Recibido: 20/12/2021
Aceptado: 06/07/2022

REVISTA DE TELEDETECCIÓN
Asociación Española de Teledetección
(2022) 60, 17-28
ISSN 1133-0953
EISSN 1988-8740
https://doi.org/10.4995/raet.2022.16907

Improved rainfall and temperature satellite dataset in areas with scarce weather stations data: case study in Ancash, Peru

Eduardo E. Villavicencio [*1], Katy D. Medina [1,2], Edwin A. Loarte [1,2], Hairo A. León [1]

[1] *Faculty of Environmental Science (FCAM), Santiago Antunez de Mayolo National University (UNASAM), 200 Centenario avenue, Huaraz, Peru.*

[2] *National Institute for Research on Glaciers and Mountain (INAIGEM), 887 Juan Bautista Mejía street, Huaraz, Peru.*

Abstract: Rainfall and temperature variables play an important role in understanding meteorology at global and regional scales. However, the availability of meteorological information in areas of complex topography is difficult, as the density of weather stations is often very low. In this study, we focused on improving existing satellite products for these areas, using Tropical Rainfall Measuring Mission (TRMM) and Global Precipitation Measurement (GPM) data for rainfall and Modern Era Retrospective Analysis for Research and Applications Version 2 (MERRA-2) data for air temperature. Our objective was to propose a model that improves the accuracy and correlation of satellite data with observed data on a monthly scale during 2012-2017. The improvement of rainfall satellite data was performed using 4 regions: region 1 Santa (R1Sn), region 2 Marañón (R2Mr), region 3 Pativilca (R3Pt) and region 4 Pacific (R4Pc). For temperature, a model based on the use of the slope obtained between temperature and altitude data was used. In addition, the reliability of the TRMM, GPM and MERRA-2 data was analyzed based on the ratio of the mean square error, PBIAS, Nash-Sutcliffe efficiency (NSE) and correlation coefficient. The final products obtained from the model for temperature are reliable with R^2 ranging from 0.72 to 0.95 for the months of February and August respectively, while the improved rainfall products obtained are shown to be acceptable (NSE≥0.6) for the regions R1Sn, R2Mr and R3Pt. However, in R4Pc it is unacceptable (NSE<0.4), reflecting that the additive model is not suitable in regions with low rainfall values.

Key words: TRMM, GPM, MERRA-2, weather stations, Ancash.

Mejora de los datos satelitales de precipitación y temperatura en áreas con baja disponibilidad de estaciones meteorológicas: caso de estudio en Ancash, Perú

Resumen: Las variables de precipitación y temperatura desempeñan un papel importante en la comprensión de la meteorología a escala global y regional. Sin embargo, disponer de información meteorológica en zonas de topografía compleja es difícil, ya que la densidad de estaciones meteorológicas suele ser muy baja. En este estudio, nos centramos en mejorar los productos satelitales existentes para estas zonas, empleando datos de la *Tropical Rainfall Measuring Mission* (TRMM) y *Global Precipitation Measurement* (GPM) para la precipitación y los datos *Modern Era Retrospective Analysis for Research and Applications* Version 2 (MERRA-2) para la temperatura. Nuestro objetivo fue proponer un modelo que mejore la precisión y la correlación de los datos satelitales con los datos observados a escala mensual durante el 2012-2017. La mejora de los datos satelitales de precipitación se

To cite this article: Villavicencio, E.E., Medina, K.D., Loarte, E.A., León, H.A. 2022. Improved rainfall and temperature satellite dataset in areas with scarce weather stations data: case study in Ancash, Peru. *Revista de Teledetección*, *60*, 17-28. https://doi.org/10.4995/raet.2021.16907

* *Corresponding author: evillavicenciog@unasam.edu.pe*

realizó utilizando 4 regiones: región 1 Santa (R1Sn), región 2 Marañón (R2Mr), región 3 Pativilca (R3Pt) y región 4 Pacífico (R4Pc). En la temperatura se utilizó un modelo basado en el uso de la pendiente obtenida entre los datos de temperatura y altitud. Además, se analizó la fiabilidad de los datos TRMM, GPM y el MERRA-2 basándose en la relación del error cuadrático medio, PBIAS, la eficiencia de Nash-Sutcliffe (NSE) y el coeficiente de correlación. Los productos finales obtenidos del modelo para la temperatura son fiables, con R^2 entre 0,72 y 0,95 para los meses de febrero y agosto, respectivamente, mientras que los productos mejorados de precipitación obtenidos son aceptables (NSE≥0,6) para las regiones R1Sn, R2Mr y R3Pt. Sin embargo, en R4Pc es inaceptable (NSE<0,4), lo que refleja que el modelo aditivo empleado no es adecuado para regiones con bajos valores de precipitación.

Palabras clave: TRMM, GPM, MERRA-2, estaciones meteorológicas, Ancash.

1. Introduction

Climate in mountain systems is often considerably complex due to the existence of microclimatic features, topographic gradients and the influence of different atmospheric circulation patterns (Vicente-Serrano et al., 2017). Several studies have described that these characteristics generate a strong uncertainty in temperature and rainfall values associated with the complexity that exists in the climate and terrain (Beniston et al., 1997; Condom et al., 2011; Garreaud et al., 2003; Vicente-Serrano et al., 2017). Associated with this, there is also limited availability of meteorological data in these areas, since most weather stations are located below 3000 m a.s.l., as is the case in the department of Ancash (Peru). Therefore, the need to have a continuous record of information that adequately represents the characteristics of the climate in these gray zones has prompted the task of improving the spatial representation models of different climatic variables. Currently, several spatialized models obtained from remote sensors are available, being the TRMM (1997-2014) and GPM (2014-present) mission the most widely used for the spatial representation of rainfall and which stand out for not presenting range problems or regional sensitivity variations and which are able to generate better rainfall descriptions than ground-based radars (Ouatiki et al., 2017). In high mountain areas TRMM and GPM products have proven to be more accurate than other satellite products. For example, in the mountainous regions of Nepal it was found that TRMM obtained a better capture of the dependence that exists between mean rainfall and elevation, a behavior that was found in data from weather stations in the area (Krakauer et al., 2013). While in the Cordillera Blanca (mountainous area of Ancash), Mourre et al. (2016) founded that TRMM data have a correct representation of the spatial patterns of rainfall at the annual scale. However, Condom et al. (2011) found that at the monthly level it is necessary to adjust TRMM rainfall values with data from weather stations through an additive model. Regarding temperature, the MERRA-2 project aimed to reanalyze data obtained by different satellites, obtaining data from 1980 to the present on a fixed grid ranging from a surface pressure of 1000 hPa to the top of the model of 0.1 hPa (McCarty et al., 2016; Gelaro et al., 2017). Although these data are useful, many of them may present erroneous measurements due to the spatial scale or to a malfunction of the algorithm used to obtain the temperature (Vicente-Serrano et al., 2017).

