

Article

# High-Resolution Precipitation Datasets in South America and West Africa based on Satellite-Derived Rainfall, Enhanced Vegetation Index and Digital Elevation Model

Guido Ceccherini <sup>1</sup>, Iban Amezttoy <sup>1</sup>, Claudia Patricia Romero Hernández <sup>2</sup>  
and Cesar Carmona Moreno <sup>1,\*</sup>

<sup>1</sup> Institute for Environment and Sustainability, DG Joint Research Centre, European Commission, Ispra 21027, Italy; E-Mails: guido.ceccherini@ext.jrc.ec.europa.eu (G.C.)

iban.amezttoy@jrc.ec.europa.eu (I.A.)

<sup>2</sup> Facultad de Ingeniería Ambiental, Universidad Santo Tomás, Bogota 5878797, Colombia; E-Mail: claudia.romero@usantotomas.edu.co

\* Author to whom correspondence should be addressed;

E-Mail: cesar.carmona-moreno@jrc.ec.europa.eu;

Tel.: +39-332-783-920; Fax: +39-332-789-073.

Academic Editors: Richard Gloaguen and Prasad S. Thenkabail

Received: 15 December 2014 / Accepted: 15 May 2015 / Published: 22 May 2015

---

**Abstract:** Mean Annual Precipitation is one of the most important variables used in water resource management. However, quantifying Mean Annual Precipitation at high spatial resolution, needed for advanced hydrological analysis, is challenging in developing countries which often present a sparse gauge network and a highly variable climate. In this work, we present a methodology to quantify Mean Annual Precipitation at 1 km spatial resolution using different precipitation products from satellite estimates and gauge observations at coarse spatial resolution (*i.e.*, ranging from 4 km to 25 km). Examples of this methodology are given for South America and West Africa. We develop a downscaling method that exploits the relationship among satellite-derived rainfall, Digital Elevation Model and Enhanced Vegetation Index. Finally, we validate its performance using rain gauge measurements: comparable annual precipitation estimates for both South America and West Africa are retrieved. Validation indicates that high resolution Mean Annual Precipitation downscaled from *CHIRP* (Climate Hazards Group Infrared Precipitation) and *GPCC* (Global Precipitation Climatology Centre) datasets present the best ensemble of performance statistics for both South America and West Africa. Results

also highlight the potential of the presented technique to downscale satellite-derived rainfall worldwide.

**Keywords:** satellite-derived precipitation; downscaling; EVI; DEM; geographically weighted regression; developing countries; South America; West Africa; mean annual precipitation

---

## 1. Introduction

Quantifying the spatial distribution of annual rainfall is crucial for water resource management in developing countries such as those of South America and West Africa. Precipitation regulates hydrological and agricultural process, and an accurate estimate of the Mean Annual Precipitation (hereafter *MAP* or annual precipitation) is necessary to different disciplines including, among others, hydrological modeling, agricultural resources management, ecology and meteorology.

Beside its importance for an improved knowledge of the water cycle, annual precipitation is also crucial for mitigation strategies of natural hazards and Disaster Risk Reduction: the work by Hosking *et al.* [1] on *L-moments* and regional frequency analysis identified statistically significant relationship between rainfall extremes and annual precipitation (for further reference see Schaefer [2]).

However, the main issue is the difficulty with obtaining precipitation data with an adequate spatial resolution. Many watersheds, especially in developing countries, are poorly instrumented and it becomes difficult to generate maps of annual precipitation using only data from rain-gauge stations. The sparse observational data coverage also means that use of advanced hydrological analysis is not allowed.

In the last 30 years, precipitation estimates are also available at global scale from satellite sensors. Satellite-derived precipitation products generally exploit the Thermal Infrared (*TIR*), the microwave channel and ground measurements (for further reading see Tapiador *et al.* [3]). The launch in 2014 of the *NASA* Global Precipitation Measurements mission (*GPM*, [4]) is an evidence of the central importance of rainfall estimation from satellite for the hydrological and geophysical research.

Nevertheless, satellite-derived precipitation is at coarse spatial resolution (see Table 1 in Section 2), while for many hydrological applications it is recommended to use a finer spatial scale. For example, the Tropical Rainfall Measuring Mission (*TRMM*, see Section 2.2.2) has a spatial resolution of ~25 km, that may not capture the spatial heterogeneity of precipitation fields. Generally, the finer the spatial resolution, the better the estimation of runoff volumes and other hydrological fluxes [5]. This issue is amplified at low latitudes, dominated by convective rain events with high spatial variability [6].

