatistics (Dumitrescu, Brabec, & Matreata, 2020; Berndt, Rabiei, & Haberlandt, 2013; Yang & Ng, 2019).

After this review, before the present work, GPM PP estimates had not been downscaled and integrated with gauge measurements using geostatistics, and this kind of downscaling and integration had not been done and evaluated for tropical areas outside Asia and Europe. In this research, geostatistical methods were applied to obtain a suitable spatial pattern of yearly aggregated PP-values at the local (basin) scale. These kinds of PP-values are valuable indicators of processes, such as water availability (Tapia-Silva & Gómez-Reyes, 2020), desertification (Morin, Marra, & Armon, 2020), water use, and water extraction (Ruiz-Alvarez, Singh, Enciso-Medina, Ontiveros-Capurata, & Corrales-Suastegui, 2020), and water management and the identification of regions vulnerable to climate change (Rata, Douaoui, Larid, & Douaik, 2020).

Ordinary kriging (OK) was applied to downscale the satellite data (GPM and TRMM) and RK to integrate the results with gauge data. These RK integration results were compared with the interpolation results of gauge data using OK and universal kriging (UK). IDW was included as a criterion for comparison purposes since it offers limited capacities to

capture the variable's spatial pattern. Leave-one-out cross-validation (Lou-CV), principal components analysis, a correlation matrix, and a heat map with cluster analysis helped to evaluate the results and to define similar values among them.

As a main result, a geostatistical downscaling and integration approach between satellite PP-estimates from GPM and TRMM and all available gauge data was developed and evaluated. The objective of this work was to answer the following questions: Is it possible to generate, through geostatistical methods, a downscaled PP spatial layer of satellite estimates of PP which can be integrated with all available gauge data for obtaining a spatial pattern suitable for the local scale, at aggregated annual values for a particular year? How do the RK results compare with other interpolation methods such as OK, and UK?

Materials and methods

Study region

The Bajo Grijalva basin (Figure 1) is located southeast of Mexico and covers an area of 9 830 km^2 (Conagua, 2015). A 30 km buffer was generated on the basin polygon obtained from the Instituto Nacional de Estadística y Geografía (INEGI) (INEGI, 2010) to delimit the study area (Figure 1).

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

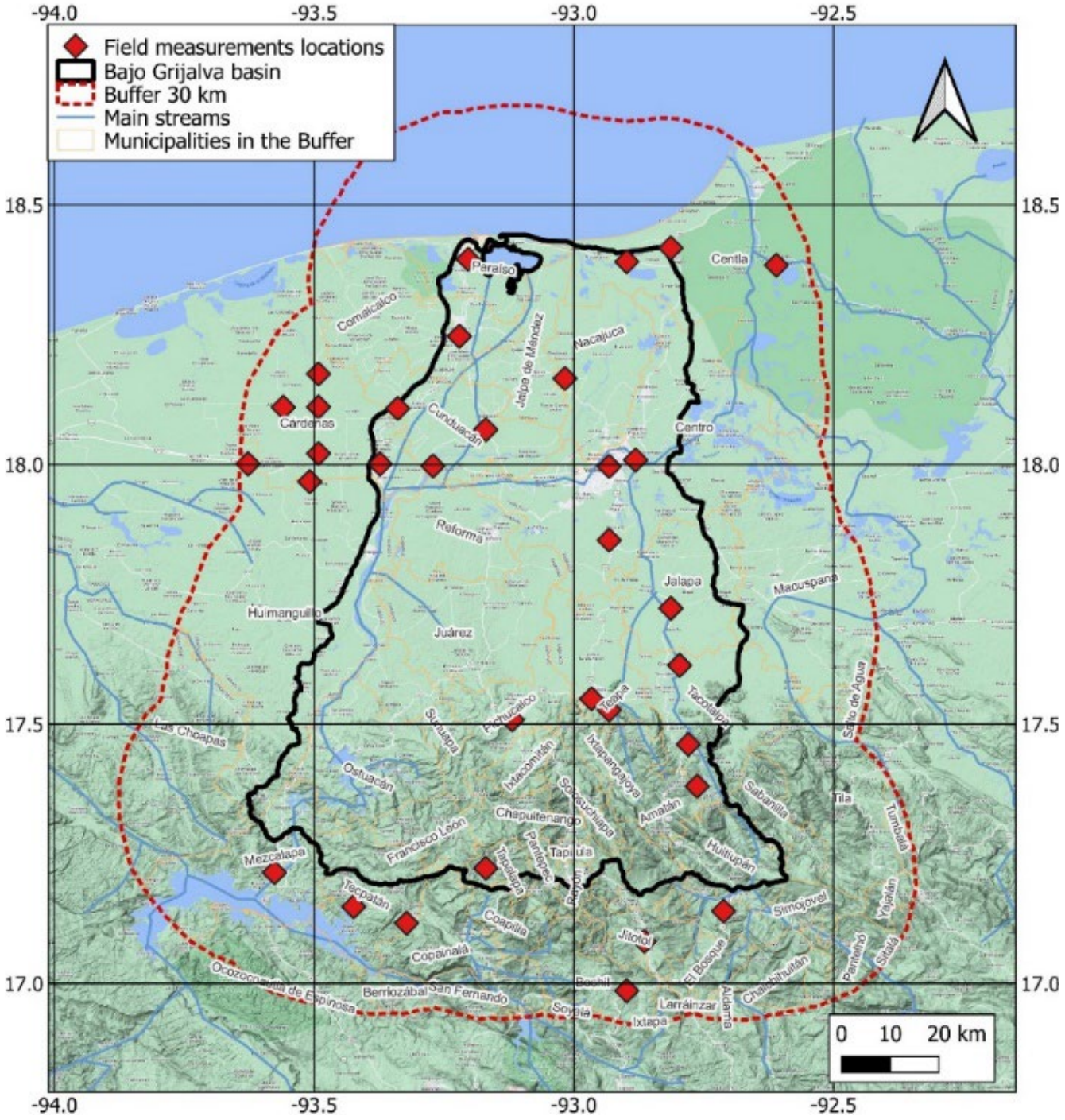

Figure 1. Bajo Grijalva Basin and its municipalities, with a 30 km buffer and the gauge measurement locations.

This basin includes 12 municipalities in Tabasco and 20 in Chiapas. The largest river system in the country, the Grijalva-Usumacinta, converges in this area. The runoff of its rivers is the largest in the Mexican Republic, with around 3 700 m^3/s on average annually (Conagua, 2015). The plains of this basin have recurrent flooding due to runoff generated by heavy rains, mainly conducted by the Sierra River (Cepal & Cenapred, 2008).

Data

The 2001 PP gauge data of the stations located in the 30 km buffer were obtained from the database provided by the SMN (Gobierno de México, 2020). As mentioned, the selected time interval (annual values) is suitable for modeling activities. It allows the estimation of water availability at the local scale in the same basin (Tapia-Silva & Gómez-Reyes, 2020). The gauge data were revised to avoid missing records and were grouped by year. On the other hand, the 2001 aggregates of TRMM (TMPA/3B43) (Kummerow *et al*., 1998; Kummerow *et al*., 2000) were downloaded from GES DISC (2011). The corresponding GPM data (IMERG Final Precipitation) (Huffman, Stocker, Bolvin, Nelkin, & Tan, 2019) were downloaded from Google Earth Engine (GEE©) using the script included in the appendix.

The statistical properties (skewness, coefficient of variation, normality fit, mean, and confidence intervals) of the used data are given

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

in Table 1. As can be observed, the parameters indicate many differences among them. For example, the mean of the gauge measurements was lower than the same value of the GPM estimates and higher than the corresponding value of the TRMM estimates. The log-transformed gauge measurements could be assumed normal, and its skewness was very close to zero.

Table 1. Statistical properties (skewness, coefficient of variation, normality fit, mean and confidence intervals) of used data.

2001-yearly aggregated PP data set	**Skewness**	**Coefficient of variation**	***P*-value of Shapiro Wilcoxon test (if >0.05 normality can be assumed)**	**Mean**	**Mean confidence intervals**	
					Left	**Right**
Gauge measurements	1.06	48.3	0.0013	1601.5	1393.3	1809.6
Log-transformed gauge measurements (used to interpolate)	0.009	6.5	0.753	3.157	3.102	3.212
TRMM satellite estimates	0.25	36.19	0.0005	508.1	478.1	538.1
GPM satellite estimates	-0.67	0.64	8.9e-11	2174.2	2140.2	2208.2

Methodology overview

Figure 2 includes a flow chart showing an overview of the developed methodology. The 2001 aggregated gauge data were interpolated using OK, UK and IDW. As a downscaling method, the values of the TRMM cells at the original resolution of approx. 30 km, were interpolated using OK. The same procedure was implemented using the GPM cell centers at their original resolution (approx. 11 km). The gauge data were interpolated with RK using the TRMM data at the original resolution (30 km) as an auxiliary variable. To improve the spatial pattern of the resulting PP layer, RK of gauge data was implemented with the TRMM OK-downscaled values, in one case, and with the GPM Ok-downscaled values in the other, as secondary information. This is the approach proposed in this work to integrate gauge data with satellite estimates of PP. In RK, the mean was estimated from the linear relationship with the auxiliary variable (TRMM or GPM OK-downscaled layers). Since only a linear relationship between the gauge measurements and satellite estimated PP-values was found and not with elevation, the latter variable was not included in the RK interpolation. All resulting layers were visually inspected to observe spatial discontinuities. Only the layer from the RK integration of gauge data and TRMM at original resolution (without downscaling) had them and was discarded.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