Therefore, a few years ago Aybar et al. (2017) developed a spatialized product of temperature and rainfall for all Peru using satellite and weather station data. This gridded product is known as Peruvian Data Interpolated from SENAMHI Climatological and Hydrological Observations (PISCO) and was built based on CHIRPS (Climate Hazards Group InfraRed Rainfall with Station data) satellite data. However, the spatial scale of the work (5×5 km grid) may generate uncertainties in smaller areas that are already beginning to suffer certain conditions at the local level that may generate a variation in the values obtained due to the fact that the spatio-temporal prediction of rainfall will have greater reliability in places where there is more station information (Fernández-Palomino et al., 2022), which may increase the uncertainty values in mountainous areas where the density of stations is very low. Likewise, the temperature values were adjusted

to the topography using an elevation model with 5 km of spatial resolution, which makes the altitudinal changes less noticeable and therefore the temperature prediction in areas with more rugged topography less accurate (Vicente-Serrano et al., 2017).

In this study, we present two improved monthly-scale products for rainfall and temperature in the department of Ancash, which presents a high climatic and topographic complexity (see study area section), during the period 2012-2017. We relied on the TRMM and GPM satellite dataset for rainfall and MERRA-2 for temperature. In addition, observations from all available weather stations were used. The objectives of the study were (1) to develop two models for the correction of TRMM, GPM and MERRA-2 data and (2) obtain two improved products reflecting the spatial and temporal variation of rainfall and temperature in the study area.

2. Study area

The study area is located in the northern sector of the Peruvian Andes and is delimited by the departmental limit of Ancash that goes from 76°43' to 78°39' W and between 08°02' and 10°47' S, which includes the Cordillera Blanca, Santa River basin, Pativilca river basin, Marañón river basin and the Pacific coastal zone (Figure 1). Its altitudinal level varies from 0 in the coastal zone to 6757 m a.s.l. in the cordillera. The highlands are characterized by a semi-dry and semi-cold climate, while the coast has a very warm climate with a high humidity content. These characteristics contribute to the climatic complexity of the area, which is mainly controlled by the Andes, as the orography acts as a topographic barrier to the flow of humidity, causing the formation of strong pluviometric gradients on the eastern flanks of the Andes. To address this complexity in rainfall, the study area was divided into 4 regions: region 1 Santa (R1Sn) is characterized by an annual rainfall of more than 400 mm and is mainly dominated by convective processes. Region 2 Marañón (R2Mr) presents values higher than 500 mm due to its proximity to the rainforest. Region 3 Pativilca (R3Pt) reaches values between 200 to 600 mm; and region 4 Pacífico (R4Pc) presents low rainfall due to the influence of cold and dry air masses coming from the Humboldt Current System, causing drier conditions on the coast (<200 mm, Aybar et al., 2019). For its part, temperature was regionalized using the entire polygon of the study area, using the correlation with altitude. This relationship, known as altitudinal gradient, shows mean annual temperatures of ~23 °C at the lowest elevations and ~7 °C above 4000 m a.s.l., presenting a rate of change in temperature of -0.007 °C/m (Motschmann et al., 2020).

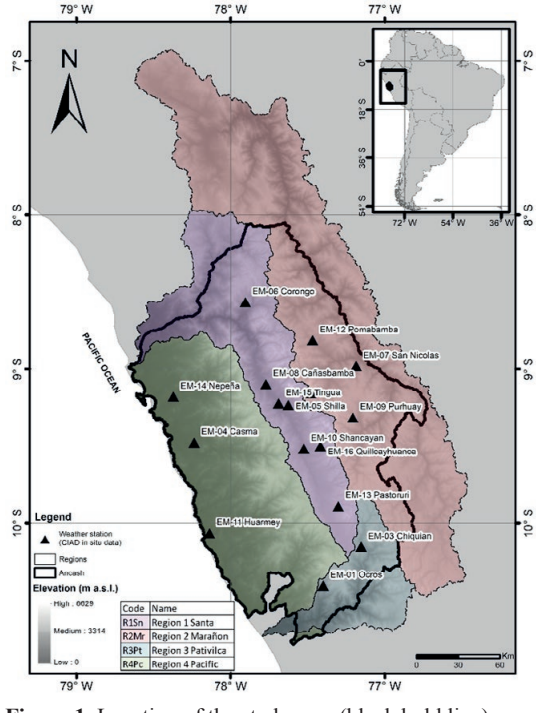

Figure 1. Location of the study area (black bold line), spatial distribution of the weather stations and the four regions delimited in our study.

3. Data

3.1. TRMM and GPM

In this study, TRMM product 3B43 with a spatial resolution of 0.25°×0.25° at monthly scale and covering between 50°N and 50°S was used. The 3B43 product provides a good estimate of monthly rainfall because it combines multiple satellite data from the Global Precipitation Climatology Center (GPCC) (Lu et al., 2018). These multiple satellite data come from passive microwave (PMW) sensors and infrared (IR)-based observations. The IR and PMW data are obtained every 3 hours, then summed for the calendar month, and then the rain gauge data are used to apply a large-scale

bias adjustment to the multi-satellite estimates (Huffman et al., 2007). While the GPM products show an improvement over its predecessor, as high frequency channels (165.6 and 183.3 GHz) are added to the GMI, and a Ka-band (35.5 GHz) is added to the DPR. This increase in high-frequency channels improves the ability to detect solid rainfall. Therefore, the GPM DPR is more sensitive in measuring light rainfall and high-latitude snowfall, providing a better understanding of global water circulation (Liu & Zipser, 2015). For this study, the 3IMERGM product was used which is an hourly dataset with a spatial resolution of 0.1°×0.1° covering an area from 60°N to 60°S, then these data are combined with GPCC rain gauge data to obtain the monthly satellite product, similar to the 3B43 products.