A common answer within the hydrological community consists in the spatial downscaling/disaggregation of low-resolution rainfall from satellite sensors [7]. Precipitation fields vary depending on altitude and wind direction, and they influence vegetation greenness. Many studies thereof hypothesize a linear relationship between precipitation and satellite-based products such as Digital Elevation Model (hereafter *DEM*), and vegetation indices such as the Normalized Difference Vegetation Index (*NDVI*) or the Enhanced Vegetation Index (*EVI*) [8]. For example, Immerzeel *et al.* [9] assumed that *NDVI* is a proxy for cumulated precipitation over the Iberian Peninsula, and consequently the *NDVI* has been

regressed against satellite-derived precipitation. Hunink *et al.*, Jia *et al.* and Park [10–12] developed linear regression models employing the *DEM* and the *NDVI* to downscale rainfall, considering also the topographical effects. Most of approaches use global regression analysis, assuming that the relationship between precipitation, elevation and vegetation greenness does not vary spatially [13]. However, this assumption is inconsistent with many studies that contradict the spatial stationarity of the relationship between precipitation and *NDVI* [14]. Especially for large area (*i.e.*, analysis at regional/continental scale), when the non-stationarity in the relationship among rainfall and other environmental variables is modeled by stationary models, possible wrong conclusions might be drawn.

A recent and original approach is the one implemented by Chen *et al.* [15], where satellite-derived precipitation is downscaled at 1 km spatial resolution using a Geographically Weighted Regression (*GWR*), a local form of linear regression used to model spatially varying relationships (for further details on *GWR* see Section 2.2).

The goal of this study is to develop a downscaling scheme to obtain annual precipitation at 1 km spatial resolution across South America and West Africa. To address this question we created datasets of *MAP* at a spatial resolution of 1 km in South America and Western Africa, downscaling the estimates of annual precipitation acquired by satellites and gauge observations.

In the present study we exclude “local rain gauge” datasets since we want to employ only well-established and long-standing precipitation datasets. Precipitation information is mainly derived by satellite instruments, and to a lesser extent by gauge observations (see Section 2.1.): henceforward, estimates of annual precipitation acquired by satellites or gauge observations that we want to downscale will be referred to as “satellite-derived *MAP*”.

The originality of this study lies in the high resolution datasets of annual precipitation, rather than in the methodology *per se*. As previously mentioned, the downscaling scheme implemented in this study has been proposed by Chen *et al.* [15] across North China. However, the proposed methodology offers some elements of novelty compared to the models that have inspired it, such as the “global” character of the methodology and the use of the *EVI* instead of the *NDVI* (for further discussion, see Discussion section).

The downscaling method presented below implements the procedure in which, once we have chosen the country and the rainfall dataset: (1) a Geographically Weighted Regression (*GWR*) is performed among the satellite-derived *MAP*, *DEM* and average annual *EVI* at coarse spatial resolution, (2) the residual of the regression—*i.e.*, the amount of satellite-derived precipitation that could not be explained by the model- is interpolated at 1 km resolution and (3) the *GWR* is applied to *DEM* and average annual *EVI* at the original resolution of 1 km, and the high-resolution residual is added.

This technique is applied for each of the seven satellite-derived precipitation products available (see Sections 2.1.1 to 2.1.7). For each product, validation analysis with ground weather stations has been carried out to determine the reliability of rainfall estimates at high spatial resolution.

Combining these datasets and techniques the aim of the study is to: (1) improve retrievals of rainfall at high spatial resolution, and (2) provide constraints on the uncertainty in precipitation estimates from remote sensing across West Africa and South America. This paper describes the methodology and presents results over South America and over West Africa, even if it can be applied worldwide.

## 2. Materials and Methods

### 2.1. Materials

In this section we present the rainfall datasets used in the present study, the Enhanced Vegetation Index, the Digital Elevation Model and the validation dataset. Table 1 summarizes the rainfall datasets used. The latter are described in Subsections 2.1.1 to 2.1.7, whereas Enhanced Vegetation Index, the DEM and the validation dataset are described in Subsections 2.1.8 to 2.1.10.

**Table 1** Overview of widely known precipitation products from satellite estimates (Sat.) and ground rain gauge observations (Gr.).

Name	Data Input	Resolution	Start/End Date	Spatial Extent
GPCC	Gr.	0.25°	1951/2000	World-Wide
TRMM 3B43	Sat. + Gr.	0.25°	1998/present	50° S–50° N; all long.
PERSIANN CDR	Sat. + Gr.	0.25°	1983/present	50° S–50° N; all long.
CHIRP	Sat. + Gr.	0.05°	1981/present	50° S–50° N; all long.
CMORPH	Sat. + Gr.	0.25°	1998/present	50° S–50° N; all long.
RainFall Estimator	Sat. + Gr.	0.1°	2000/present	40° S–40° N; –20W–55E
TAMSAT	Sat. + Gr.	0.0375°	1983/2012	40° S–40° N; –20W–55E

For all practical purposes, (1) all datasets are first corrected removing obvious outliers, then (2) the daily/weekly/monthly products are aggregated on annual basis, and (3) then the annual precipitation is calculated for the entire reference period of the dataset.