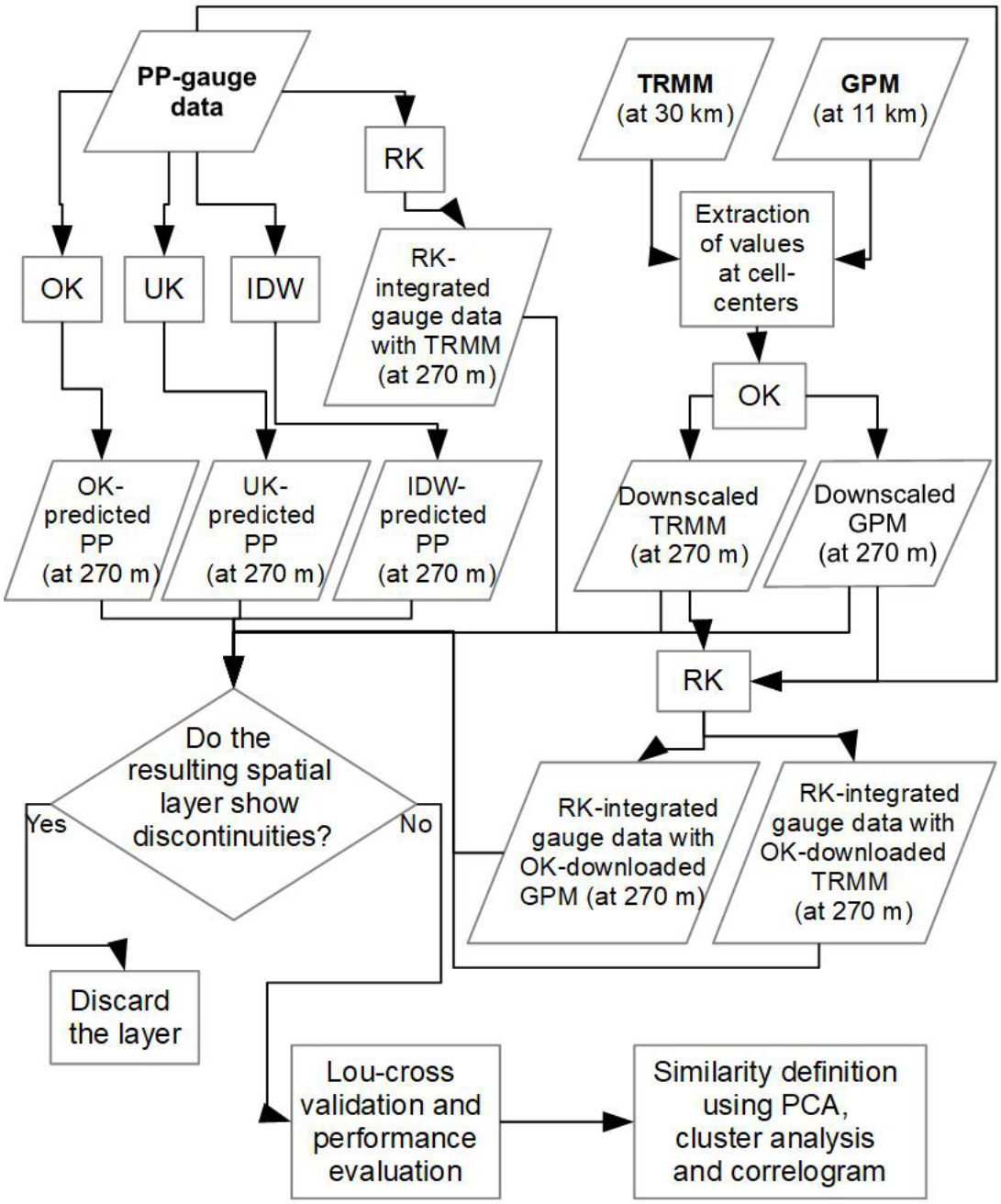

Figure 2. Methodology overview.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

As the gauge measurements were skewed (Table 1), they were log-transformed. The variogram model (s. Equation (6) in section Kriging) was fitted to the empirical values using an optimization procedure in R (R Core Team, n.d.) with the 'fit.variogram' function of the package gstat (Pebesma, 2004).

The 270 m resolution was selected because it is useful for obtaining PP spatial layers for planning and management activities, such as annual hydrological balances at the local scale of basins (*e.g.*, Tapia-Silva & Gómez-Reyes, 2020). The theory of applied kriging methods and procedures such as Lou-CV, performance evaluation, and definition of similarity of the resulting values, are described in the following paragraphs.

Kriging

According to Hengl (2009), the original idea of kriging came from the mining engineer D. G. Krige and the statistician H. S. Sichel. Matheron (Matheron, 1963; Matheron, 1965) followed the empirical work of Krige and established the Theory of Regionalized Variables, which is the base of geostatistics (Oliver & Webster, 2015). Since the mathematical formulation of kriging can be very extensive in the literature (*e. g.*, Goovaerts, 1997), a summary of the mathematical basis of the applied kriging techniques is provided in this section.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Kriging accounts for local variations in the mean by limiting the domain of stationarity to a local neighborhood, Ω, around the position, x, where the variable is to be estimated. Let $Z(x) = Y(x) + m(x)$ be a stochastic process with a variable mean that is determined by $m(x)$ and the covariance function $C(h)$. As such, $Y(x)$ is a stochastic process with a null mean. A linear estimator is a linear combination of measurements $Z(x_1)$, $Z(x_2)$,…, $Z(x_n)$ at positions x_1, x_2,…, $x_n \in \Omega$. Specifically (Goovaerts, 1997, p. 126):

$$\hat{Y}(\mathrm{x}) = \sum_{k=1}^{n} \lambda_k(x) Y(x_k) \quad (1)$$

or:

$$\hat{Z}(x) = m(x) + \sum_{k=1}^{n} \lambda_k(x)[Z(x_k) - m(x_k)] \quad (2)$$

If the mean is constant in domain Ω, then it can be eliminated from the equation above by forcing the kriging weights to sum to one, in which case, the estimator is called OK, and is expressed as (Goovaerts, 1997, p. 133):

$$\hat{Z}_{KO}(x) = \sum_{k=1}^{n} \lambda_k(x) Z(x_k) \quad (3)$$

with:

$$\sum_{k=1}^{n} \lambda_k(x) = 1 \quad (4)$$

The Lagrange multiplier method (Goovaerts, 1997, p. 133) is used to obtain the optimal weights that minimize the estimation of the error variance, which results in the following system of equations:

$$\begin{cases} \sum_{k=1}^{n} \lambda_k(x) C(x_j - x_k) + \mu(x) = C(x_j - \mathrm{x}), j =, \ldots\ldots, n \\ \sum_{k=1}^{n} \lambda_{k(x)=1} \end{cases} \quad (5)$$

where μ denotes the Lagrange multiplier. Alternatively, when considering the relation between the covariance function and the semivariogram function γ(h), i.e., C(h)=C(0)-γ(h), then the above system can be written as:

$$\begin{cases} \sum_{k=1}^{n} \lambda_k(x) \gamma(x_j - x_k) + \mu(x) = \gamma(x_j - \mathrm{x}), j =, \ldots\ldots, n \\ \sum_{k=1}^{n} \lambda_{k(x)=1} \end{cases} \quad (6)$$

OK assumes a stationary mean, that is, it is a constant of the random function *Z(x)*, of the real underlying value. However, it is often not constant throughout the entire study area. When that is the case, a non-stationary regionalized variable has two components: the drift (which is the average or the expected value of the regionalized variable, called the structured component) and the residual (which is the difference

between the values of the parameter that are considered real and the drift, called the random component) (Matheron, 1971, p. 5).

UK divides the random function into a linear combination of deterministic functions: the smooth and non-stationary trend (drift or mean) $\mu(x)\ \varepsilon\ R$, and the residual random function $Y(x): = Z\ (x)\ -\ \mu\ (x)$ (Wackernagel, 2003, p. 300). UK assumes that $\mu(x)$ is a function of the spatial location, and this can be approximated by the following model (Kumar, 2007):

$$\mu(x) = \sum_{i=1}^{n} \alpha_i f_i(\mathrm{x}) \tag{7}$$

Where α_i is the coefficient to be estimated based on the data, f_i = the basic function of drift as a function of the spatial coordinates, and n = the number of functions used in the drift model.

As with OK, the weights in UK are obtained by minimizing the variance of the prediction error subject to the unbiasedness restriction. The Lagrange multiplier is applied once again, taking into consideration the spatial autocorrelation structure to obtain the optimal weights.

RK is the Best Linear Unbiased Prediction (BLUP) model for spatial data, and all other techniques, such as OK, IDW, etc. can be seen as its special cases (Hengl, 2009, pp. 29-30; Hengl *et al*., 2007). In matrix notation, RK is commonly written as (Hengl, 2009, p. 28):

$$\hat{Z}_{RK}(x_o) = q_o^T \cdot \hat{\beta}_{GLS} + \lambda_o^T \cdot \left(Z - q \cdot \hat{\beta}_{GLS}\right) \tag{8}$$

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

$\hat{Z}_{RK}(x_o)$ is the predicted value at the location x_o, q_o is the vector of p + 1 predictors, $\hat{\beta}_{GLS}$ are regression coefficients estimated with OLS (Ordinary Least Squares) or optimally with GLS (Generalized Least Squares), and λ_o is the vector of the *n* kriging weights used to interpolate the residual. It has a prediction variance that reflects the position of new locations (extrapolation) in both geographical and feature spaces (Hengl, 2009, p. 28; Hengl *et al*., 2007):

$$\hat{\sigma}^2_{RK}(x_o) = (C_o - C_1) - c_o^T \cdot C^{-1} \cdot c_o + (q_o - q^t \cdot C^{-1} \cdot c_o)^T \cdot (q^t \cdot C^{-1} \cdot q)^{-1} \cdot (q_o - q^t \cdot C^{-1} \cdot c_o) \quad (9)$$

Where *Co* + *C1* is the sill variation and c_o is the vector of the covariances of residuals at the unvisited location.