3.2. MERRA-2 reanalysis

The MERRA-2 reanalysis covers the period of satellite Earth observation from 1980 to the present. MERRA-2 is downloaded through the Goddard Earth Science Data and Information Services Center (GES DISC). MERRA-2 datasets contain different geophysical variables, including air temperature and humidity on a fixed pressure grid consisting of 42 standard pressure levels from a surface pressure of 1000 hPa to the top of the model of 0.1 hPa (Gelaro et al., 2017). The spatial resolution of MERRA-2 varies between 0.6° and 0.5° in the range of longitude and latitude. MERRA-2 was improved by including data from the Cross Track Infrared Sounder (CrIS), the Advanced Technology Microwave Sounder (ATMS) aboard the Suomi-NPP satellite, the visible and infrared imager aboard the MeteoSat satellite, and other sensors. One of the most important improvements is the measurement of air temperature at different atmospheric pressure, which facilitates the study of this variable at different altitudinal levels. For more details of the improvements included in MERRA-2 and their applications see McCarty et al. (2016).

3.3. *In situ* air temperature and rainfall data

Hourly temperature and rainfall records for the 16 weather stations were provided by the Centro de Investigación Ambiental para el Desarrollo (CIAD). Rainfall data were grouped according to the location of the meteorological stations available for each of the 4 regions (R1Sn, R2Mr, R3Pt and R4Pc) and temperature data were grouped according to the polygon of the department of Ancash. The distribution of the stations, the regions, the SRTM digital elevation model and the Ancash department boundary are shown in Table 1.

In the analysis of the temperature and rainfall time series, missing or erroneous data were detected,

Table 1. Location of the weather stations and number of years between 2012 and 2017 with observations.

Weather station name	Region	Longitude (W)	Latitude (S)	Elevation (m a.s.l.)	T (°C/year)	Rainfall (mm/year)
EM-05 Shilla	R1Sn	77°37’	9°14’	3210	14.8	773
EM-06 Corongo	R1Sn	77°54’	8°33’	3207	11.8	509
EM-08 Cañasbamba	R1Sn	77°46’	9°05’	2367	19.0	423
EM-10 Shancayan	R1Sn	77°31’	9°30’	3135	15.4	686
EM-13 Pastoruri	R1Sn	77°18’	9°53’	4194	6.7	760
EM-15 Tingua	R1Sn	77°41’	9°13’	2614	18.0	626
EM-16 Quillcayhuanca	R1Sn	77°24’	9°29’	3830	9.0	945
EM-02 Chacas	R2Mr	77°28’	9°09’	3964	9.4	1046
EM-07 San Nicolas	R2Mr	77°11’	8°58’	2573	18.2	526
EM-09 Purhuay	R2Mr	77°12’	9°18’	3450	12.4	944
EM-12 Pomabamba	R2Mr	77°28’	8°48’	3024	15.4	1152
EM-01 Ocros	R3Pt	77°23’	10°24’	3514	12.3	286
EM-03 Chiquian	R3Pt	77°09’	10°09’	3367	13.4	620
EM-04 Casma	R4Pc	78°14’	9°28’	167	22.56	7
EM-11 Huarmey	R4Pc	78°08’	10°03’	12	21.13	11
EM-14 Nepeña	R4Pc	78°22’	9°10’	126	22.1	13

possibly due to mishandling of the data by field personnel or meteorological impacts on the measurements (Franchito et al., 2009). Some of the missing data were recovered, as the stations generate a report that is sent to an electronic platform in the cloud. Those missing data that were not recovered were also not completed, as we sought to capture the actual behavior of rainfall and temperature over the dates, while erroneous data were filtered out using a data quality control process (See section 4.1). Finally, it is important to note that the stations were installed in March 2012, so the first months of this year do not present very dense records of information.

4. Methodology

4.1. Quality control (QC) of temperature and rainfall *in situ* data

Daily maximum and minimum temperature values were calculated with the hourly temperature records, while daily accumulated values were calculated for rainfall. These daily data were then subjected to a QC to remove some extreme or anomalous values. This removal was performed by employing an upper and lower limit to the entire daily time series of temperature and rainfall, following the Equation (1) proposed by Zhang et al. (2018),

$$Lim = \bar{x} \pm 3 \cdot S \quad (1)$$

where *Lim* is the maximum and minimum limit for the temperature and rainfall time series, $\bar{x}$ is the mean and S is the standard deviation of the temperature or rainfall. The QC was applied using a script developed in Python and with these values the monthly cumulative rainfall and monthly mean temperature were calculated for each weather station.

4.2. Pre-processing TRMM, GPM and MERRA-2

The comparison between *in situ* and raster data (TRMM, GPM and MERRA-2) was performed by extracting the value of the cells of each raster to a vector composed of 16 points that represent the geolocation of the weather stations over the study area. Thus, a vector of points with monthly precipitation and temperature data was obtained from each raster data for the period 2012 to 2017. The TRMM and GPM data, being homologous, were aggregated into a single time series for comparison with the monthly *in situ* rainfall data, while the MERRA-2 values were converted to degrees Celsius for comparison with the *in situ* data. The comparison was possible because the data extracted from the raster and *in situ* data were indexed according to the date of the data (month and year). Then the monthly rainfall data were grouped according to the distribution of weather stations in each region (Table 1). In the case of temperature, the stations were grouped for the entire study area.