This study has been carried out for the well-established and long-standing precipitation datasets available in literature. However, for the sake of simplicity, not all the global precipitation datasets available have been used. For example, Brocca *et al.* [16] recently proposed a promising “bottom-up” approach that, by doing inverse modeling, employs soil moisture estimates from microwave satellite to infer the preceding rainfall amounts. However, the rough spatial resolution (~100 km) effectively renders it useless for this study. Precipitation from Global Climatic Models (GCM) [17] are also excluded because of their rough spatial resolution (again, in the region of 100 km).

If possible, we consider the longest time series of precipitation even if their timespans are not consistent with each other, in order to have a more reliable estimate of the MAP. As a result, we assume that temporal trends in annual precipitation are negligible. This is a strong assumption, but trend analysis and trend removal at continental scale, along with the spatial interpolation of low-resolution datasets would only make the final outcome more inaccurate. Moreover, trend analysis is beyond the scope of the study.

#### 2.1.1. GPCC

The Global Precipitation Climatology Centre (GPCC) provides the mean monthly global land-surface precipitation for every month of the year and the annual total [18]. This climatology is available on a regular latitude/longitude grid with different spatial resolutions: 0.25°, 0.5°, 1.0°, and 2.5° (*i.e.*, ~25, 50, 100 and 250 km, respectively). The GPCC dataset at 0.25° is the best for complete spatial coverage, whereas the GPCC dataset at 0.50° has the highest accuracy. We decided to use in our study

the 0.25° product, preferring the spatial coverage and the spatial resolution to the accuracy. The *GPCC* rainfall product at 0.25° spatial resolution (hereafter *GPCC*) is based on the 67,200 stations world-wide that feature record durations of 10 years or longer, for the target reference period January 1951 to December 2000 [19].

#### 2.1.2. TRMM 3B43

*TRMM 3B43* dataset is based on the Tropical Rainfall Measuring Mission (*TRMM*) and the monthly accumulated *GPCC* rain gauge dataset. *TRMM* is a joint U.S.-Japan satellite mission to monitor tropical and subtropical precipitation, which employs radars, passive microwave along with visible/infrared radiometers. The purpose of *TRMM 3B43* is to produce the “*TRMM and Other Data*” best-estimate precipitation estimates. Basically, *TRMM* rainfall data are summed for the calendar month, and then the *GPCC* rain gauge data are used to perform a large-scale bias correction [20]. The final gridded precipitation data have a monthly temporal resolution and a 0.25° spatial resolution. *TRMM 3B43* dataset (hereafter *TRMM*) is available for the period January 1998 to April 2015. Spatial coverage extends from 50° S to 50° N latitude (for further information see [21]).

#### 2.1.3. PERSIANN CDR

The *PERSIANN CDR* (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks Climate Data Record) consists of daily precipitation estimates derived from long wave infrared satellite data [22]. The *PERSIANN CDR* dataset (hereafter *PERSIANN*) provides precipitation data beginning in 1983 at 0.25° spatial resolution. Satellite-derived rainfall are adjusted using the *GPCC* monthly product to maintain consistency of the two datasets throughout the entire record [23].

#### 2.1.4. CMORPH

*CMORPH* refers to the name of a technique (MORPHing technique) to produce global precipitation maps from microwave satellite source at high spatial and temporal resolution. At present, the National Oceanic and Atmospheric Administration (*NOAA*) released a *CMORPH* product, called *CMORPH* Version 1.0 [24] incorporating precipitation estimates derived from the passive microwaves aboard the *NASA's AQUA* and *TRMM* spacecraft (for further information see Joyce *et al.* [25]). Please note that, although the dataset is also available at 8 km spatial resolution, the resolution of the individual satellite estimates is coarser than that. Therefore the finer resolution is obtained via interpolation. We used in the present study the *CMORPH* gauge-satellite blended precipitation product from 1998 onwards, at 0.25° spatial resolution (hereafter *CMORPH*). A routine has been implemented to: (1) aggregate the daily rainfall to annual time step and (2) calculate the annual precipitation throughout the entire record period.

#### 2.1.5. CHIRP

Climate Hazards Group Infrared Precipitation with Station data (hereafter *CHIRP*) is a 30-year quasi-global rainfall dataset [26]. Spanning 50° S–50° N (and all longitudes), starting in 1981 to

present, *CHIRP* incorporates 0.05° resolution satellite imagery with *in-situ* station data to create gridded monthly time series for seasonal drought monitoring and water resource management [27].

#### 2.1.6. RFE

The Rainfall Estimate (hereafter *RFE*) is an operational product created by the Climate Prediction Center for the United States Agency for International Development Famine Early Warning System project to assist in drought monitoring and flood forecasting over Africa [28]. The spatial resolution is 0.1° and the spatial domain spans −40° S to 40° N in latitude, and 20° W to 55° E in longitude, encompassing thus the whole Africa continent. Daily precipitation maps are created by merging gauge observations and three kinds of satellite estimates (*i.e.*, *GPI*, *SSM/I* and *AMSU*) using the method described by Xie and Xiong [29]. Data is available from 2000 onwards [30].