According to Hengl (2009, p. 29), if the residuals show no spatial autocorrelation (pure nugget effect), the RK converges to pure multiple linear regression given that the covariance matrix (*C*) becomes an identity matrix. Hengl (2009, p. 29) indicated that, if the target variable shows no correlation with the auxiliary predictors, the RK model reduces to OK because the deterministic part equals the (global) mean value.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Evaluation and similarity of results

All interpolation results were evaluated using Lou-CV. The evaluation parameters were the determination and correlation coefficients (R^2 and r, respectively) and z-scores (z_i) calculated as (Bivand, Pebesma, & Gómez-Rubio, 2013, p. 225):

$$z_i = \frac{Z(x_i) - \hat{Z}_{[i]}(x_i))}{\sigma_{[i]}(x_i)} \quad (10)$$

with $Z_{[i]}(x_i)$ and $\hat{Z}_{[i]}(x_i)$ as the cross-validation prediction for x_i and $\sigma_{[i]}(x_i)$ as the corresponding kriging standard error. According to Bivand *et al*. (2013 p. 225) z_i it is a standardized residual, and, if the variogram model is correct, it should have mean and variance values close to 0 and 1.

The *Lou-CV*-residuals were analyzed calculating their mean value and the root mean square error (RMSE) was calculated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Z(x_i) - \hat{Z}(x_i))^2}{n}} \quad (11)$$

Nash-Sutcliffe Efficiency (NSE) and Ratio of Standard Deviation (RSD) were also included in the evaluations of the Lou-CV results. They were calculated as:

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Z(x_i) - \hat{Z}(x_i))^2}{\sum_{i=1}^{n}(Z(x_i) - \hat{Z}(x_i))^2} \quad (12)$$

$$RSD = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z(x_i) - \hat{Z}(x_i))^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Z(x_i) - \bar{Z})^2}} \quad (13)$$

Following Burgan and Aksoy (2022), the ranges of these parameters and their remarks are the following: -∞ ≤ NSE ≤ 1, NSE approaching 1 shows better performance; 0 ≤ RMSE ≤ ∞, RMSE has the dimension as the variable, and this parameter approaching zero shows better performance; 0 ≤ RSD ≤ ∞, RSD approaching zero shows better performance. For NSE and RSD, Burgan and Aksoy (2022) provide the following criteria of performance metrics: very good: 0.75 < NSE ≤ 1.00, 0.00 ≤ RSD ≤ 0.50; Good 0.65 < NSE ≤ 0.75, 0.50 < RSD ≤ 0.60; adequate: 0.50 < NSE ≤ 0.65 0.60 < RSD ≤ 0.70; and finally inadequate: NSE ≤ 0.50 RSD > 0.70.

Additionally, to define the similarity of the values of the obtained spatial layers and their grouping, a correlation analysis including a heatmap and a dendrogram, as well as a principal components analysis (PCA), were implemented. All validation procedures were performed in the free statistical software R (R Core Team, n.d.) using its package gstat (Pebesma, 2004).

Results

OK, UK, and IDW prediction and Lou-CV results

The resulting evaluation parameters obtained from the Lou-CV, for all interpolations are in Table 2.

Table 2. The Lou-CV resulting parameters of the interpolations of the 2001 PP yearly values.

Interpolation approach	Adj. R^2	*r*	*p*-value	Residuals mean	NSE	RSD	RMSE	Z_i-	
								mean	variance
OK of gauge measurements	0.31	0.57	9.258e-06	1.08	0.23	0.87	671.8	0.01	1.33
UK of gauge measurements	0.29	0.54	1.666e-05	1.083	0.19	0.89	690.2	0.01	1.38
IDW of gauge measurements	0.2	0.44	0.0005	111.05	0.19	0.89	688.5	na	na
OK-downscaling of TRMM	0.91	0.95	2.2e-16	0.53	0.91	0.30	135.91	-0.001	1.33
OK-downscaling of GPM	0.99	0.99	2.2e-16	-0.02	0.99	0.1	37.11	-0.020	2.98
RK-integration of gauge measurements and OK-downscaled TRMM	0.35	0.59	1.546e-06	1.07	0.29	0.83	644.5	-0.01	1.53
RK-integration of gauge measurements and OK-downscaled GPM	0.41	0.64	1.266e-07	1.06	0.40	0.76	590.8	-0.001	1.41

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

The prediction and Lou-CV results of the interpolation of gauge measurements using the OK, UK, and IDW methods are shown in Figure 3. For the OK interpolation (part a in Figure 3), a gaussian semivariogram model was fitted. The results of the OK interpolation showed a smoothed spatial pattern with the highest PP-values in the western part of the study area followed by the eastern part. The Lou-CV results from the OK prediction were the following: adjusted R^2 of 0.31, r = 0.57, and p-value = 9.26e-06. The CV-residual mean and the RMSE were 1.08 and 671.8, respectively. The parameters NSE and RSD had the values 0.23 and 0.87, respectively, indicating an inadequate performance (Burgan & Aksoy, 2022). However, the obtained Lou-CV-z_i values had a mean of 0.009 and a variance of 1.33, which, according to Bivand *et al*. (2013, p. 225) indicates a good fit for the semivariogram model. The map of z_i-values shows the locations with the lowest values (with a minimum of -4.87) located at the center and the southeast part of the area of study. The locations with the highest z_i-values (with a maximum of 2.26) were in the middle and the south of the area.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Tecnología y Ciencias del Agua
OPEN ACCESS
Check for updates

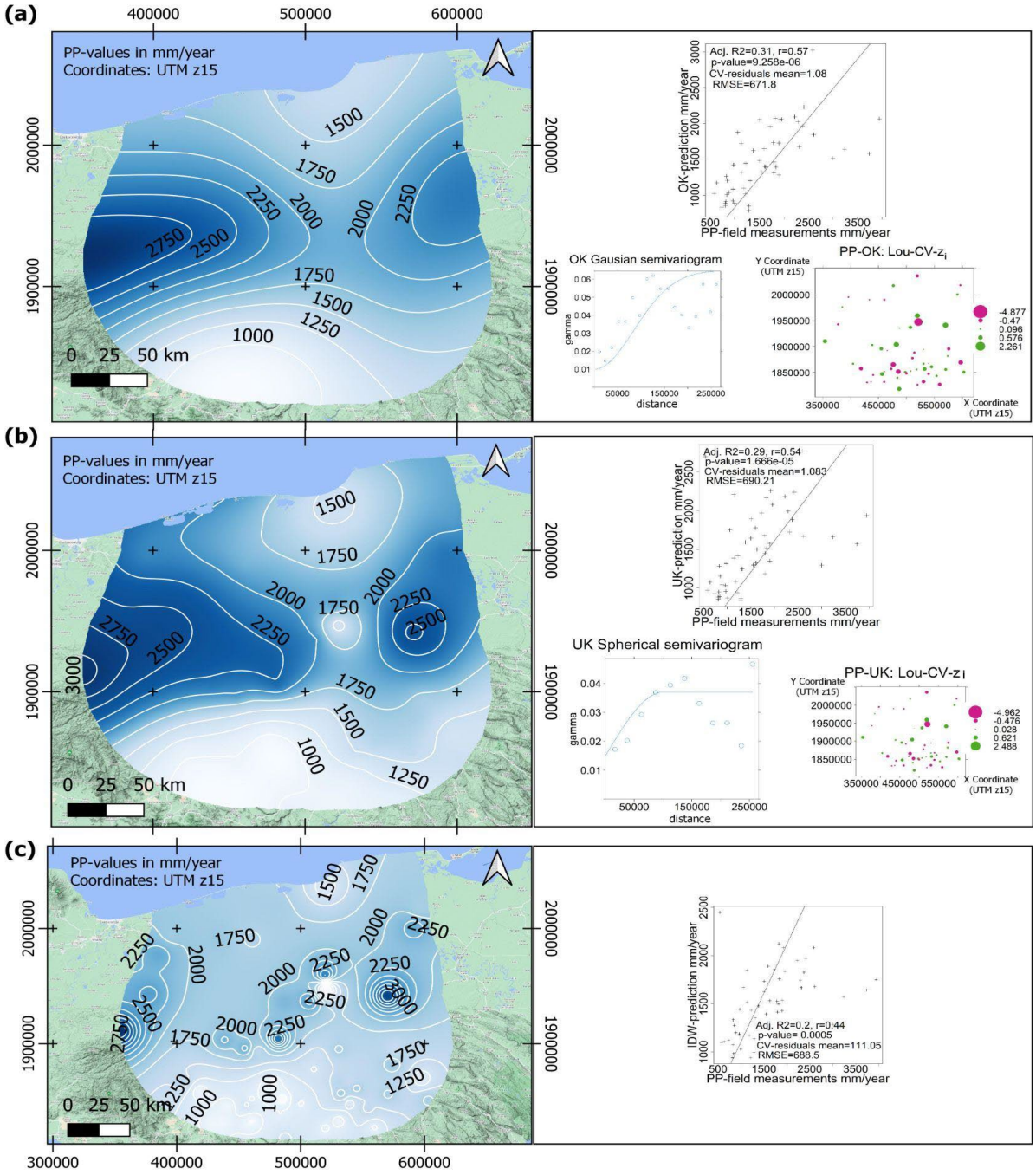

Figure 3. The prediction and leave-one-out cross validation (Lou-CV) results of the ordinary kriging (OK) (a), universal kriging (UK) (b), and inverse distance weighted interpolation (IDW) (c) interpolations of the

annual values of 2001 precipitation (PP) gauge measurements; (a) and (b) include plots of the fitted semivariogram as well as a map showing the spatial distribution of the z_i obtained from the Lou-CV; (a), (b), and (c) includes a scatterplot that compares the predicted values and the gauge measurements and shows R^2 and r, the p-value, the mean of the CV residuals, and the RMSE.