4.3. Correction model for TRMM 3B43 and GPM 3IMERGM

In the analysis by regions, we could observe that only a few stations were available, so we decided not to perform a spatial interpolation. However, we propose the correction of the rainfall raster (TRMM and GPM) using an additive model based on the calculation of a correction factor (F_1) that is applied to each cell value of the raster data (Condom et al., 2011), This factor represents the difference between the raster and *in situ* data values of the stations located in each region, assuming that the *in situ* data are correct. For this purpose, we used the monthly rainfall values extracted from the raster data found in the attribute table of the point vector and using Equation (2) we calculated the cumulative F_1 for each month (i) of the year (j) in each region of the study area,

$$F_{1i} = \frac{\sum_{t=1}^{j} (RASTER_{i,j} - INSITU_{i,j})}{j} \quad (2)$$

where $RASTER_{i,j}$ is the rainfall estimated by TRMM and GPM over the (x, y) cell and $INSITU_{i,j}$ is the data measured by the stations located over the (x, y) cell. Consider that (x, y) is given by the geographical coordinates of the weather stations (Table 1). Then, we apply the pixel-level correction factor for all the original TRMM and GPM raster data according to Equation (3),

$$RASTER\_C_{i,j} = \sqrt[F1i]{RASTER\_O_{i,j} + 1} - 1 \quad (3)$$

where $RASTER\_C_{i,j}$ is the raster product with the corrected values of TRMM and GPM for month

i of year j, F_{1i} is the correction factor for each month and $RASTER\_O_{i,j}$ is the original raster of the rainfall estimated by TRMM and GPM. This equation was implemented using a Python script that allows to perform mathematical operations with raster or gridded data. For more details on additive models and raster data correction, please refer to the bibliography of Condom et al. (2011) and Lavado-Casimiro et al. (2009).

4.4. Correction model for MERRA-2

MERRA-2 data were corrected by applying a model based on the altitudinal variation and temperature gradient proposed by Fries et al. (2012). However, for this study we have made a modification in the altitude estimation for the reference level (Z_{Det}). The proposed modification consisted of performing the difference between the *in situ* temperature and the MERRA-2 temperature, thus identifying stations with similar temperatures and establishing the altitude of the reference level (Z_{Det}) based on this similarity. Subsequently, a linear regression was performed between temperature and elevation to calculate the temperature gradient with altitude (r); this process was carried out with the *in situ* and MERRA-2 temperature. The altitude of all values was reduced to the reference altitude level using the calculated gradients, thus eliminating vertical gradients in our study area. This reduction was calculated by applying Equation (4),

$$T_{Det} = T_m + \left(r * (Z_{Det} - Z_m)\right) \quad (4)$$

where T_{Det} is the monthly average temperature of the reference level, T_m is the average monthly temperature, and Z_m is the altitude of the weather station with *in situ* data and data extracted from MERRA-2. We then applied Equation (5) to replace the original altitude values of the MERRA-2 data using the temperature gradient obtained between the *in situ* values and the elevation, obtaining a MERRA-2 temperature with an initial correction,

$$T_{(x,y)} = T_{Det} + r * \left(Z_{(x,y)}^{DEM} - Z_{Det}\right) \quad (5)$$

where $T_{(x,y)}$ is the temperature at an altitude and location determined by the $DEM_{(x,y)}$, T_{Det} is the temperature determined in the Equation 5, $Z_{(x,y)}^{DEM}$ is the altitude of a cell in the DEM, and Z_{Det} is the altitude at which the temperature was determined. Finally, a comparison was made between the two linear equations obtained between the *in situ* data and MERRA-2 versus elevation. Both equations showed differences, with the MERRA-2 data being slightly underestimated, so the MERRA-2 data equation was added to the existing difference with the *in situ* data to obtain the final temperature MERRA-2.

4.5. Difference between corrected products and *in situ* data

The comparison of the satellite temperature and rainfall data with the *in situ* data from weather stations (Condom et al., 2011) was applied through Equation (6), which allows us to observe the difference between the corrected products and *in situ* data,

$$\Delta TPs = \left(TPs_{i,j} - EM_{i,j}\right) \quad (6)$$

where ΔTPs is the variation between the corrected products with the *in situ* data, $TPs_{i,j}$ are MERRA-2, TRMM and GPM corrected by the model, $EM_{i,j}$ are data from the weather stations, j number of years (2012-2017) and i number of months (January-December).

4.6. Evaluation metrics of corrected products

We analyzed the probabilistic behavior between the satellite and *in situ* data, using the root mean square error (RMSE) to verify that the estimated value is the lowest possible value in terms of its standard deviation (Despotovic et al., 2016). Also, the relative root mean square error (RRMSE) was used to indicate the precision of the satellite data with respect to the *in situ* data, being determined in percentage, if <50% is determined as reliable values, if ≥50% the estimate is considered unreliable (Condom et al., 2011). The evaluation of the covariation was carried out using the coefficient of determination (R^2). For a better interpretation, the R^2 values were classified as very good >0.95, good from 0.85 - 0.95, satisfactory from 0.65 - 0.85, and unsatisfactory <0.65 (Lujano-Laura et al., 2015). We also calculated the percentage BIAS (7) and NSE (Nash and Sutcliffe Efficiency) (8), values greater than 0.90 are very

good, good from 0.80 - 0.90, acceptable between 0.65 - 0.80 and unsatisfactory <0.65 (Ritter and Muñoz-Carpena, 2013),

$$BIAS = \left[\frac{\sum_{i=1}^{N}(TPs_i - EM_i)}{\sum_{i=1}^{N} EM_i}\right] \cdot 100 \qquad (7)$$

$$NSE = 1 - \frac{\sum_{i=1}^{N}(EM_i - TPs_i)^2}{\sum_{i=1}^{N}(EM_i - EM)^2} \qquad (8)$$

where EM_i is the multi-year monthly station value, TPs_i is the corrected estimated value (MERRA-2, TRMM and GPM), and i is the monthly average of all years (January-December).

