Nature Environment and Pollution Technology
An International Quarterly Scientific Journal

p-ISSN: 0972-6268
(Print copies up to 2016)
e-ISSN: 2395-3454

Vol. 23 | No. 2 | pp. 1035-1045 | 2024

Original Research Paper

https://doi.org/10.46488/NEPT.2024.v23i02.037

Open Access Journal

# Statistical Performance of Gridded Rainfall Datasets Over Ungauged Jalaur River Basin, Philippines

**Christsam Joy S. Jaspe-Santander*(**)† and Ian Dominic F. Tabañag*(***)**

*School of Engineering, University of San Carlos, Talamban, Cebu City, 6000, Philippines

**College of Engineering, Central Philippine University, Jaro, Iloilo City 5000, Philippines

***Philippine Council for Industry, Energy, and Emerging Technology Research and Development, Department of Science and Technology (DOST-PCIEERD), Bicutan, Taguig City 1631, Philippines

† Corresponding author: Christsam Joy S. Jaspe-Santander; cjjsantander@cpu.edu.ph

Nat. Env. & Poll. Tech.
Website: www.neptjournal.com

*Received:* 07-09-2023
*Revised:* 15-12-2023
*Accepted:* 22-12-2023

**Key Words:**
Gridded rainfall dataset
Jalaur river basin
Ungauged basin

## ABSTRACT

The study presented aims to find the most appropriate climate dataset for the data-scarce Jalaur River Basin (JRB), Iloilo, Philippines, by evaluating the statistical performance of five rainfall datasets (APHRODITE, CPC NOAA, ERA5, SA-OBS, and PGF-V3) with resolutions of 0.25° and 0.5° having a time domain of 1981 to 2005. Bilinear interpolation implemented through Climate Data Operator (CDO) was used to extract and process grid climate datasets with Linear scaling as bias correction to minimize product simulation uncertainties. The datasets were compared to the lone meteorological station nearest to JRB investigated at monthly and annual timescales using six statistical metrics, namely, Pearson's correlation coefficient ($r$), coefficient of determination ($R^2$), modified index of agreement ($d_1$), Kling-Gupta efficiency, Nash-Sutcliffe efficiency (NSE), and RMSE-observations standard deviation ratio (RSR). The results indicate a strong positive correlation with the observed data for both rainfall and temperature ($r > 0.8$; $R^2$, $d_1 > 0.80$). Although graphical observation shows an underestimation of rainfall, goodness-of-fit values indicate very good model performance (NSE, KGE > 0.75; RSR < 0.50). In terms of temperature, variable responses are observed with significant overestimation for maximum temperature and underestimation for minimum temperature. SA-OBS proved to be the best-performing dataset, followed by ERA5 and PGF-V3. These key findings supply useful information in deciding the most appropriate gridded climate dataset for hydrometeorological investigation in the JRB and could enhance the regional representation of global datasets.

## INTRODUCTION

High-resolution gridded climate data are necessary input parameters for atmospheric, climatic, hydrological, and ecological studies to establish the hydrological characteristics of a watershed. They are particularly important in modeling and simulation for the evaluation of numerical forecast datasets and impact studies. However, due to the geographical conditions and the shortage of meteorological observations in the Philippines, it is challenging to obtain accurate regional area rainfall amounts in river basins.

The availability of rainfall data is still a major issue in water resource studies (Nassaj et al. 2022, Ang et al. 2022). The quality of rainfall data is critical in the development of a reliable hydrological model to generate streamflow estimates and trends over extended periods (Try et al. 2020). In regions where rainfall data is limited and coarsely distributed, gridded satellite and reanalysis climate datasets have been widely used to supply the lack of long-term, high-spatial-resolution data (Peralta et al. 2020). However, these datasets vary in spatial resolution and temporal coverage and are highly influenced by their method of production (Ayoub et al. 2020).

Numerous studies have used gridded datasets as an alternate input for gauged data in hydrological modeling applications. Ang et al. (2022) evaluated seven gridded meteorological datasets of rainfall and air temperature covering the data-sparse Tonle Sap Lake Basin in Cambodia. Schumacher et al. (2020) described the performance of five gridded datasets in reproducing rainfall and/or temperature over the complex terrain in the high Chilean Andes, and Try et al. (2020) assessed the performances of various gridded rainfall datasets in rainfall-runoff and flood-inundation modeling of the Mekong River Basin.

This study focused on the statistical evaluation of five rainfall datasets (APHRODITE, CPC NOAA, ERA5, SA-OBS, and PGF-V3) and four temperature datasets (CPC NOAA, ERA5, SA-OBS, and PGF-V3) to investigate their reliability as alternative climate data source to characterize the climate conditions in JRB. These datasets were compared with the lone ground-based observation station in the basin to identify the most accurate data source that could be used in further evaluation and analysis to characterize the climate condition of the study area.

## MATERIALS AND METHODS

### Study Area

The Jalaur River Basin (JRB) lies between 10°45' to 11°10' North latitude and 122°12' to 122°55' East longitude in Panay Island, Philippines. The basin area is 1650 sq km, receiving approximately 2100 mm of rainfall annually. An estimated population of 430,000 reside within the basin, with agriculture as the main source of livelihood. A sizable percentage of land is dedicated for agricultural purposes (45%). The remaining part is classified as grassland/shrubland (39%), woodland/forest (15%) and wetland (1%). Long-term rainfall and temperature data are recorded in the lone meteorological station near the basin, while five (5) flow gauging stations are located along the main river and its tributaries (Fig. 1).

The JRB is projected to be in critical condition in water availability and accessibility by 2025 due to climate and anthropogenic changes. It is crucial to study its hydrologic characteristics and develop a quantitative prediction model to assess its hydrologic response to these changes (Arceo et al. 2018).

### Climate Datasets

The observed gauge-based rainfall dataset in the JRB is limited. High-quality daily observation of rainfall and maximum and minimum temperature is available for only one station acquired from the Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA) dataset for the period 1951-2009. Considering temporal availability and spatial resolution and from previous studies in the Asian region, five rainfall datasets (APHRODITE, CPC NOAA, ERA5, SA-OBS, and PGF-V3) and four temperature datasets (CPC NOAA, ERA5, SA-OBS, and PGF-V3) are investigated (Table 1).