For the UK interpolations (part b in Figure 3), a spherical semivariogram model was fitted, and the gauge measurements were log-transformed. The spatial pattern of the resulting spatial layer can be defined as smooth as it resulted from the OK interpolation results. The highest predicted values were also located in the east and west parts of the study area. A difference between the OK and UK interpolation results can be observed in the middle part of the UK-prediction layer, in which a small zone with the lowest values was present. This was not observed for the OK prediction. The results of the Lou-CV for the UK prediction were the following: an adjusted R^2 of 0.29, an r of 0.54, and a p of 1.66e-05. The obtained CV-residual mean and the RMSE values were 1.08 and 690.2, respectively. The parameters NSE and RSD had values of 0.19 and 0.89, respectively, indicating an inadequate performance (Burgan & Aksoy, 2022). However, the z_i-mean was 0.011 and the z_i-variance 1.38, which indicates good fit of the semivariogram model (Bivand *et al*., 2013, p. 225). The Lou-CV parameters of the UK prediction were slightly lower than the corresponding parameters of the OK prediction.

The spatial distribution of the locations with the lowest and highest z_i-values from the UK prediction was like the OK prediction. In the case

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

of the UK prediction, the lowest z_i -value was -4.96, and the highest was 2.48. In both cases (OK and UK), the mean and variance of the z_i -values were close to 0 and 1, respectively, which, according to Bivand *et al*. (2013, p. 225), indicates a good fit of the semivariogram model with a slightly better performance of the OK interpolation.

The IDW interpolation results are shown in part c of Figure 3. For this interpolation, the effect of the measurement values from individual stations, that generate circular influence zones around them, could be observed. This spatial pattern can be considered incorrect, as the PP patterns are distributed over a regional or global context more than a local range around the gauge stations (Smalley & L'Ecuyer, 2015). As expected, the parameters of Lou-CV for this interpolation R^2=0.2 and r=0.44 (p-value=0.0005) were the lowest compared to the OK and UK results. The parameters NSE and RSD had the values 0.19 and 0.89, respectively, indicating an inadequate performance (Burgan & Aksoy, 2022).

Evaluation of OK downscaling of satellite estimates

The results of the geostatistical downscaling procedure of satellite estimates of TRMM and GPM using OK are presented in Figure 4.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

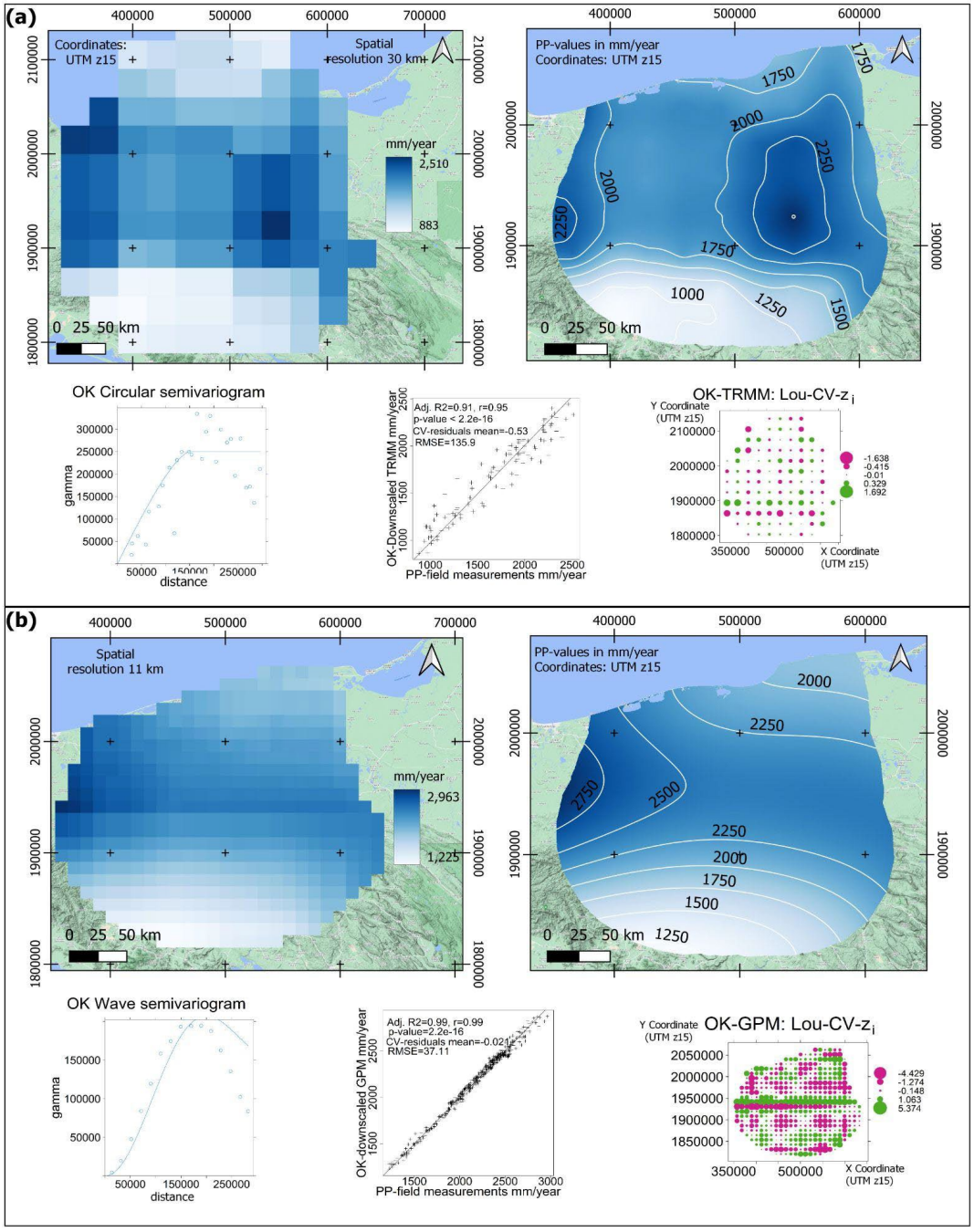

Figure 4. Prediction and Lou-CV results of the OK downscaling procedure of TRMM satellite estimates (a) and results of the OK-downscaling procedure of GPM satellite estimates (b) for 2001 annual PP

values. Each part of the figure includes maps of the data at the original resolution and maps of the downscaled values, the fitted semivariogram to these data, the results of the Lou-CV with two graphs: A scatterplot between the OK predictions and the original values (cell center values) including the values of R^2 and r, p, the mean of the CV residuals, and the RMSE; and map showing the spatial distribution of z_i-values.

The gauge measurements had values from 540 to 3 948 mm, the original TRMM estimates (part (a) to the left) from 883 to 2 510 mm, and the GPM estimates (part (b) to the left) from 1 225 to 2 963 mm (ranges of 3 408, 1 627, and 1 738 mm, respectively). The satellite estimates have the lowest ranges because they are considered spatial smoothers due to the large grid cell size as compared to point (gauge) measurements (Toté *et al*., 2015).

For the OK downscaling of TRMM (Figure 4, part a), it was not necessary to transform the data, and it was possible to fit a circular semivariogram model. The obtained spatial layer had a smoother pattern than the original TRMM layer, which is more suitable for the local scale.

Lou-CV parameters of the OK downscaling of TRMM (part a of Figure 4, below the maps) were adjusted R^2=0.91, r = 0.95, and p-value = 2.2e-16. The CV-residual mean and RMSE values were 0.53 and 135.9, respectively. The parameters NSE and RSD had the values 0.91 and 0.30, respectively, indicating very good performance (Burgan & Aksoy, 2022).

The resulting z_i-values had a mean of 0.001 and a variance of 0.43, indicating a good fit for the semivariogram model (Bivand *et al*., 2013, p.

225). The z_i-values had a minimum of -1.64 and a maximum of 1.69. The resulted parameters indicated a close fit between the predicted and the measured PP-values.

A wave semivariogram model was fitted for the OK downscaling of the GPM values (part b of Figure 4). According to the Lou-CV, an almost perfect fit between the gauge measurements and predicted values was observed (adj. R^2=0.99, r=0.99, and p-value=2.2e-16). The CV-residual mean and the RMSE values were -0.021 and 37.11. The parameters NSE and RSD had the values 0.99 and 0.1, respectively, indicating very good performance (Burgan & Aksoy, 2022). The z_i-values had a mean of -0.0204 and a variance of 2.98, which indicates a good fit for the semivariogram model (Bivand *et al*., 2013, p. 225). The obtained evaluation parameters were the best compared to the other interpolation approaches (for integration or downscaling).