5. Results and discussion

5.1. Evaluation of satellite data

TRMM. Rainfall values measured in the region R1Sn indicate that during the wet months (November-March) they fluctuate between 40 to 202 mm, while the driest months (April-October) vary between 0 to 80 mm. In the R2Mr, the rainfall values in the wet season range between 50 to 213 mm and in the dry season between 0 to 87 mm. In R3Pt the rainfall values reach values of 10 to 156 mm in the wet season and minimum values in the dry season between 0 to 40 mm and finally, R4Pc is the driest region since the rainfall values reach values of 3 to 124 mm in the wet season and minimum values in the dry season between 0 to 22 mm (Figure 2).

GPM. The GPM rainfall values indicate that the R1Sn region fluctuates between 36 to 218 mm (wet season), while in the dry months the rainfall varies between 0 to 59 mm. In R2Mr between 55 to 210 mm in the wet months and between 0 to 83 mm in the dry months. R3Pt reaches rainfall values between 10 to 167 mm in the wet season and minimum values in the dry season between 0 to 48 mm. Finally, R4Pc as in the TRMM data, this region presents the lowest rainfall values between 10 to 105 mm in the wet season and minimum values in the dry season between 0 to 23 mm (Figure 2). The rainfall data generated by the satellites have a high degree of relationship with respect to the *in situ* data having a maximum of TRMM 0.83 for region 3 and in GPM of 0.81 for region 1. These values are higher than others calculated for zones similar to our study area, for example, in the Santa River basin an R^2 value between 0.7 (Condom et al., 2011) and 0.86 (Lujano-Laura et al., 2015) was estimated.

MERRA-2. Temperature data show that below 1000 m a.s.l., they have a temperature between

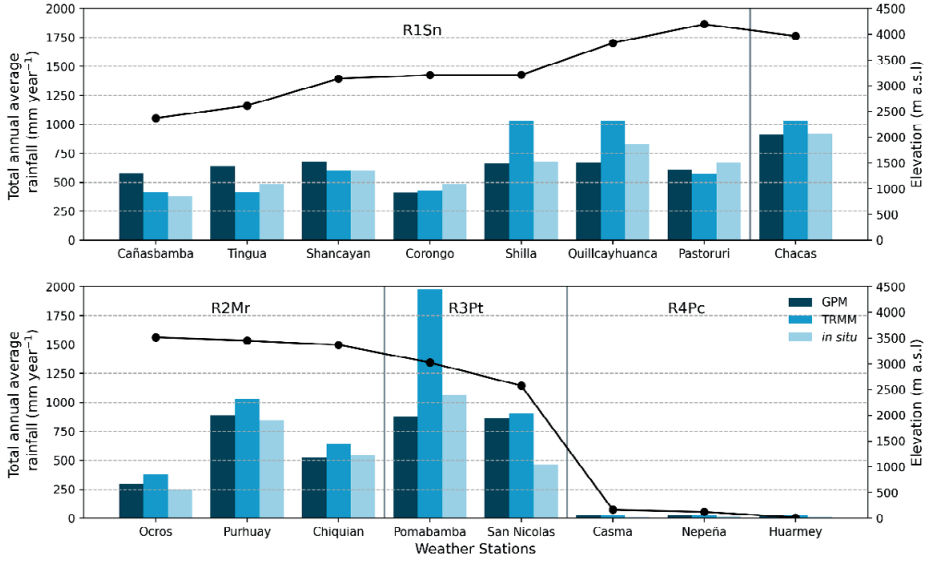

Figure 2. Comparison of rainfall variability as a function of altitude, region, weather station and satellite data (TRMM and GPM).

21 °C to 26 °C. On the other hand, in areas above 2000 m a.s.l. temperatures are between 7.4 °C and 14.8 °C, while in areas above 3500 m a.s.l. temperatures vary between 2.15 °C and 5.8 °C. The MERRA-2 data have an annual monthly correlation with the altitude shown in Figure 3, where there is a minimum of 0.91 in October, while reaching a maximum of 0.94 in March. This correlation of the temperature data has a behavior similar to the results obtained from the study of Ninyerola et al. (2000) and Rau & Condom (2010).

5.2. Analysis of satellite, *in situ* and model products

Rainfall. The rainfall data has a quite differentiated behavior, observing that in R1Sn, R2Mr and R3Pt between the months of May to September they are lower than in the rest of the months (rainfall values between 0 and 125 mm) but that they May present anomalies with values close to 250 mm (as in the case of R2Mr). With the month of March being the rainiest (>350 mm) in these three regions. This behavior coincides with the rainfall patterns of the western-central part of mountain areas found by (Endara-Huanca, 2016). On the other hand, R4Pc presents the driest months between the months of April to November, while the months of December to March present a little rainfall, which coincides with the patterns found in the North Coast by Endara-Huanca (2016). The corrected products for rainfall were corrected using Equations (2) and (6). The corrected values show that the existing difference with the station values is about 200 mm (Figure 4), both for TRMM and GPM. This difference is similar to the one calculated by Condom et al. (2011) for some areas of Peru, and it is statistically acceptable, since the difference is close to 0, indicating that both data are related to a monthly scale. Therefore, this model allowed obtaining rainfall data without being affected by the low density of weather stations and mountain topography, fitting these conditions, since the application of the Thiessen polygon or other interpolation methods require a large availability of *in situ* data and usually do not represent the spatial gradients of rainfall in our study area.