APHRODITE is a long-term daily rainfall and temperature dataset developed by the Research Institute for Humanity and Nature and the Meteorological Research Institute of Japan Meteorological Agency. These rainfall records are produced from dense rain-gauge observation networks covering domains of Monsoon Asia, the Middle East, and Northern Eurasia, available at 0.25° and 0.5° resolutions, and for Japan

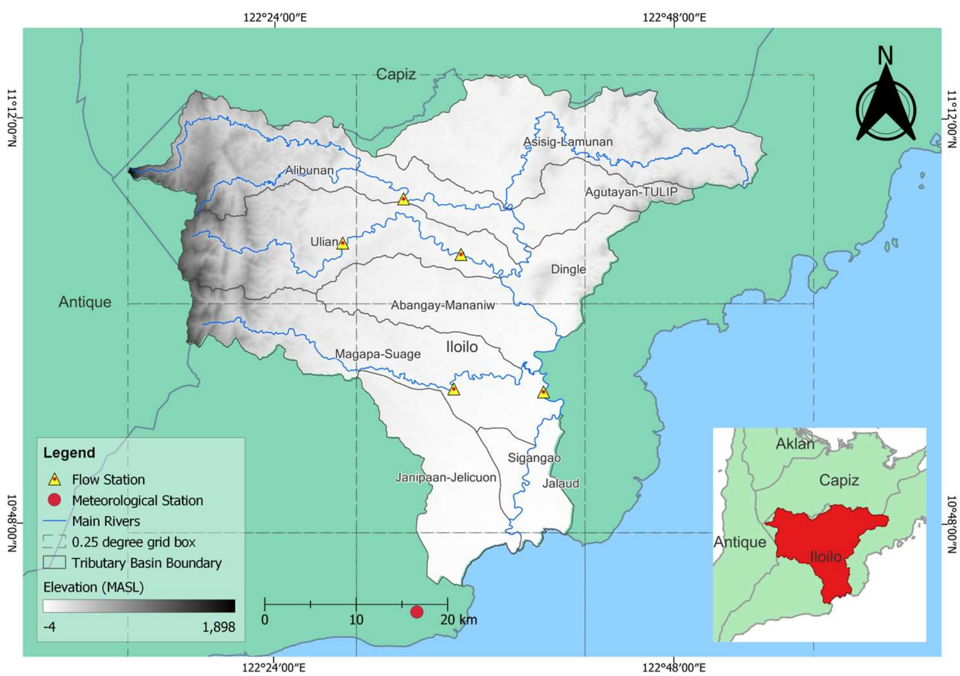



Fig. 1: Location Map of the Jalaur River Basin. The dashed boxes represent 0.25 x 0.25° grid cells. The inset indicates the location of the JRB in Panay Island, Philippines.

at 0.05° resolution (Yatagai et al. 2012).

CPC NOAA is a global reanalysis dataset by the American National Oceanic and Atmospheric Administration (NOAA). The Global Telecommunication System (GTS) network's interpolated station-based observation data records were originally used to create this collection (Fan & Van Den Dool 2008). CPC NOAA is now accessible at 0.5° resolutions from 1948 to the present day and is obtained from over 30,000 gauged data provided by various national and international institutions.

ERA5 or ECWMF Reanalysis v5 is the replacement for ERA-Interim reanalysis. It is an hourly estimate of atmospheric, land, and oceanic climate variables produced by the Copernicus Climate Change Service (C3S). ERA5 uses modeling and data assimilation tools to aggregate enormous volumes of historical information into global estimations. The dataset is available at 0.25° resolutions from 1959 to the present day (Hersbach et al. 2020).

PGF-V3 is a hybrid of observational and reanalysis datasets created by integrating NCEP/NCAR reanalysis data and global observation-based datasets. This dataset, developed by the Department of Civil Engineering, Princeton University, is available at 0.25° resolutions from 1948 to 2016 (Sheffield et al. 2006).

SA-OBS is a high spatial resolution daily gridded data that covers the Southeast Asian region and is based on the station data collated by the Southeast Asia Climate Assessment and Dataset (SACA&D). Meteorological stations throughout Southeast Asia supply these data. The dataset is available at 0.25° and 0.5° resolutions and spans the period 1981 to 2017 (Van Den Besselaar et al. 2017).

## Data Processing

The bilinear interpolation method through Climate Data Operators (CDO) was used to extract data to the location of the ground-based meteorological station. This method is recommended for grid location-based continuous datasets, by calculating the point value using the distance-weighted value of the four nearest points (Schulzweida et al. 2012).

Linear Scaling was applied as a bias correction method (Shrestha 2015). The mean monthly values of the corrected and observed dataset are perfectly matched. The monthly correction factor is developed based on the differences between observed and raw data. The correction factor for rainfall is developed as a multiplier, while the temperature correction factor is an additive.

$$P_{cor,m,d} = P_{raw,m,d}\ x\ \frac{\mu(P_{obs,m})}{\mu(P_{raw,m})} \quad \ldots(1)$$

$$T_{cor,m,d} = T_{raw,m,d} + [\mu(T_{obs,m}) - \mu(T_{raw,m})] \quad \ldots(2)$$

Where $P_{cor,m,d}$ and $T_{cor,m,d}$ are corrected rainfall and temperature on the *dth* day of *mth* month, and $P_{raw,m,d}$ and $T_{raw,m,d}$ are the raw rainfall and temperature on the *dth* day of *mth* month. $\mu$ represents the correction factor (represents the mean value of observed rainfall at a given month *m*) (Fang et al. 2015).

## Performance Evaluation of Gridded Rainfall and Temperature Datasets

The reliabilities of the rainfall and temperature datasets were evaluated by using the reference gauge method, where the grid climate datasets were directly compared with the observation at the reference gauge (Tian et al. 2021). Statistical and graphical evaluation was conducted in monthly and annual timescales from 1981-2000 for rainfall, and 1995-2000 for air temperature. Data validation for both datasets was done for the period 2001-2005.

Statistical evaluation indices used to assess the reliability of the gridded dataset were based on the evaluation by Legates & McCabe (1999) and Moriasi et al. (2007): Pearson's correlation coefficient ($r$), coefficient of determination ($R^2$) modified index of agreement ($d_1$), Kling-Gupta efficiency, Nash-Sutcliffe efficiency (NSE), and RMSE-observations

Table 1: Description of rainfall and temperature datasets.

| Parameter | Dataset* | Spatial Resolution | Temporal Resolution | Time Domain | Source Institution** |
|---|---|---|---|---|---|
| Rainfall | APHRODITE | 0.25° | Daily | 1951–2015 | RISH, Kyoto University |
| Rainfall & Temperature | CPC NOAA | 0.5° | Daily | 1979-Present | NCEP/CPC |
| | ERA5 | 0.25° | Hourly | 1979-Present | |
| | PGF-V3 | 0.25° | 3-hourly, daily, monthly | 1948-2016 | Princeton University |
| | SA-OBS | 0.25° | Daily | 1981–2017 | RNMI |

*APHRODITE: Asian Precipitation - Highly-Resolved Observational Data Integration Towards Evaluation, CPC: Climate Prediction Center, ERA5: ECWMF Reanalysis v5, PGF-V3: Princeton Global Forcing v3, SA-OBS: Southeast Asia-Observational

**RISH = Research Institute for Sustainable Humanosphere, NCEP = National Centers for Environmental Prediction, RNMI = Royal Netherlands Meteorological Institute

Table 2: Statistical evaluation indices and their optimal values.