Evaluation of RK integration of gauge measurements and satellite estimates

The results of the RK integration with TRMM at the original resolution are shown in part a of Figure 5. A gaussian semivariogram was fitted. A linear relationship between gauge measurements and the corresponding TRMM values was observed with an adjusted R^2 of 0.48 (r = 0.69, p-value = 3.77e-09). However, the original TRMM resolution generated a spatial pattern with many discontinuities in the resulting spatial layer.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

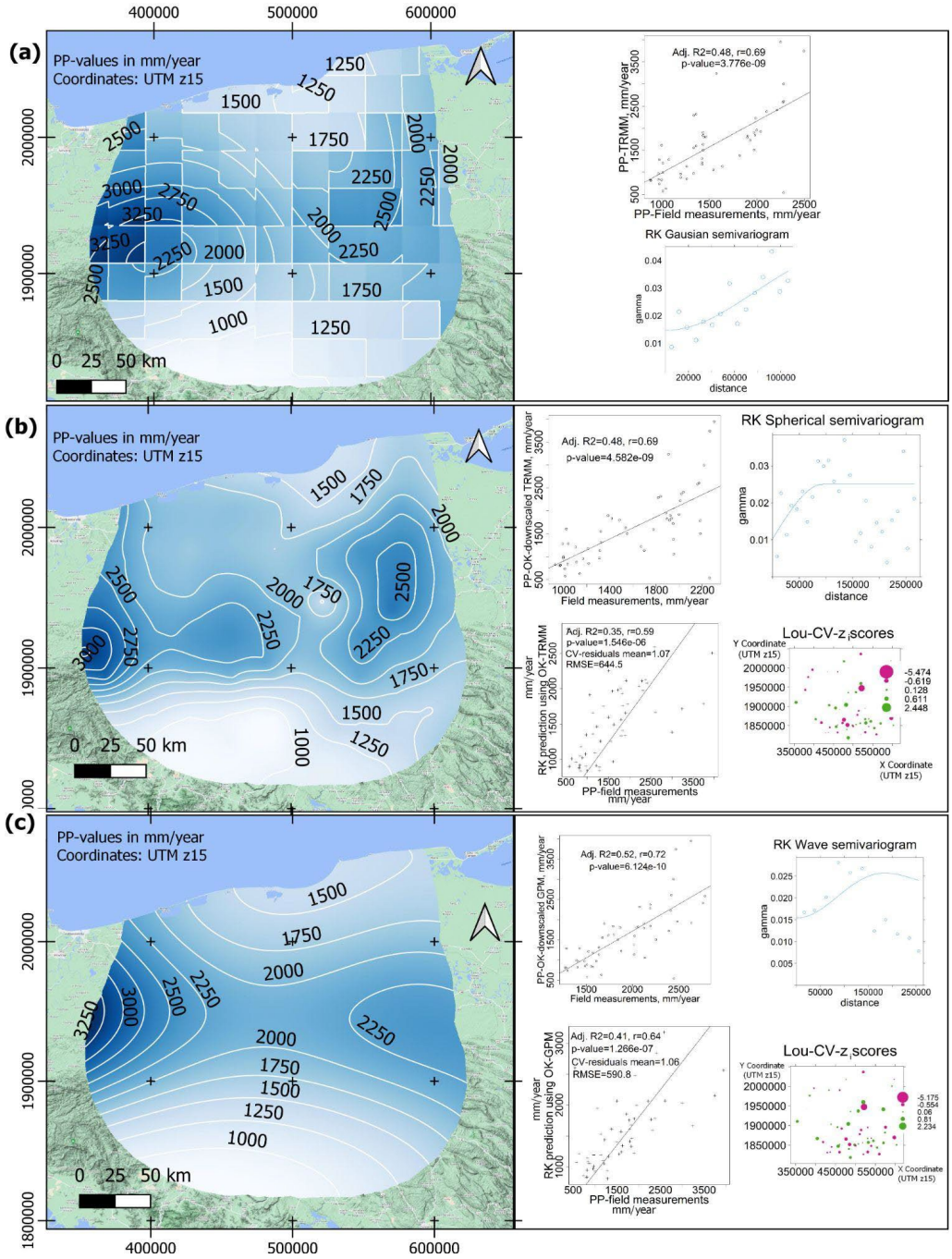

Figure 5. Prediction and Lou-CV results of the RK integration of yearly 2001 aggregated values of PP obtained from gauge measurements and satellite estimates. Part a: results of the integration with TRMM at the

original resolution. Part b: results of the integration with the OK-downscaled TRMM PP estimates. Part c: results of the integration with the OK-downscaled GPM PP estimates. Each part of the figure includes the map of the predicted PP-values, the scatterplot that shows the linear relationship between the integrated values and the fitted semivariogram. Parts (b) and (c) include the Lou-CV results with two graphs: a scatterplot between the OK-predictions and the measured values showing the values of R^2, r, p, the mean of the CV residuals, and the RMSE, and a map displaying the z_i spatial distribution.

The results of the RK-integration of gauge measurements using OK-downscaled TRMM as an auxiliary variable are shown in part b of Figure 5. A linear relationship (R^2 = 0.48, r = 0.69, p = 4.582e-09) between the gauge measurements values and the OK-downscaled TRMM values enabled to use of these downscaled values to obtain the trend of the predicted values in the applied RK model. A spherical semivariogram fitted the log-transformed gauge measurement values. The resulting spatial pattern was like the UK prediction (Figure 3, part b). The interpolation Lou-CV parameters R^2 = 0.35, r = 0.59, p-value = 1.54e-06, mean of the CV-residuals = 1.07, and RMSE = 644.5 indicated a moderate fit between the predicted and the measured PP-values. The parameters NSE and RSD had values 0.29 and 0.83, respectively, indicating an inadequate performance (Burgan & Aksoy, 2022). The R^2 and r values were higher for this interpolation compared with the corresponding values of the OK (Figure 3, part a) and UK (Figure 3, part b) interpolations of gauge measurements. This can be also observed in Table 2. The z_i-values had a

mean of -0.009 and a variance of 1.53. They indicated a good fit for the RK-semivariogram model (Bivand *et al*., 2013, p. 225). The z_i map shows that the locations with the lowest values (with a minimum of -5.47) were located at the center and the southeast parts of the area of study. The locations with the highest z_i-values (with a maximum of 2.23) were in the center and the middle of the area of study.

In the case of the RK integration using OK-downscaled GPM as auxiliary information (part b of Figure 5), a linear relationship (adjusted R^2 = 0.52, r = 0.72, and p = 6.124e-10) between the gauge measurements and these downscaled values enabled to obtain the trend of the predicted values in the applied RK model. A wave semivariogram was fitted. The Lou-CV parameters (R^2 = 0.41, r = 0.64, p-value = 1.26e-06, mean of the CV-residuals = 1.07, and RMSE = 590.8) indicated a moderate linear relationship between the integrated PP-values. The parameters NSE and RSD had the values 0.40 and 0.76, respectively, indicating an inadequate performance (Burgan & Aksoy, 2022). However, the mean (-0.009) and variance (1.53) of the z_i-values indicated a good fit of the semivariogram model (Bivand *et al*., 2013, p. 225). The lowest values of the z_i-map (with a minimum of -5.17) were located at the center and the southeast part of the study area. The highest values (with a maximum of -2.45) were observed also in the center and southeast parts. These results indicate an acceptable performance of the predictive model to obtain a locally suitable spatial layer. Among the methods that interpolated gauge measurements, the RK-integration with these measurements with OK-downscaled GPM-values resulted in the best Lou-CV parameters (Table 2).

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

PCA and correlation comparison among layers of predicted PP-values

The heatmap, the correlogram, and the PCA biplot to define similarity among the values of the obtained PP-layers are included in Figure 6. The results of RK-integration of gauge measurements with OK-downscaled GPM values (shown in the figure as RK_OK.GPM) and OK-downscaling of GPM (OK_GPM) were the closest to each other by forming a cluster with hot colors in the correlation heatmap (shown in the figure as Figure 6, part a). OK-downscaled TRMM (shown in the figure as OK_TRMM) and IDW were the more separated results from all the others. They were not grouped in any cluster and had the coldest colors in the heatmap. The same situation was observed in the PCA biplot (Figure 6, part c). The results of OK-downscaled TRMM were the most separated from the other results, followed by IDW. As observed in this biplot, the results of OK-downscaled GPM (shown in the figure as RK_OK.TRMM) and RK-integration of gauge measurements with OK-downscaled GPM values were close to each other. The results shown in the three parts of Figure 6 verified the similarity between the predictions of OK and UK of gauge measurements. They were close to each other in the PCA biplot and the heatmap. The heatmap showed warm colors between them with an *r* of 0.96, but they were separated from the other interpolation results.