Temperature. The analysis of the temperature as a function of the altitude shows that there is a non-constant behavior throughout the year, with periods of greater correlation being *in situ* in the months of January to March (coinciding with the summer), while there is a decrease in the correlation for April - August months (Figure 5) that agrees with the

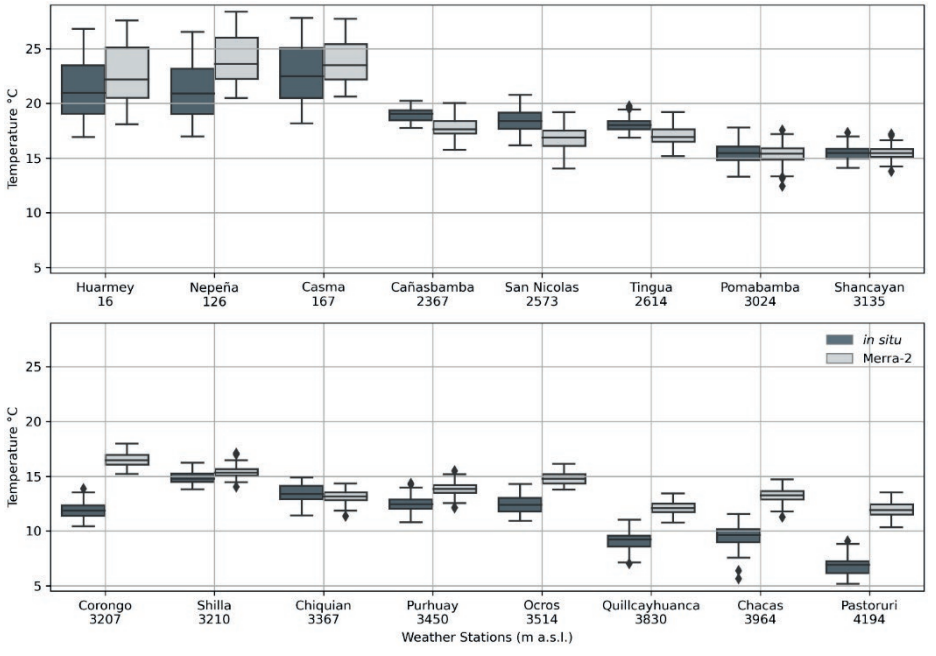

Figure 3. Comparison between the mean temperature at each weather station and the MERRA-2 data.

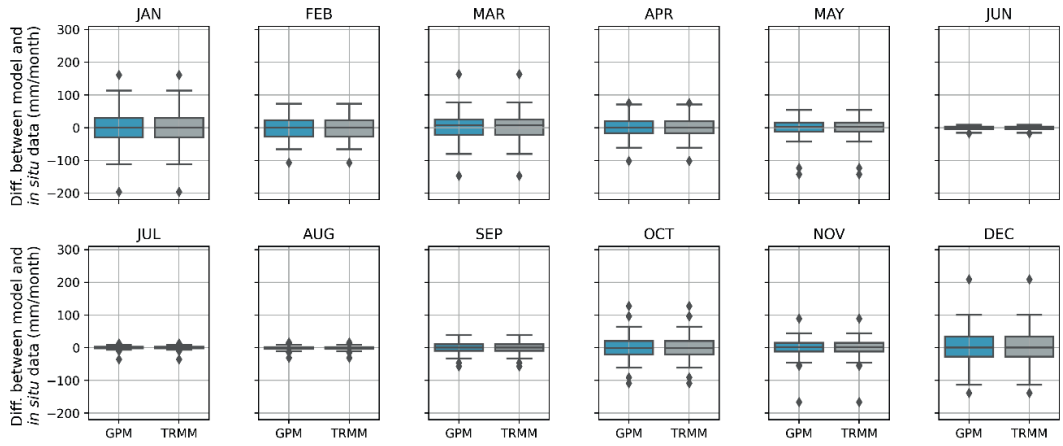

Figure 4. Variability of TRMM and GPM corrected product and *in situ* data.

months of greatest rainfall in the coastal zone and that corresponds to the dry season in the mountain areas, as described by Quevedo and Sánchez (2009). The MERRA-2 data were corrected, since they have a constant variation and a behavior repeated throughout the year with an underestimation of -4.8 °C on average.

5.3. Validation and accuracy

Corrected products for the TRMM and GPM (Table 2), in R1Sn and R2Mr, improve considerably the correlation, becoming satisfactory. The quadratic error also reaches acceptable values, and the NSE indicates that the additive model works well. However, for R1Pt the quadratic error is very high, despite the fact that the predictor model works well, while for R4Pc most of the statistics calculated are unacceptable, that is, the satellite products do not improve despite their correction using *in situ* data. According to Condom et al. (2011), the TRMM and GPM data are not adequate to be used on the Peruvian coast, something that is evidenced in this study as well. Adding other variables that reflect the behavior of rainfall could support the improvement of TRMM and GPM satellite data, for example a variable such as the NDVI, which yielded good results according to Georganos et al. (2017).

Regarding the model implemented for temperature, an improvement in the correlation between the *in situ* data and the satellite data has been obtained, obtaining the improvement of the correlation in a maximum of 0.26, while the RMSE has maximum values of 1.94 and minimum

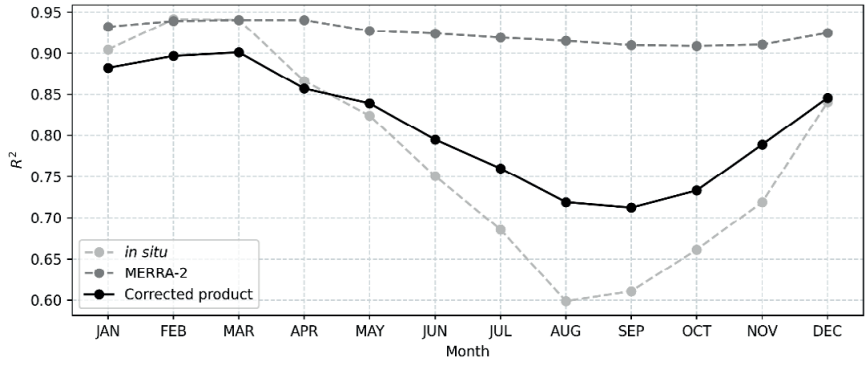

Figure 5. Correlation of temperature-elevation from *in situ* data, MERRA-2 and corrected product per months.

Table 2. Metrics of corrected products for TRMM and GPM rainfall.