| Index | Formula | Value Range | Optimal Value |
|---|---|---|---|
| r | $R=\frac{\sum(P_i-\overline{P})(O_i-\overline{O})}{\sqrt{\sum(P_i-\overline{P})^2\sum(O_i-\overline{O})^2}}$ | -1 to 1 | 1 |
| $R^2$ | $R^2=\left[\frac{\sum_{i=1}^{N}(O_i-\overline{O})(P_i-\overline{P})}{\sqrt{\sum_{i=1}^{N}(O_i-\overline{O})^2}\sqrt{\sum_{i=1}^{N}(P_i-\overline{P})^2}}\right]^2$ | 0 to 1 | 1 |
| $d_1$ | $d_1=1-\frac{\sum_{i=1}^{N}(O_i-P_i)}{\sum_{i=1}^{N}(\lvert P_i-\overline{O}\rvert+\lvert O_i-\overline{O}\rvert)}$ | 0 to 1 | 1 |
| NSE | $NSE=1-\left[\frac{\sum_{i=1}^{N}(O_i-P_i)^2}{\sum_{i=1}^{N}(O_i-\overline{O})^2}\right]$ | $-\infty$ to 1 | 1 |
| KGE | $KGE=\sqrt{(r-1)^2+(\frac{\sigma_P}{\sigma_O}-1)^2+(\frac{\overline{P}}{\overline{O}}-1)^2}$ | $-\infty$ to 1 | 1 |
| RSR | $RSR=\frac{\sqrt{\sum_{i=1}^{N}(O_i-P_i)^2}}{\sqrt{\sum_{i=1}^{N}(O_i-\bar{O})^2}}$ | 0 to $+\infty$ | 0 |

* $P_i$ and $O_i$ denote predicted and observed, respectively, daily rainfall or temperature of the $i^{th}$ day, $\overline{P}$ and $\overline{O}$ denote predicted and observed, respectively, mean daily rainfall or temperature, $\sigma_P$ and $\sigma_O$ denote predicted and observed, respectively, standard deviation for daily rainfall or temperature, and N is the amount of data at daily time series.

Table 3: General performance efficiency ratings for a monthly time step.

| Performance Rating | NSE | KGE | RSR |
|---|---|---|---|
| Very Good | $0.75 < NSE \leq 1.00$ | $0.75 < KGE \leq 1.00$ | $0.00 < RSR \leq 0.50$ |
| Good | $0.65 < NSE \leq 0.75$ | $0.50 < KGE \leq 0.75$ | $0.50 < RSR \leq 0.60$ |
| Satisfactory | $0.50 < NSE \leq 0.65$ | $0.00 < KGE \leq 0.50$ | $0.60 < RSR \leq 0.70$ |
| Unsatisfactory | $NSE \leq 0.50$ | $KGE \leq 0.00$ | $RSR > 0.70$ |

standard deviation ratio (RSR). Table 2 and Table 3 presents the optimal ranges and benchmarking categories (Moriasi et al. 2007) of the statistical indices.

Pearson's correlation coefficient ($r$) is a statistical test that measures the magnitude of association and direction of relationship between two continuous variables. The values of $r$ range from -1 to 1, with $r = 0$ having no linear relationship. The coefficient of determination ($R^2$) measures the variance of the grid climate dataset that can be explained by the observed climate data. It quantifies the strength of linear relationship between variables, thus, being called the *goodness of fit model*. The modified index of agreement ($d_1$) is the modified version of the index of agreement proposed by Willmott (1981) to detect additive and proportional differences in the data. It is used as a standardized measure of the degree of model prediction error (Ahmed et al. 2019, Willmott 1984). Higher values of $R^2$ and $d_1$, i.e., close to 1, indicate higher accuracy of gridded datasets in estimating rainfall and air temperature.

Nash-Sutcliffe efficiency (NSE) determines the relative magnitude of the residual variance compared to the measured variance. It is used to measure how well a simulation can predict the outcome variable (Nash & Sutcliffe 1970). Kling-Gupta Efficiency (KGE) compares observed data with the grid data, considering the three elements of model error of Nash-Sutcliffe efficiency (NSE), i.e., correlation, bias, the ratio of variances, or coefficients of variation (Gupta et al. 2009). The value of NSE and KGE ranges from $-\infty$ to its optimal value of 1.

Root-mean-square error (RMSE) evaluates the quality of predictions by showing how far the predicted value is from the measured true values. A model evaluation statistic by Singh et al. (2004) standardizes RSME values using the RMSE-observations standard deviation ratio (RSR). The value of RSR is determined by the ratio of the RMSE and standard deviation of measured data. The values vary from 0, which indicates a perfect model simulation of observed data, to large positive values. Lower RSR values indicate lower RMSE and better climate simulation performance.

## RESULTS AND DISCUSSION

### Rainfall

The grid-based climate products are evaluated and validated in monthly and annual timescales. The obtained time series are then compared along with the corresponding station-based time series. Scatter plots between the observed and grid rainfall at a monthly time scale, together with statistical indices ($r, R^2, d_1, NSE, KGE, RSR$), are shown in Fig. 3. Graphical examination shows an underestimation of grid rainfall datasets for the evaluation period and slight overestimation for the validation period. High values of $r$, $R^2$, and $d_1$ for both periods indicate a high positive correlation between gridded and observed rainfall datasets. Exceptional goodness-of-fit is particularly observed for SA-OBS, followed by PGF-V3 and APHRODITE.

The values of NSE, KGE, and RSR allow us to make explicit interpretations of gridded datasets in estimating rainfall and air temperature. Gridded rainfall products show exceptionally good performance in simulating observed rainfall datasets for the evaluation period. The same can be said for the validation period, except for ERA5 having satisfactory performance with values of NSE = 0.63, and RSR = 0.61.

Gridded rainfall dataset performance varies for the annual timescale presented in Fig. 2. SA-OBS and PGF-V3 showed the most acceptable association with the observed rainfall dataset with the lowest RSR values. Although. APH, CPC NOAA, and ERA5 showed good statistical performance when comparing monthly periods, underestimation in high rainfall years and overestimation in low rainfall years can be observed when comparing in annual periods.

### Maximum and Minimum Temperature

Scatter plots between the observed and grid maximum and minimum temperature at a monthly time scale, together with statistical indices ($r, R^2, d_1, NSE, KGE, RSR$), are shown in Fig. 4 and Fig. 5, respectively. Based on the correlation statistics on the monthly time period, SA-OBS showed exceptional results for both temperature parameters in the evaluation and validation period. The lowest correlation values for maximum temperature can be observed for CPC NOAA ($r$=0.89 $R^2$=0.79, $d_1$=0.75) and ERA5 for minimum temperature ($r$=0.78 $R^2$=0.66, $d_1$=0.66).

SA-OBS showcased exceptional accuracy and prediction performance for maximum temperature (NSE=0.99, KGE=0.97, RSR=0.12)and minimum temperature (NSE=0.98, KGE=0.95, RSR=0.14). CPC NOAA and ERA 5 provided the opposite for the same statistical indicators.

The mean annual maximum and minimum temperature variation of the gridded rainfall dataset with the observed temperature is presented in Fig. 6. All datasets follow the annual temperature trend of the observed data, showing significant overestimation starting in 1999 for maximum temperature and underestimation from the same year for minimum temperature.