Tecnología y Ciencias del Agua

OPEN ACCESS

Check for updates

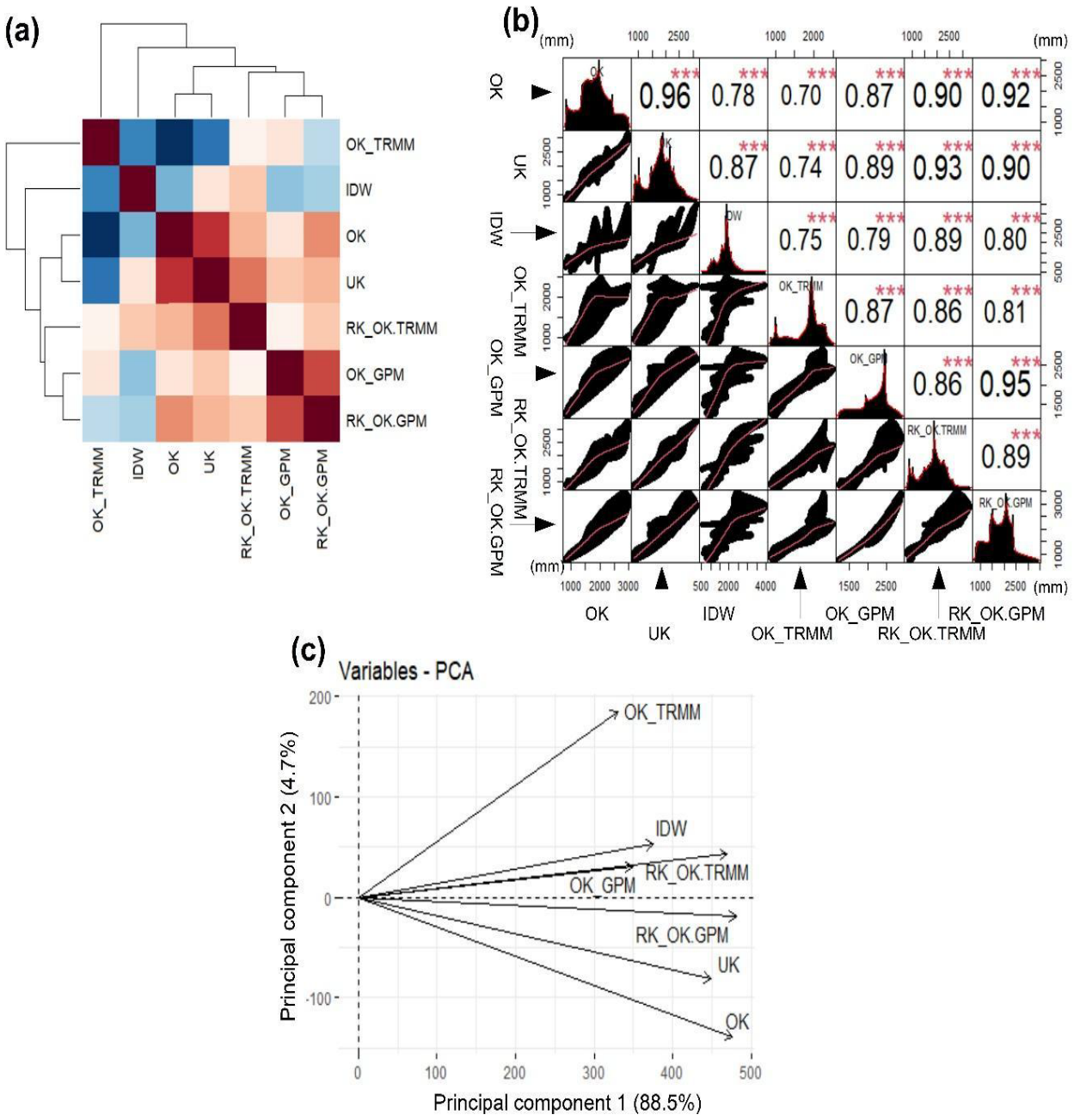

Figure 6. (a): Heatmap of correlation analysis showing clustering by dendrogram, (b): correlogram with scatterplots of the bivariate comparisons, *r* and significances (shown with the asterisks, three of

them mean high significance), as well as histograms of each prediction result, and (c): biplot of the PCA of the predicted PP values. The compared results are predictions using OK, UK and IDW of gauge measurements, OK downscaled TRMM (OK_TRMM), OK downscaled GPM (OK_GPM), RK integration of gauge measurements with OK-downscaled TRMM values (RK_OK.TRMM), and RK-integration of gauge measurements with OK-downscaled GPM values (RK_OK.GPM).

The closeness between the results of OK-downscaled GPM and the results of the RK-integration of these with gauge measurements with OK-downscaled GPM values was confirmed in the correlogram (Figure 6, part b) with an r of 0.95. All correlations were highly significant (indicated by three asterisks in the correlation matrix). IDW showed $r < 0.8$ with all the other interpolation results except for a $r = 0.89$ with RK-integration of gauge measurements and OK-downscaled TRMM values, confirming its low performance and the closeness between both results.

Regarding the scatterplots of the bivariate comparison (Figure 6, part b), the results were different, particularly in the higher PP-values. A trend line could be fitted for their low values, but it presented an abrupt slope reduction for their higher values. That happened, for example, between the RK-integration of gauge measurements with OK-downscaled GPM values and IDW and between OK-downscaled GPM and OK-downscaled TRMM. The histograms of the prediction results are shown in Figure 6, part b, as a diagonal of squares, formed from the upper left corner to the lower right corner. They indicated the dissimilarity of IDW

with respect to all the other predictions. This dissimilarity was observed in terms of the shape and scale of the histogram.

Discussion

The focus of the present work was to evaluate geostatistics to downscale satellite estimates (TRMM and GPM) and to integrate the results with gauge measurements to obtain a spatial layer of annual PP-values suitable to be used at the local scale in a tropical basin, taking a year as a case of study. This approach is part of the work´s novelty, since, after the performed literature review, no work had been done about the geostatistical downscaling of GPM PP estimates and its integration using RK with gauge measurements in a tropical basin and outside Asia and Europe.

The studied satellite estimates integrate, since its production, gauge data of very few stations (three at the most) for the studied zone (GPCC, 2012). Therefore, in the present study, all the available gauge data were integrated with the satellite estimates to obtain a local suitable spatial PP layer. The OK downscaling procedure of GPM satellite estimates resulted in a substantial improvement in the local spatial pattern of the resulting layer.

Among the interpolation approaches of gauge measurements, the best-Lou-CV evaluated method was the RK integration of these measurements with OK-downscaled GPM values. These results can be

suitable to be used at a local scale because they integrated all the available gauge data, considered the reference values (New *et al*., 2001), and the GPM estimates, which can capture the spatial variability of the PP. However, the parameters NSE and RSD indicated an inadequate performance of this geostatistical integration approach. Therefore, more research considering aggregated values of other years and other climatic conditions is required to confirm these results. The OK and UK interpolations had lower values of the Lou-CV parameters than the RK results. OK was slightly better than UK. The observed spatial pattern of the predicted values from all interpolation approaches was different. This pattern can be the first element to select a method to obtain a PP spatial layer useful at the local scale. Given that the PP patterns are distributed over a regional or global context more than a local range around the gauge stations (Smalley & L'Ecuyer, 2015), the resulted IDW spatial pattern was considered incorrect.

According to the Lou-CV parameters, the RK-integration of gauge measurements with OK-downscaled GPM outperforms the integration of gauge measurements with OK-downscaled TRMM. The closeness between the results of OK-downscaled GPM and the RK-integration of gauge measurements with OK-downscaled GPM values, observed in the PCA and the correlation analysis (described in the previous section), is interesting since the former includes measurements, and the latter does not. Additionally, Figure 4 (part b) and Figure 5 (part c) show a similar spatial pattern between both predictions but different from the other results. This observation confirms that GPM satellite estimates can be downscaled

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

utilizing OK to obtain a layer suitable at the local scale with or without integration with field measurements.

Collocated cokriging (CK) was not included in this work to make the integration because it was developed for situations where the auxiliary information is not spatially exhaustive (Knotters, Brus, & Voshaar, 1995; Hengl *et al*., 2007). According to Hengl *et al*. (2007), when applying RK, the auxiliary variable should not account for 100 % of the variability on the value to interpolate. The condition is that both variables are correlated. No linear relationship between the gauge values and altitude was found. However, for other climatic conditions and geographic zones, elevation can be a factor that partly defines the spatial variability of PP. Therefore, its inclusion can improve the estimates within the here proposed geostatistical integration scheme, as found by other studies (Goovaerts, 2000; Rata *et al*., 2020). The effects of the pixel size, methods that determine parameter uncertainty, such as Bayesian kriging, and the inclusion of other auxiliary variables, such as the distance to the coast, can be further investigated. All the mentioned are considerations to be included in future research.

Previously performed research (Abdollahipour *et al*., 2022; Park *et al*., 2017; Chen *et al*., 2020; Chen *et al*., 2019) included performance metrics to validate with field measurements but not the determination of the similarity of the predicted PP values using PCA, cluster and correlogram analysis, taking a low-performance interpolation approach as reference. In this research, IDW of gauge measurements was considered the reference for low performance. This way, a similarity was found between the results of IDW and the RK integration of gauge

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

measurements with OK downscaled TRMM and between the results of IDW and OK downscaled TRMM by itself. These results, and the lowest Lou-CV parameter values of IDW, indicated that TRMM estimates were not suitable at the local scale; although, the TRMM OK downscaling became high Lou-CV evaluation parameters (zi-values, RMSE, NSE, and RSD). These findings are relevant since TRMM satellite estimates at different periods have been the most integrated with field measurements using kriging (Abdollahipour *et al*., 2022; Park *et al*., 2017; Chen *et al*., 2020; Chen *et al*., 2019).

Conclusion

The present work shows that geostatistical techniques can help to downscale satellite PP data and integrate them with field values to generate PP spatial layers that can be useful at the local scale for tropical basins as the studied area. The evaluation of results, which usually involves performance metrics from cross-validation procedures, should be complemented with methods to define the similarity of the resulted values to a reference interpolation result which can be of low performance, as done in the present research. The reason for this is that, although an interpolation result can have good validation performance metrics, at the same time, it can have a high degree of similarity to the low-performance layer. It means that the applied interpolation method can perform well predicting the observed values, while it can perform not well doing the same at not measured locations, like the reference (low performance)

layer. As previous research found, geostatistical techniques can be applied to downscale TRMM satellite data and integrate it with field values. However, this study found that the TRMM results can be like IDW, and therefore, in such cases, it is not advisable to use them at the local scale for tropical basins as the study zone. Instead, OK downscaled GPM can be used with or without RK geostatistical integration with gauge measurements to obtain a PP layer suitable to the local scale. The developed research lets to conclude that OK performs well to downscale PP GPM satellite estimates.

The RK integration of OK-downscaled GPM satellite estimates with gauge measurements can substantially improve the Lou-CV indicators compared to the predicted values from UK and OK predictions of gauge measurements. The resulting spatial layer can be considered an improved product in terms of being more suitable for the local scale, at which the decision-making process can be made. According to the PCA, cluster, and correlogram analysis, the spatial layers obtained from OK downscaled GPM data, as well as the spatial layer resulting from its RK integration with gauge measurements, show similar values meaning that they are suitable for the local scale.

This research showed that geostatistical methods are effective for downscaling satellite estimates and for integrating them with rain gauge measurements to obtain spatial layers capable of supporting the decision-making process at the local level.