#### 2.1.7. TAMSAT

*TAMSAT* stands for Tropical Applications of Meteorology using SATellite data and ground-based observations and it is developed at the department of meteorology of the University of Reading, UK. The *TAMSAT* group provides a 30-year rainfall climatology (from 1983 to 2012) covering Africa with a spatial resolution of 0.0375° [31]. Ten-daily and monthly rainfall estimates, along with the corresponding anomalies, are derived from archived Meteosat thermal infrared imagery, calibrated against rain gauge records collected from numerous African agencies [32].

#### 2.1.8. Vegetation Index: EVI

Vegetation indices are dimensionless variable, generally varying between zero and one, useful for many applications, such as agriculture and hydrological modeling, which can be monitored from space. Vegetation indices are attractive because they can provide information on catchment water balance, and because their spatial patterns are driven by water availability [33].

The most widely used vegetation index is the Normalized Difference Vegetation Index (*NDVI*), which consists of a normalized ratio of the near infra-red and red spectral bands. However, *NDVI* often saturates, it is affected by atmospheric and soil conditions, and it might be no longer sensitive to changes in vegetation [34]. To overcome these limitations, new vegetation indices, such as the Enhanced Vegetation Index (*EVI*), have been created. *EVI* is a Vegetation Index that effectively characterizes bio-physical/ biochemical states and processes from vegetated surfaces [8]. *EVI* minimizes canopy background variations and maintains sensitivity over dense vegetation conditions. The *EVI* also employs the blue spectral band, particularly sensitive to aerosols, to account for atmospheric effects. *EVI* is calculated from these measurements as follows:

$$EVI = G \cdot \frac{(NIR - RED)}{(NIR + C_1 \cdot RED - C_2 \cdot BLUE + L)} \quad (1)$$

where *RED*, *BLUE* and *NIR* stand for the surface reflectance measurements acquired in the visible (red and blue) and near-infrared regions, respectively; whereas *C*<sub>1</sub>, *C*<sub>2</sub> are the coefficients describing the aerosol resistance, and *L* is the canopy background adjustment parameter.

The most successful applications of *EVI* are reported in areas having high biomass such as the Amazon forest [8,35] and Africa rainforest [36]. The *EVI*, in the same manner as the *NDVI*, is sensitive to the cumulated rainfall, and its average annual value can be used as a proxy for the mean annual precipitation [9,15].

The *EVI* dataset employed in this study is derived from monthly observations captured by Moderate Resolution Imaging Spectroradiometer (*MODIS*) on board the Terra satellite at spatial resolutions of 1 km (*MODIS* product MOD13A3) from 2000 to present [37]. *EVI* dataset has been averaged at annual scale, and then we have calculated the average annual *EVI* for the entire reference period of the dataset. This average annual *EVI*—a kind of “climatology” of annual *EVI* with 14 years of observations—is the product that we use consecutively to downscale satellite-derived precipitation.

It has not escaped our notice the issue of the temporal consistency between *EVI* and rainfall datasets which often span different or successive periods (e.g., *GPCC* refers to the 1951–2000 period). However, we assume that in ~10 years *EVI* dataset provides us with a good estimate of the climatology of vegetation, filtering out anomalies and outliers. In order to obtain a consistent *EVI* measurement through a longer time, we should have processed *GIMMS* (Global Inventory Modeling and Mapping Studies) dataset [38] instead. However, *GIMMS* dataset has been discarded since it presents a coarse resolution (*i.e.*, ~8km), combining different sensors spanning successive periods and does not employ the blue spectral band, particularly sensitive to aerosols, not accounting thus for atmospheric correction.

One of the main problem facing with *EVI* products from *MODIS* is represented by rivers, lakes and water bodies, where the *EVI* has negative and/or low values. These areas have to be identified, filtered out and interpolated to avoid problems with the subsequent regression. Identification of water bodies is not straightforward [39] since we cannot use an a priori threshold to mask *EVI* value corresponding to water bodies. In order to have a global and portable method, we use the *SWBD* dataset (SRTM Water Bodies Data, [40]) to mask water bodies and to remove the corresponding *EVI* values. *EVI* values are subsequently filled with a simple bilinear interpolation.

#### 2.1.9. DEM

Information on the elevation (*i.e.*, DEM) is available from the *HydroSHEDS* product (Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales [41]), derived from the Shuttle Radar Topography Mission (*SRTM*, [42]). The spatial resolution of the *DEM* is 1 km. *HydroSHEDS* product provides hydrographic information in a consistent and comprehensive format for regional and global-scale hydrological applications. *HydroSHEDS* dataset includes stream networks, watershed boundaries and drainage directions. The original *SRTM* data have been hydrologically conditioned to void filling and remove spurious sinks using a sequence of automated and manual procedures (for further details see Lehner *et al.* [43]).