	R1Sn		R2Mr		R3Pt		R4Pc	
Metrics	TRMM	GPM	TRMM	GPM	TRMM	GPM	TRMM	GPM
R^2 (a)	0.50	0.82	0.62	0.60	0.83	0.81	0.32	0.53
R^2 (b)	0.73	0.84	0.66	0.65	0.85	0.88	0.38	0.62
BIAS	-1.60	-2.01	-1.90	-2.09	-2.28	-2.05	-1.96	-1.97
RMSE	28.65	21.62	37.99	37.53	20.78	18.57	3.35	3.30
RRMSE	48%	40%	48%	48%	52%	49%	268%	257%
NSE	0.75	0.82	0.67	0.60	0.84	0.88	**0.06**	**0.37**

values of 1.25 (Table 3), which are similar to the results obtained for RMSE (1.83 °C) in the study of Benali et al. (2012), demonstrating that high precision monthly and annual climatic data can be generated following methodologies that integrate statistical techniques and GIS methods (Ninyerola et al., 2000).

Table 3. Metrics results of corrected products for MERRA-2 temperature.

Month	R^2 (a)	R^2 (b)	BIAS	RMSE
January	0.87	0.93	-0.1	1.40
February	0.90	0.95	+1.4	1.25
March	0.90	0.95	+0.1	1.29
April	0.83	0.92	0	1.39
May	0.80	0.90	+1.1	1.46
June	0.73	0.84	+1.0	1.70
July	0.67	0.79	-1.4	1.86
August	0.60	0.72	-0.8	1.94
September	0.62	0.75	-0.1	1.83
October	0.67	0.80	+0.5	1.66
November	0.73	0.84	+0.7	1.58
December	0.82	0.91	-0.5	1.41

6. Conclusions

The monthly rainfall and temperature in the 16 weather stations was characterized by strong seasonality in the values recorded during the wet period (October to March) and the dry season (April to September). One of the main problems found is that the TRMM and GPM products do not allow the prediction of rainfall in the Peruvian coast, due to the scarce quantity of weather stations. The MERRA-2 data tend to underestimate the actual temperature values, which shows that data must be corrected before their use. Therefore, an additive model for rainfall and a regression model for temperature can be applied to monthly satellite data. Without correction, the TRMM and GPM data showed a low correlation with the *in situ* data, but it was improved to a maximum value of 0.83 in the R1Sn, R2Mr and R3Pt regions. In the case of temperature, the correlation values improved in a range from 0.72 to 0.95 for the entire study area. Finally, the corrected data provided the possibility of having high-precision climate information in high mountain areas. However, it is still pending to explore new techniques for the estimation of these climatic variables at a higher spatial and temporal resolution in areas with few weather stations and high spatial variability.

7. Data availability

Corrected products based on TRMM, GPM and MERRA-2 from January 2012 to December 2017 are available as open source in the following links:

- The GPM corrected data is available in https://doi.org/10.6084/m9.figshare.14541720
- The TRMM corrected data is available in https://doi.org/10.6084/m9.figshare.14541903
- The MERRA-2 corrected data is available in https://doi.org/10.6084/m9.figshare.14541930

Acknowledgements

The authors acknowledge the financial support from the CONCYTEC - World Bank Project “Improvement and Expansion of the National Science Technology and Technological Innovation System Services”; 8682-PE, through its executing unit FONDECYT [Contract N°23-2018-FONDECYT-BM-IADT-MU] of Permafrost Project and from the Newton-Paulet Fund and the NERC within the framework of the call E031-2018-01-NERC & Glacier Research, through its

executing unit FONDECYT [Contract N°08-2019-FONDECYT] of PeruGROWS project. We thank Rafael Tauquino, Ciro Fernández and Ricardo Villanueva for providing the meteorological data from the Center for Environmental Research for Development (CIAD), Santiago Antunez de Mayolo National University (UNASAM).

References

Aybar, C., Lavado, W., Huerta, A., Fernández, C., Vega, F., Sabino, E., Felipe, O. 2017. Uso del Producto Grillado "PISCO" de precipitación en Estudios, Investigaciones y Sistemas Operacionales de Monitoreo y Pronóstico Hidrometeorológico. In *Nota Técnica 001 SENAMHI-DHI-2017* (pp. 1–22). SENAMHI.

Aybar, C., Fernández, C., Huerta, A., Lavado, W., Vega, F., Felipe-Obando, O. 2019. Construction of a high-resolution gridded rainfall dataset for Peru from 1981 to the present day. *Hydrological Sciences Journal*, *65*(5), 770–785. https://doi.org/10.1080/02626667.2019.1649411

Benali, A., Carvalho, A.C., Nunes, J.P., Carvalhais, N., Santos, A. 2012. Estimating air surface temperature in Portugal using MODIS LST data. *Remote Sensing of Environment*, *124*, 108–121. https://doi.org/10.1016/j.rse.2012.04.024

Beniston, M., Diaz, H.F., Bradley, R. 1997. Climatic change at high elevation sites: an overview. *Climatic Change*, *2*, 157–167. https://doi.org/10.1007/978-94-015-8905-5_1

Condom, T., Rau, P., Espinoza, J.C. 2011. Correction of TRMM 3B43 monthly precipitation data over the mountainous areas of Peru during the period 1998-2007. *Hydrological Processes*, *25*(12), 1924–1933. https://doi.org/10.1002/hyp.7949

Despotovic, M., Nedic, V., Despotovic, D., Cvetanovic, S. 2016. Evaluation of empirical models for predicting monthly mean horizontal diffuse solar radiation. *Renewable and Sustainable Energy Reviews*, *56*, 246–260. https://doi.org/10.1016/j.rser.2015.11.058

Endara-Huanca, S.M. 2016. Ciclos Horarios De Precipitación En El Perú Utilizando Información Satelital Dirección. In *Servicio Nacional de Meteorología e Hidrología*. Available at http://repositorio.senamhi.gob.pe/handle/20.500.12542/112

Fernández-Palomino, C.A., Hattermann, F.F., Krysanova, V., Lobanova, A., Vega-Jácome, F., Lavado, W., Santini, W., Aybar, C., Bronstert, A. 2022. A Novel High-Resolution Gridded Precipitation Dataset for Peruvian and Ecuadorian Watersheds: Development and Hydrological Evaluation. *Journal of Hydrometeorology*, *23*(3), 309–336. https://doi.org/10.1175/JHM-D-20-0285.1