Tecnología y ciencias del agua, ISSN 2007-2422, 15(1), 54-110. DOI: 10.24850/j-tyca-15-01-02

Acknowledgments

Thanks to Dr. Eugenio Reyes for inviting the author to participate in the CONACYT project 248719. Thanks to Dr. Charlotte Smith (Berkeley University) for her comments on the article.

This research was funded by Conacyt grant number (project) 248719 of the 2014 National Problems Proposals and FORDECYT 2018-10 grant number 297259.

Appendix

```
// Google Eearth Engine (GEE) script to download GPM data.
var dataset = ee.ImageCollection('NASA/GPM_L3/IMERG_MONTHLY_V06')
 .filterDate('2001-01-01', '2001-12-31');

// Select the max precipitation and mask out low precipitation values.
var precipitation = dataset.select('precipitation').sum();
var mask = precipitation.gt(0.01);
var precipitation = precipitation.updateMask(mask);
var precipitation = precipitation.multiply(720)
var palette = [
 '000096','0064ff', '00b4ff', '33db80', '9beb4a',
```

```
 'ffeb00', 'ffb300', 'ff6400', 'eb1e00', 'af0000'
];
var precipitationVis = {min: 0.0, max: 2000, palette: palette};
Map.addLayer(precipitation, precipitationVis, 'Precipitation');
Map.setCenter(-92, 17, 7);

var geometry = ee.Geometry.Rectangle([-94.4, 18.8, -91.5116, 15.0]);

Export.image.toDrive({
 image: precipitation,
 description: "GPM_Precipitation_2001_mmanno",
 scale: 11000,
 region: geometry,
 fileFormat: "GeoTIFF",
 });
1
```

References

Abdollahipour, A., Ahmadi, H., & Aminnejad, B. (2022). A review of downscaling methods of satellite-based precipitation estimates. *Earth Science Informatics* 15, 1-20. Recovered from https://doi.org/10.1007/s12145-021-00669-4

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research*, 116, 1-11. Recovered from https://doi.org/10.1029/2010jd014741

Agou, V. D., Varouchakis, E. A., & Hristopulos, D. T. (2019). Geostatistical analysis of precipitation in the island of Crete (Greece) based on a sparse monitoring network. *Environmental Monitoring Assessment* 191(353), 1-24. Recovered from https://doi.org/10.1007/s10661-019-7462-8

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., & Papadopoulos, A. (2010). Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates. *IEEE Transactions Geosciences and Remote Sensing*, 48, 1667-1683. Recovered from https://doi.org/10.1109/TGRS.2009.2034736

Bell, T. L. (2003). Comparing satellite rainfall estimates with rain gauge data: Optimal strategies suggested by a spectral model. *Journal of Geophysical Research*, 108(D3)4121, 1-15. Recovered from https://doi.org/10.1029/2002JD002641

Berndt, C., Rabiei, E., & Haberlandt, U. (2013). Geostatistical merging of rain gauge and radar data for high temporal resolutions and various station density scenarios, *Journal of Hydrology*, 508, 88-101. Recovered from https://doi.org/10.1016/j.jhydrol.2013.10.028

Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2013). *Applied spatial data analysis with R*. New York, USA: Springer. Recovered from https://doi.org/10.1007/978-0-387-78171-6

Bowman, K.P. (2005). Comparison of TRMM P retrievals with rain gauge data from ocean buoys. *Journal of Climate*, 18(1), 178-190. Recovered from https://doi.org/10.1175/JCLI3259.1

Burgan, H. I., & Aksoy, H. (2022). Daily flow duration curve model for ungauged intermittent sub-basins of gauged rivers. *Journal of Hydrology*, 604, 127429. Recovered from https://doi.org/10.1016/j.jhydrol.2021.127249

Cepal & Cenapred, Centro Nacional de Prevención de Desastres & Comisión Económica para América Latina y el Caribe. (2008). *Tabasco: características e impacto socioeconómico de las inundaciones provocadas a finales de octubre y a comienzos de noviembre de 2007 por el frente frío número 4. Informe LC/MEX/L.864.* Recovered from https://repositorio.cepal.org/handle/11362/25881

Cersosimo, A., Larosa, S., Romano, F., Cimini, D., Di Paola, F., Gallucci, D., Gentile, S., Geraldi, E., Teodosio-Nilo, S., Ricciardelli, E., Ripepi, E., & Viggiano, M. (2018). Downscaling of satellite OPEMW surface rain intensity data. *Remote Sensing*, 10(11), 1763, 1-16. Recovered from https://doi.org/10.3390/rs10111763

Chen, F., Gao, Y., Yiguo, W., & Li, X. (2020). A downscaling-merging method for high-resolution daily precipitation estimation. *Journal of Hydrology*, 581(124414), 1-15. Recovered from https://doi.org/10.1016/j.jhydrol.2019.124414

Chen, S., Zhang, L., She, D., & Chen, J. (2019). Spatial downscaling of tropical rainfall measuring mission (TRMM) annual and monthly precipitation data over the middle and lower reaches of the Yangtze River Basin, China. *Water*, 11(3), 568. Recovered from https://doi.org/10.3390/w11030568

Conagua, Comisión Nacional del Agua. (2015). *Estadísticas del agua en México*. Recovered from https://agua.org.mx/biblioteca/estadisticas-del-agua-en-mexico-edicion-2015/

Cressie, N. (1990). The origins of kriging. *Mathematical Geology*, 22, 239-252. Recovered from https://doi.org/10.1007/BF00889887

Curran, P. J., & Atkinson, P. M. (1998). Geostatistics and remote sensing. *Progress in Physical Geography: Earth and Environment*, 22(1), 61-78. Recovered from https://doi.org/10.1177/030913339802200103

Dumitrescu, A., Brabec, M., & Matreata, M. (2020). Integrating ground-based observations and radar data into gridding sub-daily precipitation. *Water Resources Management*, 34, 3479-3497. Recovered from https://doi.org/10.1007/s11269-020-02622-4

GPCC, Global Precipitation Climatology Centre. (2012). *Number of stations used by GPC for May 2012*. Recovered from https://climatedataguide.ucar.edu/climate-data/gpcc-global-precipitation-climatology-centre

Gobierno de México. (November 12, 2020). *Información de estaciones climatológicas.* Recovered from https://smn.conagua.gob.mx/es/climatologia/informacion-climatologica/informacion-estadistica-climatologica

GES DISC, Goddard Earth Sciences Data and Information Services Center. (2011). *Tropical Rainfall Measuring Mission (TRMM) (2011), TRMM (TMPA/3B43) Rainfall Estimate L3 1 month 0.25 degree x 0.25 degree V7. [Data file]*. Recovered from https://doi.org/10.5067/TRMM/TMPA/MONTH/7

Goovaerts, P. (1997). *Geostatistics for natural resource evaluation*. New York, USA: Oxford University Press.

Goovaerts, P. (2000). Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology*, 228(1–2), 113-129. Recovered from https://doi.org/10.1016/S0022-1694(00)00144-X

Greene, J. S., & Morryssey, M. L. (2000). Validation and uncertainty analysis of satellite rainfall algorithms. *Professional Geographer*, 52(29), 247-258. Recovered from https://doi.org/10.1111/j.0033-0124.2000.t01-1-.x

Hengl, T. (2009). *A practical guide to geostatistical mapping of environmental variables.* Luxembourg, Luxembourg: European Communities, Publications Office. Recovered from http://spatial-analyst.net/book/system/files/Hengl_2009_GEOSTATe2c1w.pdf

Hengl, T., Gerard, B., Heuvelink, M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. *Computer and Geosciences*, 33(10), 1301-1315. Recovered from https://doi.org/10.1016/j.cageo.2007.05.001

Holawe, F., & Dutter, R. (1999). Geostatistical study of precipitation series in Austria: Time and space. *Journal of Hydrology*, 219(1-2), 70-82. Recovered from https://doi.org/10.1016/S0022-1694(99)00046-3

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K.-L., Joyce, R. J., Kidd, C. E., Nelkin, J., Sorooshian, S., Stocker, E. F., Tan, J., Wolff, D. B., & Xie, P. (2020). Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). In: Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., & Turk, F. (eds.). *Satellite precipitation measurement. Advances in global change research*. Vol. 69 (pp. 3-23). Cham, Switzerland: Springer. Recovered from https://doi.org/10.1007/978-3-030-35798-6_1

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., Hong, Y., Bowman, K. P., & Stocker, E. F. (2007). The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, 8(1), 38-55. Recovered from https://doi.org/10.1175/JHM560.1

Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E. J., & Tan, J. (2019). *GPM IMERG final precipitation L3 half hourly 0.1 degree x 0.1 degree V06.* GES DISC, Goddard Earth Sciences Data and Information Services Center. Recovered from https://doi.org/10.5067/GPM/IMERG/3B-HH/06

INEGI, Instituto Nacional de Estadística y Geografía. (2010). *Red hidrográfica mexicana.* Recovered from https://www.inegi.org.mx/temas/hidrografia/default.html#Descargas

Keblouti, M., Ouerdachi, L., & Boutaghane, H. (2012). Spatial interpolation of annual precipitation in Annaba-Algeria. Comparison and evaluation of methods. *Energy Procedia*, 18, 468-475. Recovered from https://doi.org/10.1016/j.egypro.2012.05.058

Kidd, C., Takayabu, Y. N., Skofronick-Jackson, G. M., Huffman, G. J., Braun, S. A., Kubota, T., & Turk, F. J. (2020). The global precipitation measurement (GPM) mission. Integrated multi-satellite retrievals for the global precipitation measurement (GPM) mission (IMERG). In: Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., & Turk, F. (eds.). *Satellite precipitation measurement. Advances in global change research*. Vol. 69 (pp. 3-23). Cham, Switzerland: Springer. Recovered from https://doi.org/10.1007/978-3-030-24568-9_19

Knotters, M., Brus, D., & Voshaar, J. (1995). A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. *Geoderma*, 67 (3-4), 227-246.