#### 2.1.10. Validation Datasets

Rainfall data from ground weather stations across South America and West Africa have been used for validation purpose. Specifically, *in-situ* annual precipitations are provided by national meteorological institutions participating in EUROCLIMA project [44] (for further information see acknowledgment section).

Rainfall products have been merged and undergone quality test, including the removal of duplication (e.g., rainfall data are available in different source datasets), outliers, and stations having less than 15 years of records.

It is important to underline that the validation dataset is independent of satellite-derived precipitation. The use of EUROCLIMA dataset allows a relatively homogeneous spatial coverage across South America and West Africa in order to assess the reliability of rainfall estimates at high spatial resolution.

All the high resolution rainfall products are validated with the independent meteorological rainfall dataset (hereafter observed precipitation). Three statistical indicators are used for the evaluation of the performance of the high resolution rainfall products, according to the formulae given below: (1) the root-mean-square error—*RMSE*; (2) the absolute error—*Abs. Err.*; and (3) the mean percentage error—*MPE*—*i.e.*, the mean error computed as percentage.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_{downscaled} - P_{observed})^2}{n}} \quad (2)$$

$$Abs.Err. = \frac{1}{n} \sum_{i=1}^n |P_{downscaled} - P_{observed}| \quad (3)$$

$$MPE = \frac{100}{n} \sum_{i=1}^n \frac{(P_{downscaled} - P_{observed})}{P_{observed}} \quad (4)$$

where  $P_{downscaled}$  and  $P_{observed}$  represent respectively downscaled and observed precipitation, while  $n$  is the total number of validation points.

## 2.2. Methods

The downscaling scheme presented in this paper has been developed on the basis of the study of [15]. The basic assumption of this study is the existence of a linear and spatially variable relationship among *MAP*, *DEM* and *EVI*.

The downscaling scheme, shown in Figure 1, is organized through the following steps:

We select the satellite-based rainfall dataset at coarse resolution to downscale, hereafter  $MAP_{sat@CR}$  (where CR stands for “Coarse Resolution”);

- 1) *DEM* and average annual *EVI* are upscaled by pixel averaging from the original fine-scale resolution of 1 km (*i.e.*,  $DEM@1km$  and  $EVI@1km$ ) to the spatial resolution of the satellite-based rainfall dataset (e.g., 25 km for *GPCC*, 5 km for *CHIRP*), hereafter  $DEM@CR$  and  $EVI@CR$ ;
- 2) A Geographically Weighted Regression (*GWR*, see next subsection for further details) is performed to establish a relationship among  $MAP_{sat@CR}$ ,  $DEM@CR$  and  $EVI@CR$ . In essence, the *GWR* is a local form of linear regression described in the Equation (5):

$$MAP_{sat@CR} = \beta_0(u) + \beta_1(u) \cdot DEM@CR + \beta_2(u) \cdot EVI@CR + \Delta@CR \quad (5)$$

where  $\beta_0(u)$ ,  $\beta_1(u)$ , and  $\beta_2(u)$  are the intercept and the slope parameters, varying with location ( $u$ ), while  $\Delta@CR$  is the residual, representing the amount of precipitation that cannot be explained by the model.



If we write the regression without the residual component we obtain:

$$MAP_{pred@CR} = \beta_0(u) + \beta_1(u) \cdot DEM_{@CR} + \beta_2(u) \cdot EVI_{@CR} \quad (6)$$

where  $MAP_{pred@CR}$  is the annual precipitation at coarse resolution predicted/explained using the regression model of  $DEM$  and  $EVI$ . The closer the value of the predicted rainfall to the satellite-based rainfall, the better the performance of the regression and the lower the value of the residual. The  $GWR$  also supplies a per-pixel flag documenting the coefficient of determination of the regression (*i.e.*,  $r^2$ );

- 3)  $GWR$  is repeated excluding  $EVI_{@CR}$ : in this way we re-compute  $MAP_{pred@CR}$ , along with  $GWR$  parameters and a map of coefficient of determinations:

$$MAP_{pred@CR} = \beta_0(u) + \beta_1(u) \cdot DEM_{@CR} \quad (7)$$

Step (4) is performed because in some cases there is not a strong relationship between precipitation,  $DEM$  and  $EVI$ , hence it is beneficial to exclude  $EVI$ ;