Franchito, S.H., Rao, V.B., Vasques, A.C., Santo, C.M.E., Conforte, J.C. 2009. Validation of TRMM precipitation radar monthly rainfall estimates over Brazil. *Journal of Geophysical Research Atmospheres*, *114*. https://doi.org/10.1029/2007JD009580

Fries, A., Rollenbeck, R., Nauß, T., Peters, T., Bendix, J. 2012. Near surface air humidity in a megadiverse Andean mountain ecosystem of southern Ecuador and its regionalization Andreas. *Agricultural and Forest Meteorology*, *2012*, 17–30. https://doi.org/10.1016/j.agrformet.2011.08.004

Garreaud, R., Vuille, M., Clement, A.C. 2003. The climate of the Altiplano: Observed current conditions and mechanisms of past changes. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *194*(1–3), 5–22. https://doi.org/10.1016/S0031-0182(03)00269-4

Gelaro, R., McCarty, W., Suárez, M.J., Todling, R., Molod, A., Takacs, L., Randles, C.A., Darmenov, A., Bosilovich, M.G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., da Silva, A.M., Gu, W., … Zhao, B. 2017. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

Georganos, S., Abdi, A.M., Tenenbaum, D.E., & Kalogirou, S. 2017. Examining the NDVI-rainfall relationship in the semi-arid Sahel using geographically weighted regression. *Journal of Arid Environments*, *146*, 64–74. https://doi.org/10.1016/j.jaridenv.2017.06.004

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

Krakauer, N.Y., Pradhanang, S.M., Lakhankar, T., Jha, A.K. 2013. Evaluating satellite products for precipitation estimation in mountain regions: A case study for Nepal. *Remote Sensing*, *5*(8), 4107–4123. https://doi.org/10.3390/rs5084107

Lavado-Casimiro, W., Labat, D., Guyot, J.L., Ronchail, J., Ordóñez, J.J. 2009. Validation of rainfall using the TRMM for two Peruvian Amazon basins and its inclusion in monthly water balance models. *Revista Peruana Geo-Atmosferica RPGA*, *19*(1), 11–19.

Liu, C., Zipser, E.J. 2015. The global distribution of largest, deepest, and most intense precipitation systems. *Geophysical Research Letters*, *42*(9), 3591–3595. https://doi.org/10.1002/2015GL063776

Lu, X., Wei, M., Tang, G., Zhang, Y. 2018. Evaluation and correction of the TRMM 3B43V7 and GPM 3IMERGM satellite precipitation products by use of ground-based data over Xinjiang, China. *Environmental Earth Sciences*, *77*(5). https://doi.org/10.1007/s12665-018-7378-6

Lujano-Laura, E., Felipe-Obando, O.G., Lujano-Laura, A., Quispe-Aragón, J. 2015. Validación de la precipitación estimada por satélite TRMM y su aplicación en la modelación hidrológica del río Ramis Puno Perú. *Revista Investigaciones Altoandinas - Journal of High Andean Investigation*, *17*(2), 221. https://hdl.handle.net/20.500.12542/1045

McCarty, W., Coy, L., Gelaro, R., Huang, A., Merkova, D., Smith, E.B., Sienkiewicz, M., Wargan, K. 2016. *MERRA-2 Input Observations: Summary and Assessment. Technical Report Series on Global Modeling and Data Assimilation 46*. *46*(October), 51.

Motschmann, A., Huggel, C., Carey, M., Moulton, H., Walker-Crawford, N., Muñoz, R. 2020. Losses and damages connected to glacier retreat in the Cordillera Blanca, Peru. *Climatic Change*, *162*(2), 837–858. https://doi.org/10.1007/s10584-020-02770-x

Mourre, L., Condom, T., Junquas, C., Lebel, T., E. Sicart, J., Figueroa, R., Cochachin, A. 2016. Spatio-temporal assessment of WRF, TRMM and in situ precipitation data in a tropical mountain environment (Cordillera Blanca, Peru). *Hydrology and Earth System Sciences*, *20*(1), 125–141. https://doi.org/10.5194/hess-20-125-2016

Ninyerola, M., Pons, X., Roure, J.M. 2000. A methodological approach of climatological modeling of air temperature and precipitation. *International Journal of Climatology*, *20*, 1823–1841. https://doi.org/10.1002/1097-0088(20001130)20:14%3C1823::AID-JOC566%3E3.0.CO;2-B

Ouatiki, H., Boudhar, A., Tramblay, Y., Jarlan, L., Benabdelouhab, T., Hanich, L., El Meslouhi, M., Chehbouni, A. 2017. Evaluation of TRMM 3B42 V7 Rainfall Product over the Oum Er Rbia Watershed in Morocco. *Climate*, *5*(1), 1. https://doi.org/10.3390/cli5010001

Quevedo, K., Sánchez, K. 2009. Comparación de dos métodos de interpolación para la estimación de la temperatura del aire aplicando técnicas geo-estadísticas. *Revista Peruana Geo-Atmosferica*, *107*(1), 90–107.

Rau, P., Condom, T. 2010. Análisis espacio temporal de la precipitación en las zonas de montaña de Peru (1998-2007). *Revista Peruana Geo Atmosférica*, *29*(2), 16–29.

Ritter, A., Muñoz-Carpena, R. 2013. Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, *480*, 33–45. https://doi.org/10.1016/J.JHYDROL.2012.12.004

Vicente-Serrano, S.M., López-Moreno, J.I., Correa, K., Avalos, G., Bazo, J., Azorin-Molina, C., Domínguez-Castro, F., Kenawy, E., Gimeno, L., Nieto, R. 2017. Recent changes in monthly surface air temperature over Peru, 1964–2014. *International Journal of Climatology*, *38*(1), 283–306. https://doi.org/10.1002/joc.5176

Zhang, X., Feng, Y., & Chan, R. 2018. Introduction to RClimDex. *Climate Research Division Environment Canada*, 1–26.