Kumar, V. (2007). Optimal contour mapping of groundwater levels using universal kriging—A case study. *Hydrological Sciences Journal*, 52, 1038-1050. Recovered from https://doi.org/10.1623/hysj.52.5.1038

Kummerow, C., Simpson, J., Thiele, O., Barnes, W., Chang, A. T. C., Stocker, E., Adler, R. F., Hou, A., Kakar, R., Wentz, F., Ashcroft, P., Kozu, T., Hong, Y., Okamoto, K., Iguchi, T., Kuroiwa, H., Im, E., Haddad, Z., Huffman, G., Ferrier, B., Olson, W. S., Zipser, E., Smith, E. A., Wilheit, T. T., North, G., Krishnamurti, T., & Nakamura, K. (2000). The status of the tropical rainfall measuring mission (TRMM) after two years in orbit. *Journal of Applied Meteorology and Climatology*, 39, 1965-1982. Recovered from https://doi.org/10.1175/1520-0450(2001)040<1965:TSOTTR>2.0.CO;2

Kummerow, C., Barnes, W., Kozu, T., Shiue, J., & Simpson, J. (1998). The tropical rainfall measuring mission (TRMM) sensor package. *Journal of Atmospheric and Oceanic Technology*, 15, 809-817. Recovered from https://doi.org/10.1175/1520-426(1998)015<0809:TTRMMT>2.0.CO;2

Laurent, H., Jobard, I., & Toma, A. (1998). Validation of satellite and ground based estimates of precipitation over the Sahel. *Atmospheric Research*, 47-48, 651-670. Recovered from https://doi.org/10.1016/S0169-8095(98)00051-9

Lin, A., & Wang, X. L. (2011). An algorithm for blending multiple satellite precipitation estimates with in situ precipitation measurements in Canada. *Journal of Geophysical Research: Atmospheres*, 116(D21), 1-19. Recovered from https://doi.org/10.1029/2011JD016359

Massari, C., & Maggioni, V. (2020). Error and uncertainty characterization. In: Levizzani V., Kidd C., Kirschbaum D., Kummerow C., Nakamura K., & Turk F. (eds.). *Satellite precipitation measurement. Advances in global change research*. Vol. 69. Cham, Switzerland: Springer. Recovered from https://doi.org/10.1007/978-3-030-35798-6_4

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58, 1246-1266.

Matheron, G. (1965). *Les variables régionalisées et leur estimation.* (Thèse). Paris, France: Masson.

Matheron, G. (1971). *The theory of regionalized variables* (English translation). Les Cahiers du Centre de Morphologie Mathématique, Fasc 5. Paris, France: Ecole Nationale Supérieure des Mines de Paris (NSMP).

Morin, E., Marra, F., & Armon, M. (2020). Dryland precipitation climatology from satellite observations. In: Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., & Turk, F. (eds.). *Satellite precipitation measurement. Advances in global change research*. Vol. 69 (pp. 843-860). Cham, Switzerland: Springer. Recovered from https://doi.org/10.1007/978-3-030-35798-6_19

Nerini, D., Zulkafli, Z., Wang, L. P., Onof, C., Buytaert, W., Lavadocasimiro, W., & Guyot, J. L. (2015). A comparative analysis of TRMM-rain gauge data merging techniques at the daily time scale for distributed rainfall-runoff modeling applications. *Journal of Hydrometeorology*, 16, 2153-2168. Recovered from https://doi.org/10.1175/JHM-D-14-0197.1

New, M., Todd, M., Hulme, M., & Jones, P. (2001). Precipitation measurements and trends in the twentieth century. *International Journal of Climatology* 21, 1899-1922. Recovered from https://doi.org/10.1002/joc.680

Oliver, M. A., & Webster, R. (2015). *Basic steps in geostatistics: The variogram and kriging*. Heidelberg, New York, Dordrecht, London: Springer Cham. Recovered from https://doi.org/10.1007/978-3-319-15865-5

Park, N. W., Kyriakidis, P. C., & Hong, S. (2017). Geostatistical integration of coarse resolution satellite precipitation products and rain gauge data to map precipitation at fine spatial resolutions. *Remote Sensing*, 9(3), 255, 1-19. Recovered from https://doi.org/10.3390/rs9030255

Pebesma, E. J. (2004). Multivariable geostatistics in S: The gstat package. *Computer and Geosciences*, 30, 683-691. Recovered from https://doi.org/10.1016/j.cageo.2004.03.012

Pielke, R. A., Stohlgren, T., Parton, W., Doesken, N., Money, J., & Schell, L. (2000). Spatial representativeness of temperature measurements from a single site. *Bulletin of the American Meteorological Society*, 81(4), 826-830. Recovered from https://www.jstor.org/stable/26215143

R Core Team. (n.d.). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Recovered from https://www.R-project.org

Rata, M., Douaoui, A., Larid, M., & Douaik, A. (2020). Comparison of geostatistical interpolation methods to map annual rainfall in the Chéliff watershed, Algeria. *Theoretical and Applied Climatology*, 141, 1009-1024. Recovered from https://doi.org/10.1007/s00704-020-03218-z

Ruiz-Alvarez, O., Singh, V. P., Enciso-Medina, J., Ontiveros-Capurata, R. E., & Corrales-Suastegui, A. (2020). Spatio-temporal trends of monthly and annual precipitation in Aguascalientes, Mexico. *Atmosphere*, 11(5), 437. Recovered from https://doi.org/10.3390/atmos11050437

Sivasubramaniam, K., Sharma, A., & Alfredsen, K. (2019). Merging radar and gauge information within a dynamical model combination framework for P estimation in cold climates. *Environmental Modelling & Software,* 119, 99-110. Recovered from https://doi.org/10.1016/j.envsoft.2019.05.013

Skofronick-Jackson, G., Kirschbaum, D., Petersen, W., Huffman, G., Kidd, C., Stocker, E., & Kakar, R. (2018). The global p measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Quarterly Journal of the Royal Meteorological Society*, 144(Suppl. 1), 27-48. Recovered from https://doi.org/10.1002/qj.3313

Smalley, M., & L'Ecuyer, T. (2015). A Global assessment of the spatial distribution of precipitation occurrence. *Journal of Applied Meteorology and Climatology*, 54(11), 2179-2197. Recovered from https://doi.org/10.1175/JAMC-D-15-0019.1

Tapia-Silva, F. O., & Gómez-Reyes, E. (2020). Disponibilidad natural de agua en las subregiones de estudio. En: Aguilar-Benitez (ed.). *La gestión de los usos del agua en tres subregiones hidrológicas: río San Juan, Valle de México y Bajo Grijalva* (pp. 55-80), Tijuana, México: El Colegio de la Frontera Norte.

Tapia-Silva, F. O., Silván-Cárdenas, J. L., & Rosales-Arriaga, E. (2013) Análisis espacial, hacia una utilización mejorada de la información medida en campo y por satélites, para apoyar la toma de decisiones en materia hídrica y Ambiental. *Tecnología y ciencias del agua*, 4(1), 149-166. Recovered from http://www.revistatyca.org.mx/index.php/tyca/article/view/343/304

Toté, C., Patricio, D., Boogaard, H., Van der Wijngaart, R., Tarnavsky, E., & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, 7(2), 1758-1776. Recovered from https://doi.org/10.3390/rs70201758

Van der Meer, F. (2012). Remote-sensing image analysis and geostatistics. *International Journal of Remote Sensing*, 33(18), 5644-5676. Recovered from https://doi.org/10.1080/01431161.2012.666363

Verdin, A., Rajagopalan, B., Kleiber, W., & Funk, C. (2015). A Bayesian kriging approach for blending satellite and ground precipitation observations. *Water Resources Research*, 51(2), 908-921. Recovered from https://doi.org/10.1002/2014wr015963

Wackernagel, H. (2003). *Multivariate geostatistics: An introduction with applications* (3rd ed.). Berlin, Heidelberg, Germany: Springer Science & Business Media.

Wang, X. L., & Lin, A. (2015). An algorithm for integrating satellite precipitation estimates with in situ precipitation data on a pentad time scale. *Journal of Geophysical Research: Atmospheres*, 120, 3728-3744. Recovered from https://doi.org/10.1002/2014JD022788.

WMO, World Meteorological Organization. (2008). *Guide to meteorological instruments and methods of observation.* WMO-No. 8 (7th ed.). Geneva, Switzerland: World Meteorological Organization. Recovered from https://www.weather.gov/media/epz/mesonet/CWOP-WMO8.pdf

Wu, Z., Zhang, Y., Sun, Z., Lin, Q., & He, H. (2018). Improvement of a combination of TMPA (or IMERG) and ground-based P and application to a typical region of the East China Plain. *Science of the Total Environment*, 640-641, 1165-1175. Recovered from https://doi.org/10.1016/j.scitotenv.2018.05.272

Yang, P., & Ng, T. L. (2019). Fast Bayesian regression kriging method for real-time merging of radar, rain gauge, and crowdsourced rainfall data. *Water Resources Research*, 55, 3194-3214. Recovered from https://doi.org/10.1029/2018WR023857

Zulkafli, Z., Buytaert, W., Onof, C., Manz, B., Tarnavsky, E., Lavado, W., & Guyot, J.-L. (2014). A comparative performance analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean-Amazon River Basins. *Journal of Hydrometeorology*, 15, 581-592. Recovered from https://doi.org/10.1175/JHM-D-13-094.1