- 4) For each cell of the spatial domain, we select the  $MAP_{pred@CR}$ , the intercept and slope parameters of the  $GWR$  associated with the highest coefficient of determination  $r^2$ , as shown in Figure 2;
- 5) We subtract  $MAP_{pred@CR}$ , obtained in step (5), from  $MAP_{sat@CR}$ , obtaining the residual of regression model at coarse resolution,  $\Delta_{@CR}$ , which represents the amount of precipitation that cannot be explained by the regression based on  $DEM$  and  $EVI$ ;
- 6) We downscale  $\Delta_{@CR}$  to 1 km resolution by applying a cubic spline interpolation, and we obtain  $\Delta_{@1km}$ ;
- 7) We downscale the model parameters to 1 km by the nearest neighbour resampling, obtaining thus  $\beta_{0@1km}$ ,  $\beta_{1@1km}$ , and  $\beta_{2@1km}$ ;
- 8) We apply the  $GWR$  model with the  $EVI$  and  $DEM$  at the original spatial resolution of 1 km, obtaining thus the predictive value of annual precipitation with 1 km resolution, hereafter  $MAP_{pred@1km}$ :

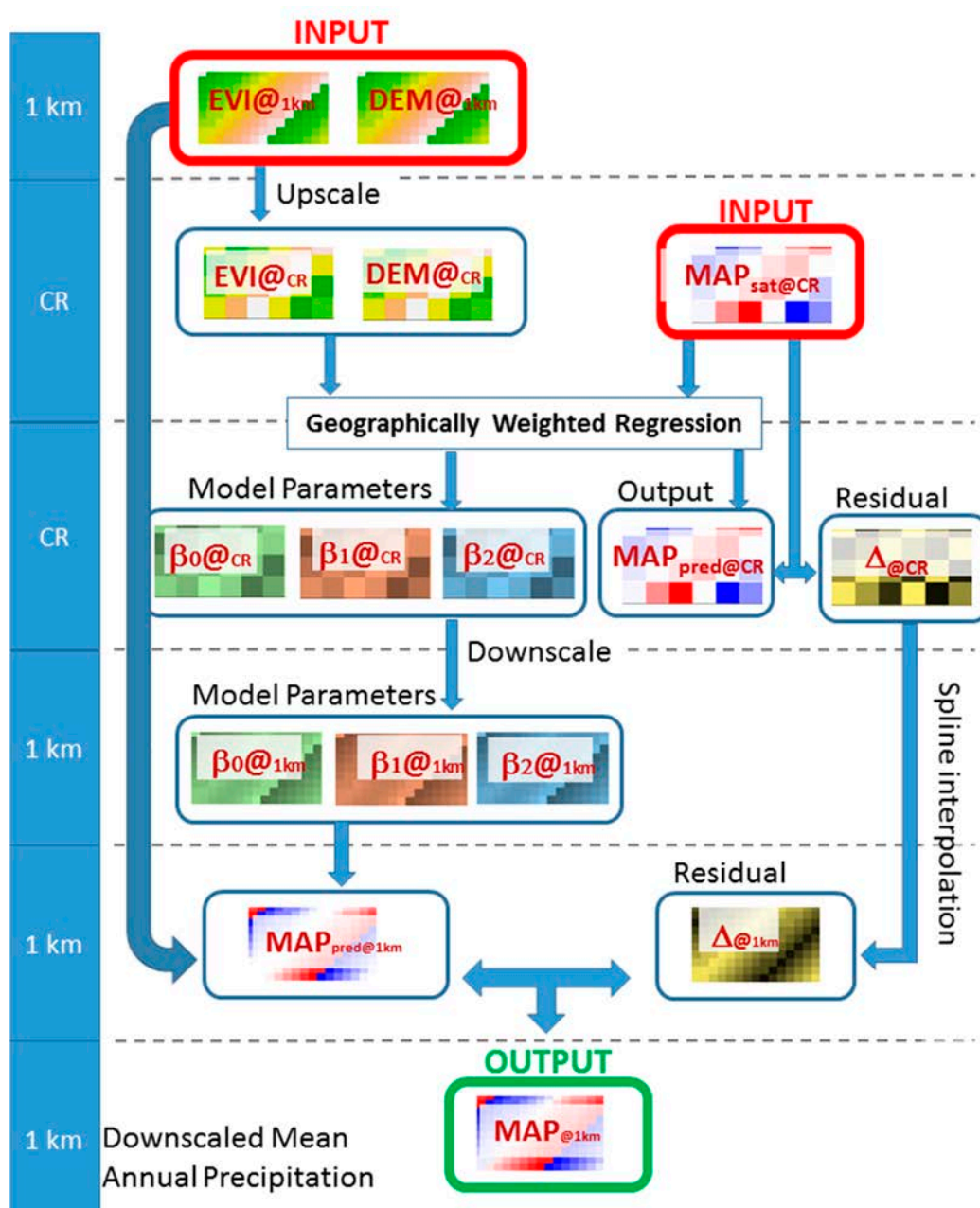
$$MAP_{pred@1km} = \beta_{0@1km}(u) + \beta_{1@1km}(u) \cdot DEM_{@1km} + \beta_{2@1km}(u) \cdot EVI_{@1km} \quad (8)$$

For the cells where  $EVI$  values are masked, the corresponding values of the coefficient  $\beta_{2@1km}$  are set equal to zero;