SA-OBS dataset consistently showed exceptional performance having the highest correlation and prediction power compared to the other datasets and better overall fit to the observed data. Additionally, the SA-OBS dataset

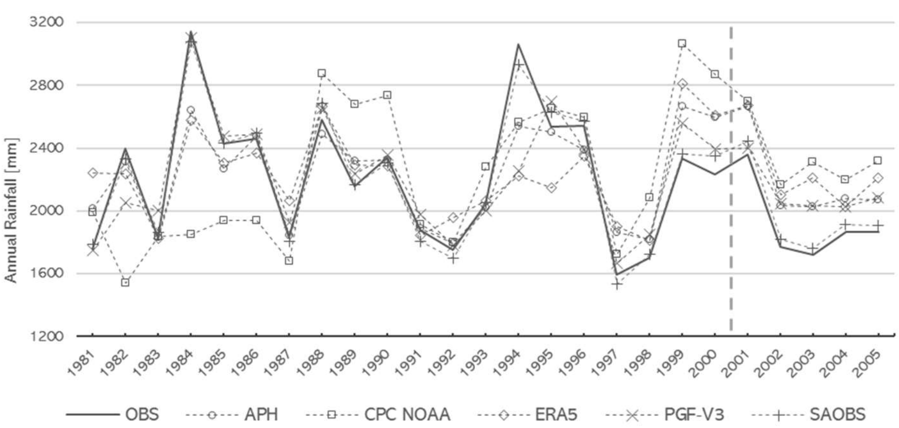
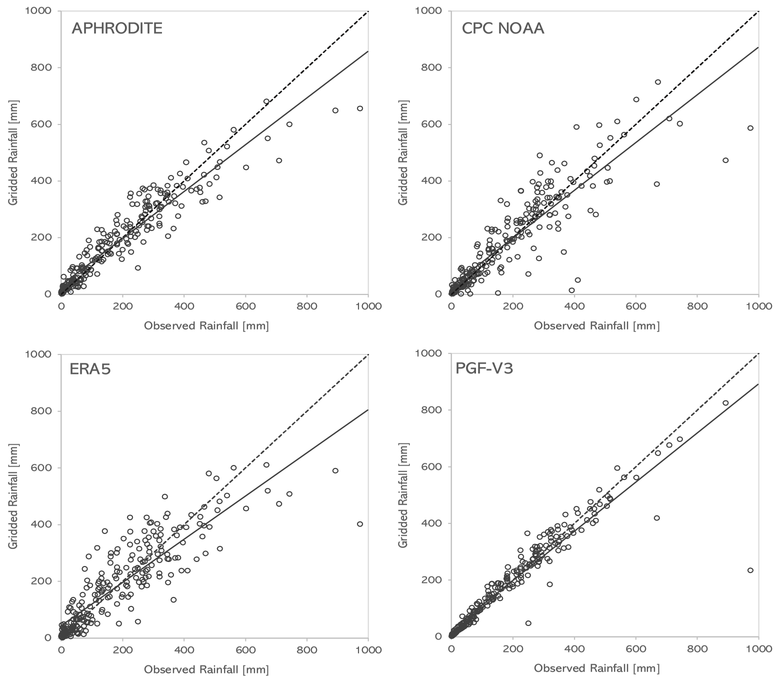
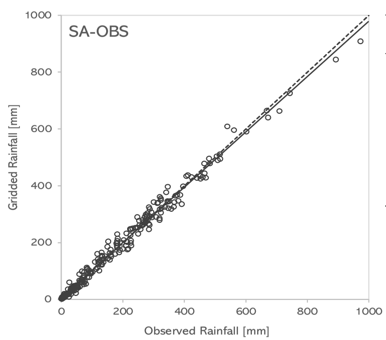


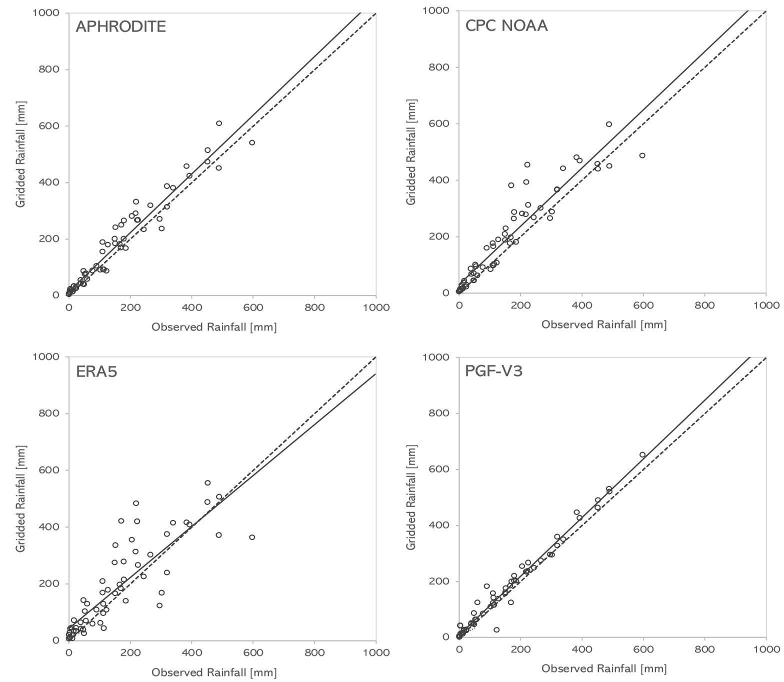

Fig. 2: Mean annual rainfall comparison for the period 1981-2005.

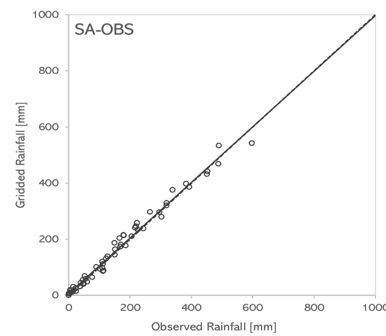



| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| APHRODITE | 0.95 | 0.89 | 0.86 | 0.89 | 0.86 | 0.33 |
| CPCNOAA | 0.89 | 0.80 | 0.83 | 0.79 | 0.88 | 0.46 |
| ERA5 | 0.87 | 0.76 | 0.79 | 0.76 | 0.82 | 0.49 |
| PGF-V3 | 0.94 | 0.89 | 0.91 | 0.89 | 0.90 | 0.34 |
| SA-OBS | 0.99 | 0.99 | 0.95 | 0.99 | 0.98 | 0.11 |

(a)