- 9) By adding this high resolution predictive precipitation data to the high-resolution residual obtained in step (8), we attain the final downscaled  $MAP$  with a 1 km resolution, hereafter  $MAP_{@1km}$ , as described below in Equation (9):

$$MAP_{@1km} = \beta_{0@1km}(u) + \beta_{1@1km}(u) \cdot DEM_{@1km} + \beta_{2@1km}(u) \cdot EVI_{@1km} + \Delta_{@1km} \quad (9)$$

Steps 4 and 5, *i.e.*, the modifications of the  $GWR$ —are introduced to improve the performance of the regression by excluding  $EVI$ . For example, where the land use is fragmented, it may happen that the relationship between  $EVI$  and  $MAP$  is not robust: the Vegetation Index is tied to land use while its relationship with rainfall is weak. In this case, the regression among annual precipitation and  $DEM$  could better explain the data than the regression that also includes  $EVI$ .



**Figure 1.** Flowchart of the scheme used to downscale different satellite-based precipitation datasets. CR on the left bar stands for Coarse Resolution.

The downscaling methodology is implemented for each country of both South America and West Africa, and for each satellite-derived dataset. For South America, we run the downscaling algorithm 65 times (13 countries multiplied by 5 rainfall datasets), 105 times for West Africa (15 countries multiplied by 7 rainfall datasets). Please note that for Africa we have two more datasets (*i.e.*, TAMSAT and RFE) that are Africa-specific.

It is worth noting that downscale schemes employed for the residual and for the parameters of the regression model (*i.e.*, steps 7 and 8) are different, as stated in Immerzeel *et al.*, Jia *et al.* and Chen *et al.* [9,11,15]. The residual is *de facto* a rainfall, thus it makes sense to employ a downscaling scheme such as the spline interpolation that considers the values of neighbouring points. Conversely, parameters of the model cannot be modified freely, since a slight variation might cause large and

unexplained variations in the output of the model. Thus it makes sense the use of a piecewise-constant interpolator such as the nearest neighbour to disaggregate the parameters of the regression model.

Also, note that we use the acronym *CR* for coarse resolution instead of a fixed number throughout the paper: the size of the low-resolution dataset depends on the choice of the original rainfall dataset, ranging from 4 km to 25 km (e.g., *TAMSAT* and *GPCC*, respectively). The proposed methodology supplies a per-pixel flag, ranged in a separate layer. Each flag documents the goodness of fit of the regression (*i.e.*,  $r^2$ ) and the variables employed in the regression (*i.e.*, both *EVI* and *DEM*, only *EVI*, only *DEM*).

The Geographically Weighted Regression is described more fully in the next sub-section.

### Geographically Weighted Regression (GWR)

The relationships among *MAP*, elevation and vegetation is intrinsically different across space: in some areas, precipitation are higher than in others having the same values of *EVI* and *DEM*. The degree to which annual precipitation varies may relate to other neglected effects such as terrain aspect, wind direction, distance from the sea or land cover which exert local downward/upward influence on precipitation.