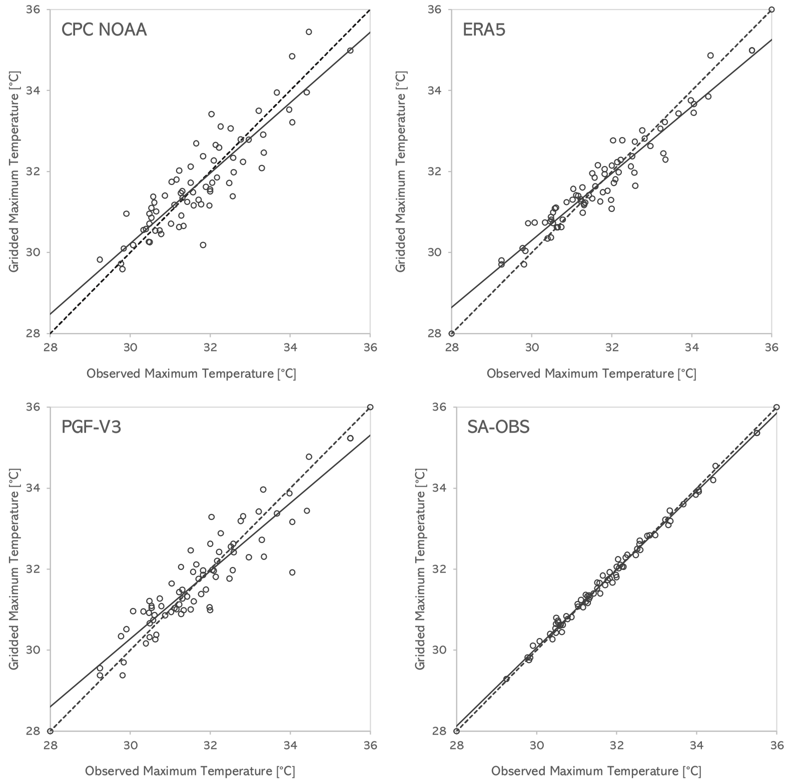



| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| APHRODITE | 0.97 | 0.94 | 0.86 | 0.91 | 0.84 | 0.30 |
| CPCNOAA | 0.94 | 0.88 | 0.82 | 0.79 | 0.75 | 0.46 |
| ERA5 | 0.85 | 0.71 | 0.75 | 0.63 | 0.76 | 0.61 |
| PGF-V3 | 0.99 | 0.97 | 0.91 | 0.95 | 0.88 | 0.21 |
| SA-OBS | 0.99 | 0.98 | 0.94 | 0.98 | 0.97 | 0.13 |

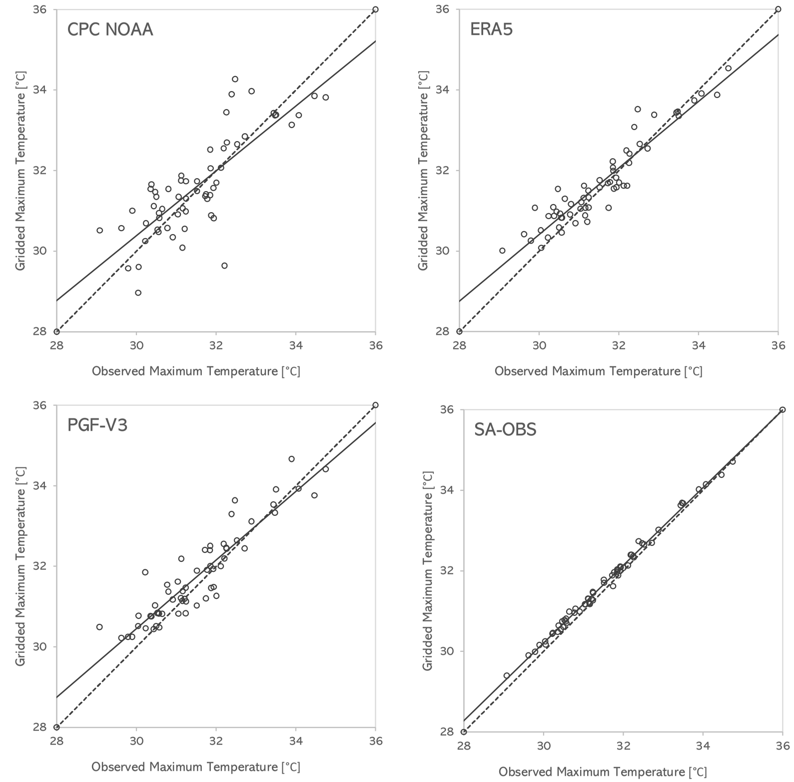
(b)

Fig. 3: Scatter plots and statistical indices of observed vs gridded daily rainfall dataset at the JRB for (a) the evaluation period of 1981 to 2000 and (b) the validation period 2001-2005. The dashed lines indicate a 1:1 line and solid lines denote linear regression.




| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| CPCNOAA | 0.89 | 0.79 | 0.75 | 0.78 | 0.89 | 0.59 |
| ERA5 | 0.95 | 0.92 | 0.82 | 0.90 | 0.86 | 0.41 |
| PGF-V3 | 0.90 | 0.83 | 0.78 | 0.81 | 0.87 | 0.56 |
| SA-OBS | 1.00 | 0.99 | 0.95 | 0.99 | 0.97 | 0.12 |

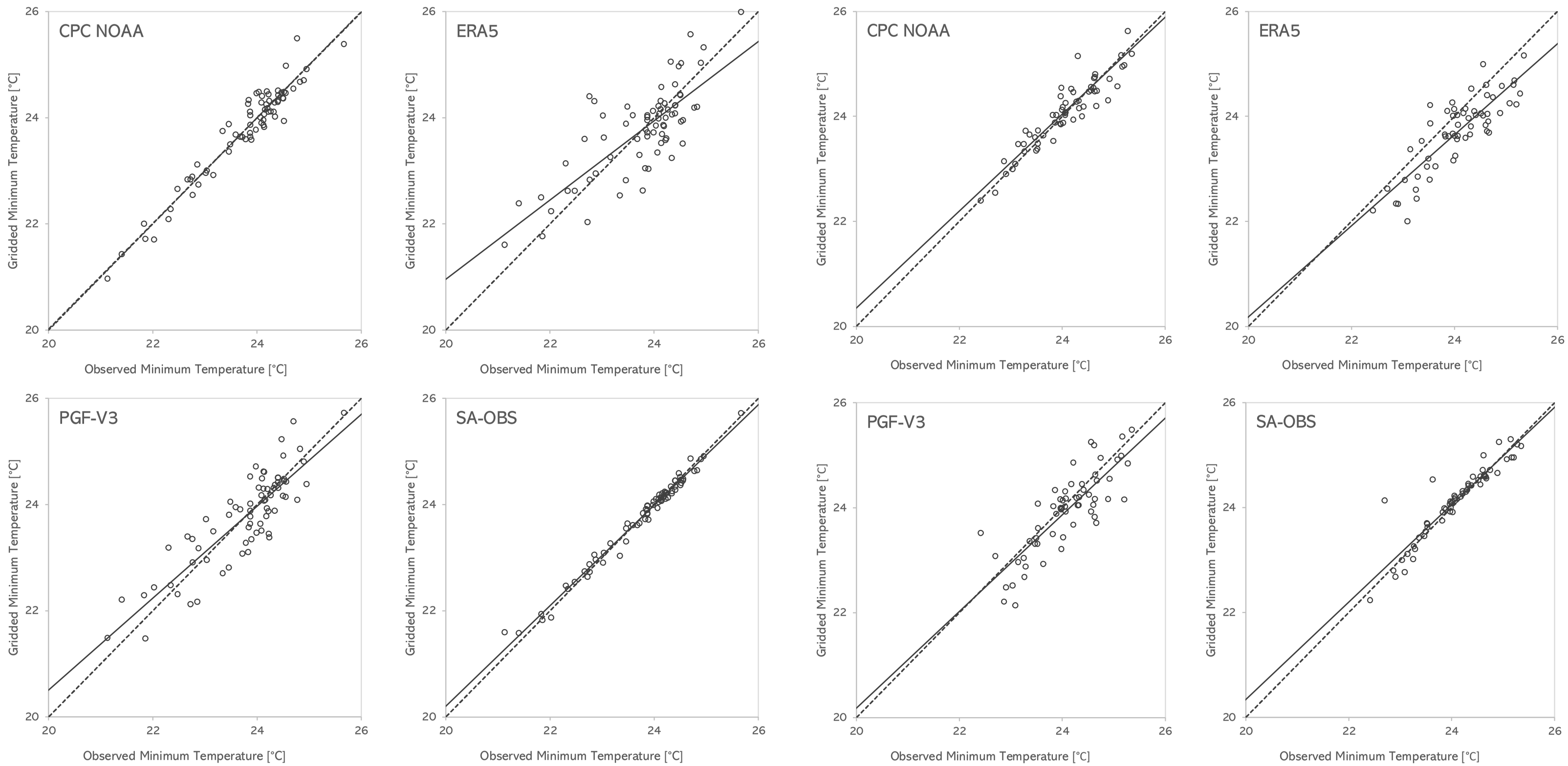
(a)




| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| CPCNOAA | 0.80 | 0.64 | 0.69 | 0.59 | 0.80 | 0.64 |
| ERA5 | 0.94 | 0.89 | 0.81 | 0.87 | 0.87 | 0.36 |
| PGF-V3 | 0.92 | 0.85 | 0.78 | 0.81 | 0.89 | 0.43 |
| SA-OBS | 1.00 | 1.00 | 0.91 | 0.98 | 0.97 | 0.15 |

(b)

Fig. 4: Scatter plots and statistical indices of observed versus gridded daily maximum temperature at the JRB for (a) the evaluation period of 1995 to 2000; and (b) the validation period 2001-2005. The dashed lines indicate a 1:1 line and solid lines denote linear regression.



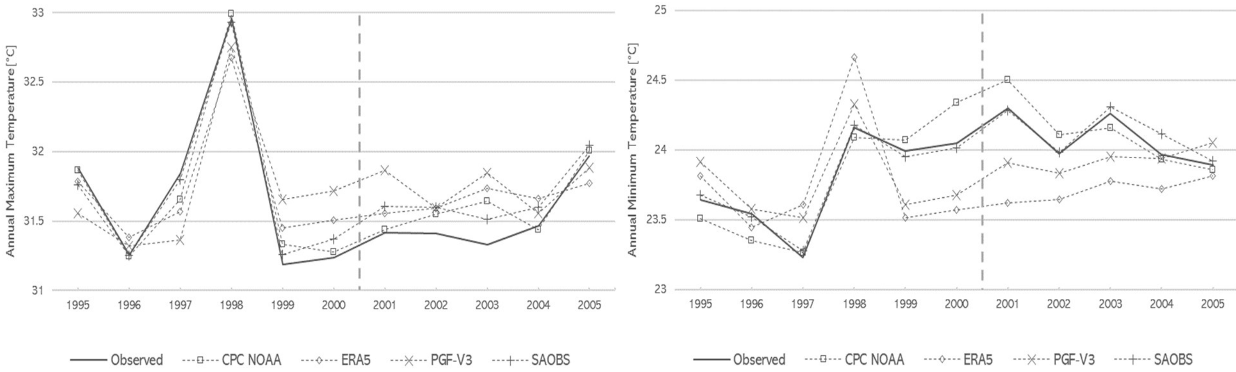

| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| CPCNOAA | 0.97 | 0.93 | 0.86 | 0.93 | 0.95 | 0.27 |
| ERA5 | 0.78 | 0.66 | 0.66 | 0.53 | 0.77 | 0.65 |
| PGF-V3 | 0.87 | 0.78 | 0.73 | 0.73 | 0.87 | 0.51 |
| SA-OBS | 0.99 | 0.99 | 0.93 | 0.98 | 0.95 | 0.14 |

(a)

| Statistical Index | $r$ | $R^2$ | $d_1$ | NSE | KGE | RSR |
|---|---|---|---|---|---|---|
| CPCNOAA | 0.91 | 0.82 | 0.81 | 0.80 | 0.90 | 0.44 |
| ERA5 | 0.85 | 0.71 | 0.60 | 0.40 | 0.84 | 0.77 |
| PGF-V3 | 0.83 | 0.68 | 0.69 | 0.55 | 0.79 | 0.67 |
| SA-OBS | 0.93 | 0.87 | 0.87 | 0.86 | 0.93 | 0.38 |

(b)

Fig. 5: Scatter plots and statistical indices of observed versus gridded daily minimum temperature at the JRB for (a) the evaluation period of 1995 to 2000; and (b) the validation period 2001-2005. The dashed lines indicate a 1:1 line and solid lines denote linear regression.




(a) (b)

Fig. 6: Mean annual (a) maximum and (b) minimum temperature comparison for the period 1995-2005.

has the lowest RMSE-observation standard deviation ratio, indicating a closer match to the variability of the observed data. Based on the performance efficiency rating presented in Table 3, SA-OBS rated "very good" in all statistical metrics. These results suggest that the SA-OBS dataset performs exceptionally well in capturing the rainfall and temperature patterns in the Jalaur River basin compared to the other gridded datasets.

## DISCUSSION

Evaluation and validation studies of gridded climate datasets in the Philippines have been conducted to augment the shortage of data availability in climate research. Salvacion et al. (2018) evaluated raw and downscaled CRU TS monthly gridded climate data in fifty-five meteorological stations throughout the country, concluding that both raw and downscaled gridded datasets showed acceptable performance with respect to rainfall data. Peralta et al. (2020) confirmed four (4) high-resolution gridded rainfall datasets, including APHRODITE, CHIRPSv2, PERSIAN_CDR, and TRMM, that represent the rainfall patterns in the country. Among these datasets, TRMM most accurately replicates the frequency and proportion of rainfall events spatially and temporally. The same dataset was used by Corporal-Lodangco & Leslie (2017) to define the climate zones in the country. However, a countrywide approach is conducted in these studies. Lee & Ahn (2022) recommend a more regional approach in the evaluation of meteorological datasets for future use in regional hydrological models.

A basin-level approach evaluated and validated five grid-based climate products for the Jalaur River Basin in this study. Results show an underestimation of grid rainfall datasets during the evaluation period and a slight overestimation during the validation period. Exceptional goodness-of-fit is observed for SA-OBS, PGF-V3, and APHRODITE. Except for the ERA5 dataset, gridded rainfall products show good performance in simulating mean monthly observed rainfall datasets. Annual performance varies, with SA-OBS and PGF-V3 showing the most acceptable association with observed rainfall.

The SA-OBS dataset demonstrated exceptional performance in capturing rainfall and temperature patterns, which is consistent with the findings of Ge et al. (2019), where SA-OBS captured spatial distributions and patterns in the Indochina Peninsula, and Ang et al. (2022), where SA-OBS dataset displayed better estimates of seasonal patterns and magnitudes for both minimum and maximum temperatures in the Tonle Slap Basin in Cambodia. Van den Besselaar et al. (2017) observed a similar exceptional simulation of SA-OBS-based rainfall. Compared to other datasets, SA-OBS has been considered one of the more accurate rainfall and temperature estimations in Southeast Asia.

ERA5 rainfall data prove to be valuable in accurately capturing rainfall patterns in Central and East Asia; however, it is highly influenced by topography (Xin et al. 2022, Jiao et al. 2021, Zandler et al. 2020). Lavers (2022) evaluated ERA5 rainfall in the European region, indicating its good overall performance but highlighting the potential underestimation of extreme rainfall events at the daily timescale.

APHRODITE and CPC NOAA underestimated the values of rainfall data for the JRB. Ang et al. (2022) noted an identical tendency of APHRODITE data to underestimate rainfall in the Tonle Sap Lake Basin of Cambodia. Saidah et al. (2019) indicated similar findings for CPC NOAA on the Island of Lombok in Indonesia, further stating that CPC NOAA has poor prediction performance and needs to be calibrated before utilization.

It is important to acknowledge that no gridded dataset can perfectly replicate the actual observation datasets. The choice of an appropriate dataset for a specific study should be guided by its capability to accurately represent the specific attributes of the data necessary for particular research.

## CONCLUSION

This research paper has addressed the critical issue of obtaining accurate and reliable high-resolution climate data for the ungauged Jalaur River Basin in the Philippines. Due to geographical constraints and limited meteorological observations, the study explored the use of gridded rainfall and temperature datasets as alternative sources of climate data. The findings are crucial for various fields, including hydrology, climate modeling, and environmental impact assessments, to better understand the hydrological characteristics of this region and its response to changing climate conditions.

The study evaluated five rainfall datasets (APHRODITE, CPC NOAA, ERA5, SA-OBS, and PGF-V3) and four temperature datasets (CPC NOAA, ERA5, SA-OBS, and PGF-V3) by comparing them with ground-based observations from a single station within the Jalaur River Basin. The statistical analysis employed various indices, including Pearson's correlation coefficient, coefficient of determination, modified index of agreement, Kling-Gupta efficiency, Nash-Sutcliffe efficiency, and RMSE-observations standard deviation ratio. These evaluations were conducted on monthly and annual timescales.

Exceptional goodness-of-fit was observed for SA-OBS, PGF-V3, and APHRODITE in terms of rainfall data, indicating a high positive correlation and accuracy in simulating observed rainfall, especially on a monthly timescale. ERA5 showed satisfactory performance with rainfall data on an annual timescale but was less accurate in capturing high and low rainfall years. For temperature data, SA-OBS consistently demonstrated exceptional performance, with the highest correlation, prediction power, and the lowest RSR values. It provided the best overall fit to the observed data for both maximum and minimum temperatures. CPC NOAA and ERA5 performed less favorably in replicating temperature patterns, particularly maximum temperature, suggesting potential challenges with these datasets in this region.

The results suggest that the SA-OBS dataset is a reliable and accurate source of climate data for this region and is well-suited for further research and analysis. The paper also emphasized that no gridded dataset can perfectly replicate observed data, and the choice of an appropriate dataset should be guided by the specific attributes required for a given research application.

The findings of this study have significant implications for water resource management and climate research in the Jalaur River Basin. Understanding the reliability of gridded climate datasets is essential for assessing the basin's hydrological response to climate and anthropogenic changes. The study also emphasizes the importance of installing weather stations to improve data collection, especially in major water resources.

In summary, this research contributes to the knowledge and methodology of assessing gridded climate dataset's performance in data-sparse regions, providing valuable insights for future hydrological modeling, climate impact studies, and water resource management efforts in the Jalaur River Basin and similar areas with limited meteorological observations.

## ACKNOWLEDGMENT

The authors would like to thank the Philippine Atmospheric, Geophysical, and Astronomical Services Administration for providing the gauged climate data. This research is supported by Central Philippine University and the Department of Science and Technology – Engineering Research and Development for Technology Program of the University of San Carlos.

## REFERENCES

Ahmed, K., Shahid, S., Wang, X., Nawaz, N. and Khan, N. 2019. Evaluation of Gridded Precipitation Datasets Over Arid Regions of Pakistan. Water, 11(2): 210.

Ang, R., Kinouchi, T. and Zhao, W. 2022. Evaluation of Daily Gridded Meteorological Datasets For Hydrological Modeling in Data-Sparse Basins of the Largest Lake in Southeast Asia. J. Hydrol. Reg. Stud., 42: 101135.

Arceo, M.G.A., Cruz, R.V., Tiburan, C., Jr., Balatibat, J. and Alibuyog, N. 2018. Modeling the Hydrologic Responses to Land Cover and Climate Changes of Selected Watersheds in the Philippines Using the Soil and Water Assessment Tool (Swat) Model. Dlsu Bus. Econ. Rev., 28: 84-101.

Ayoub, A. B., Tangang, F., Juneng, L., Tan, M. L. and Chung, J. X. 2020. Evaluation of Gridded Precipitation Datasets In Malaysia. Remote Sens., 12(4): 613.

Corporal-Lodangco, I.L. and Leslie, L.M. 2017. Defining Philippine Climate Zones Using Surface and High-Resolution Satellite Data. Proceed. Comput. Sci., 114: 324-332.

Fan, Y. and Van Den Dool, H. 2008. A Global Monthly Land Surface Air Temperature Analysis For 1948–Present. J. Geophys. Res., 113(D1): 56

Fang, G., Jian, Y., Chen, Y. and Zammit, C. 2015. Comparing Bias Correction Methods in Downscaling Meteorological Variables for A Hydrologic Impact Study in an Arid Area in China. Hydrol. Earth Syst. Sci., 19(6): 2547-2559.

Ge, F., Peng, T., Fraedrich, K., Sielmann, F., Zhu, X., Zhi, X., Liu, X., Wei-Wei, T. and Zhao, P. 2018. Assessment of Trends and Variability in Surface Air Temperature on Multiple High-Resolution Datasets Over the Indochina Peninsula. Theor. Appl. Climatol., 135(3-4): 1609-1627.

Gupta, H.V., Kling, H., Yilmaz, K.K. and Martínez, G. 2000. Decomposition of the Mean Squared Error and NSE Performance Criteria: Implications For Improving Hydrological Modeling. J. Hydrol., 377(1-2): 80-91.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, Á., Muñoz Sabater, J., Nicolas, J. P., Peubey, C., Radu, R., Schepers, D., Simmons, A. J., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M. and Thépaut, J. 2020. The Era5 Global Reanalysis. Q. J. R. Meteorol. Soc., 146(730): 1999-2049.

Ang, R., Kinouchi, T. and Zhao, W. 2022. Evaluation of Daily Gridded Meteorological Datasets for Hydrological Modeling in Data-Sparse Basins of the Largest Lake in Southeast Asia. J. Hydrol. Reg. Stud., 42: 101135.

Jiao, D., Xu, N., Yang, F. and Xu, K. 2021. Evaluation of Spatial-Temporal Variation Performance of Era5 Precipitation Data In China. Sci. Rep., 11(1): Https://Doi.Org/10.1038/S41598-021-97432-Y

Lavers, D.A., Simmons, A. J., Vamborg, F. and Rodwell, M. J. 2022. An Evaluation of Era5 Precipitation for Climate Monitoring. Q. J. R. Meteorol. Soc., 148(748): 3152-3165. Https://Doi.Org/10.1002/Qj.4351

Lee, D.G. and Ahn, K. H. 2022. Assessment of Suitable Gridded Climate Datasets for Large-Scale Hydrological Modeling Over South Korea. Remote Sens., 14(15): 3535. Https://Doi.Org/10.3390/Rs14153535

Legates, D.R. and Mccabe, G. J. 1999. Evaluating The use of "Goodness-of-Fit" Measures in Hydrologic and Hydroclimatic Model Validation. Water Resour. Res., 35(1): 233-241. Https://Doi.Org/10.1029/1998wr900018

Merino, A., García Ortega, E., Navarro, A., Fernández González, S., Tapiador, F. J. and Sánchez, J.L. 2021. Evaluation of Gridded Rain Gauge Based Precipitation Datasets: Impact of Station Density, Spatial Resolution, Altitude Gradient and Climate. Int. J. Climatol., 41(5): 3027-3043. Https://Doi.Org/10.1002/Joc.7003

Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L. 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. Asabe, 50(3): 885-900. Https://Doi.Org/10.13031/2013.23153

Nash, J. and Sutcliffe, J.V. 1970. River Flow Forecasting Through Conceptual Models Part I: A Discussion of Principles. J. Hydrol., 10(3): 282-290. Https://Doi.Org/10.1016/0022-1694(70)90255-6

Nassaj, B.N., Zohrabi, N., Shahbazi, A. and Fathian, H. 2022. Evaluating The Performance Of Eight Global Gridded Precipitation Datasets Across Iran. Dyn. Atmos. Oceans, 98: 101297. Https://Doi.Org/10.1016/J.Dynatmoce.2022.101297

Peralta, J.C., Narisma, G. And Cruz, F. 2020. Validation Of High-Resolution Gridded Rainfall Datasets For Climate Applications in the Philippines. J. Hydrometeorol., 21(7): 1571-1587. Https://Doi.Org/10.1175/Jhm-D-19-0276.1

Saidah, H., Saptaningtyas, R.S., Hanifah, L., Negara, I.J. and Widyanty, D. 2020. Noaa Satellite Performance in Estimating Rainfall Over the Island of Lombok. Iop Conf. Ser., 437(1): 012035. Https://Doi.Org/10.1088/1755-1315/437/1/012035

Salvacion, A.R., Magcale-Macandog, D.B., Cruz, P.C.S., Saludes, R.B., Pangga, I.B. and Cumagun, C.J.R. 2018. Evaluation and Spatial Downscaling of Cru Ts Precipitation Data in the Philippines. Model. Earth Syst. Environ., 4(3): 891-898. Https://Doi.Org/10.1007/S40808-018-0477-2

Schulzweida, U.S., Kornblueh, L.K. and Quast, R.Q. 2012. Cdo User's Guide: Climate Data Operators Version 1.5.9. Max-Planck-Institute For Meteorology, Hamburg, Germany.

Schumacher, V., Justino, F., Fernández, A., Meseguer-Ruiz, Ó., Sarricolea, P., Comin, A.N., Venâncio, L.P. and Althoff, D. 2020. Comparison Between Observations and Gridded Data Sets over Complex Terrain in the Chilean Andes: Precipitation and Temperature. Int. J. Climatol., 40(12): 5266-5288. Https://Doi.Org/10.1002/Joc.6518

Sheffield, J., Goteti, G. And Wood, E.F. 2006. Development Of A 50-Year High-Resolution Global Dataset Of Meteorological Forcings For Land Surface Modeling. J. Clim., 19(13): 3088–3111. Https://Doi.Org/10.1175/Jcli3790.1

Shrestha, M.S. 2015. Linear Scaling Bias Correction. Int. Cent. Integr. Mountain Dev., 6: 41-63.

Singh, J.S., Knapp, H.V.K. And Demissie, M.D. 2004. Hydrologic Modeling Of The Iroquois River Watershed Using Hspf And Swat. Illinois State Water Survey Contract Report 2004-08. Isws, Illinois, Us.

Tian, W., Liu, X., Wang, K., Bai, P., Liang, K. And Liu, C. 2021. Evaluation Of Six Precipitation Products In The Mekong River Basin. Atmos. Res., 255: 105539. Https://Doi.Org/10.1016/J.Atmosres.2021.105539

Try, S., Tanaka, S., Tanaka, K., Sayama, T., Oeurng, C., Sovannara, U., Takara, K., Hu, M. And Han, D. 2020. Comparison Of Gridded Precipitation Datasets For Rainfall-Runoff And Inundation Modeling In The Mekong River Basin. Plos One, 15(1): E0226814. Https://Doi.Org/10.1371/Journal.Pone.0226814

Van Den Besselaar, E., Van Der Schrier, G., Cornes, R., Iqbal, A. S. And Tank, A.K. 2017. Sa-Obs: A Daily Gridded Surface Temperature And Precipitation Dataset For Southeast Asia. J. Clim., 30(14): 5151–5165. Https://Doi.Org/10.1175/Jcli-D-16-0575.1

Willmott, C.J. 1981. On The Validation Of Models. Phys. Geogr., 2(2), 184–194. Https://Doi.Org/10.1080/02723646.1981.10642213

Willmott, C.J. 1984. On The Evaluation Of Model Performance In Physical Geography. Springer, Singapore, Pp. 443–460. Https://Doi.Org/10.1007/978-94-017-3048-8_23

Xin, Y., Yang, Y., Chen, X., Yue, X., Liu, Y. And Yin, C. 2022. Evaluation Of Imerg And Era5 Precipitation Products Over The Mongolian Plateau. Sci. Rep., 12(1). Https://Doi.Org/10.1038/S41598-022-26047-8

Yatagai, A., Kamiguchi, K., Arakawa, O., Hamada, A., Yasutomi, N. And Kitoh, A. 2012. Aphrodite: Constructing A Long-Term Daily Gridded Precipitation Dataset For Asia Based On A Dense Network Of Rain Gauges. Bull. Am. Meteorol. Soc., 93(9), 1401–1415. Https://Doi.Org/10.1175/Bams-D-11-00122.1

Zandler, H., Senftl, T. And Vanselow, K. A. 2020. Reanalysis Datasets Outperform Other Gridded Climate Products In Vegetation Change Analysis In Peripheral Conservation Areas Of Central Asia. Sci. Rep., 10(1). Https://Doi.Org/10.1038/S41598-020-79480-Y

## ORCID DETAILS OF THE AUTHORS

Christsam Joy S. Jaspe-Santander: https://orcid.org/0000-0003-3240-7780
Ian Dominic F. Tabañag: https://orcid.org/0000-0002-4049-1073