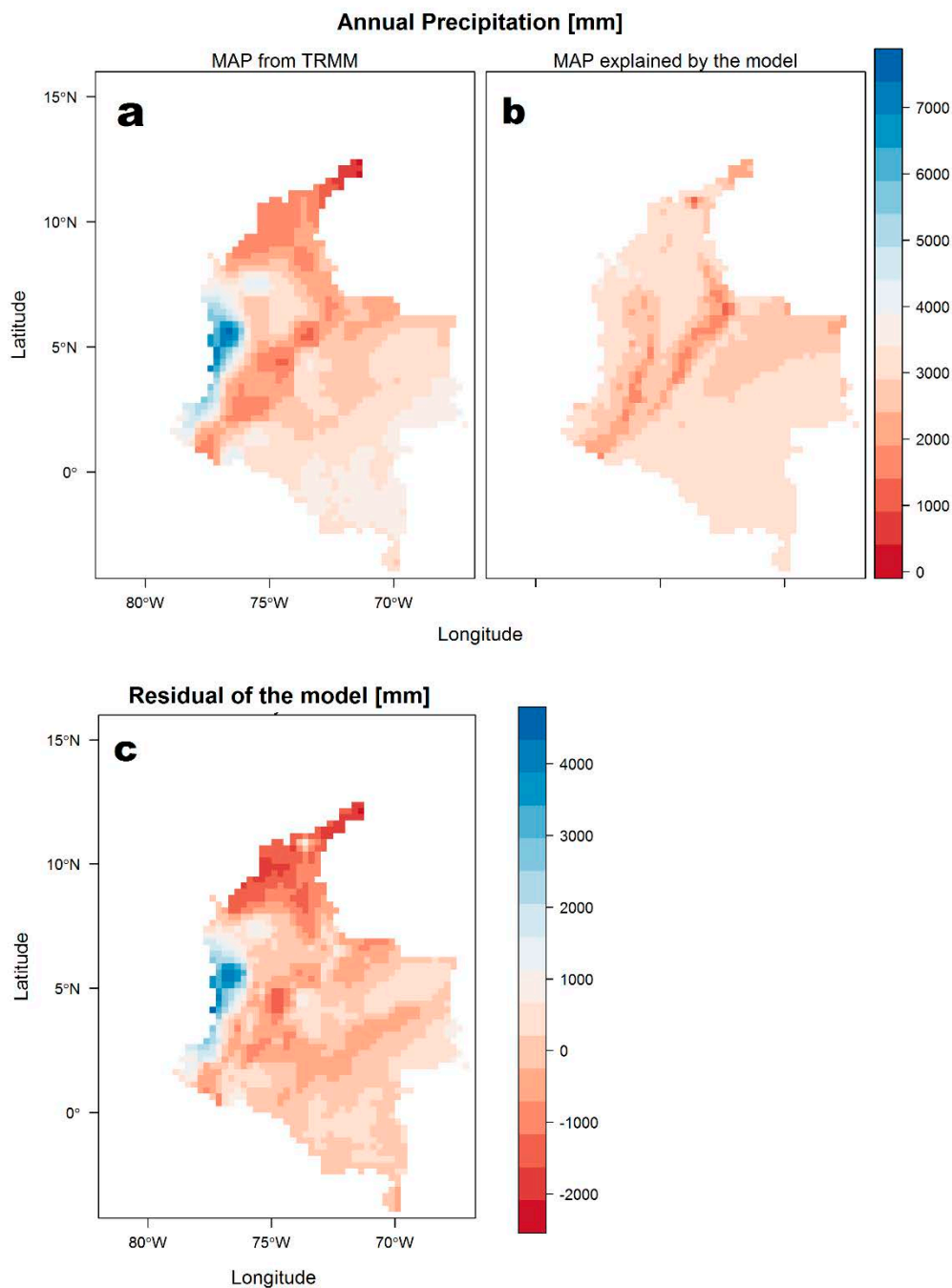
As pointed out in Chen *et al.* [15], even if we find a statistically significant relationship between annual precipitation, *DEM* and *EVI*, if we apply uniformly this relationship over the whole country, we will have a large amount of variance that is not explained by the regression. An insight on the need to characterize spatial heterogeneity is given in Figure 2.

In this example, we establish a global linear regression for the whole Colombia among *TRMM* annual precipitation, *DEM* and *EVI*, assuming that the parameters of the model (*i.e.*, slopes and intercept) are constant over space. The *p-values* computed for the global linear regression are less than 0.01, indicating thus that it is unlikely that no relationship exists between annual precipitation, *DEM* and *EVI*. The first panel shows the precipitation from *TRMM* satellite, whereas the second panel represents the annual rainfall explained by *DEM* and *EVI* with the global model. The third panel represents their difference, *i.e.*, the difference between *TRMM* and modeled rainfall. The spatially homogeneous global regression cannot capture spatial variations in the relationship between *MAP*, *DEM* and *EVI*, much to the detriment of spatially varying relationships. Spatial patterns of annual precipitation are captured by the residual instead of the model itself.

The spatial non-stationarity in this study is modeled by means of the Geographically Weighted Regression (*GWR*) [13], a linear regression where the relationship between variables varies spatially. *GWR* has been initially used as graphical tool for data exploration in social science, economy and urban economics disciplines. Only in the last decade it has been broadly applied in geoscience disciplines. The general form of the *GWR* for one independent variable is expressed in the Equation (10):

$$y = \beta_0(u) + \beta_1(u) \cdot x + \varepsilon(u) \quad (10)$$

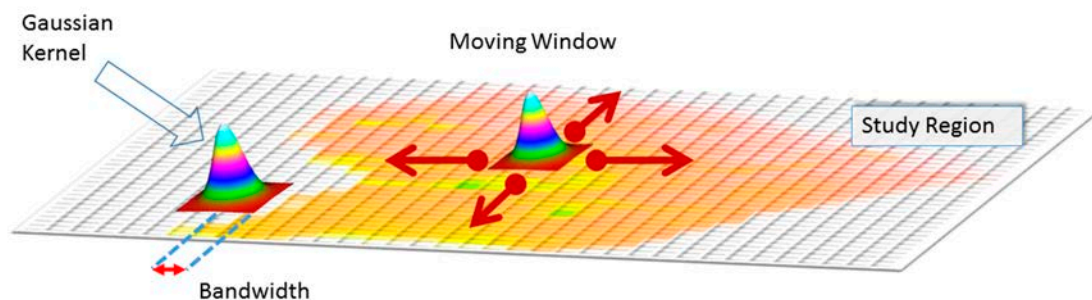
where  $x$  and  $y$  are the independent and the dependent variable,  $\beta_0$  and  $\beta_1$  are the intercept and the slope of the regression, respectively,  $\varepsilon(u)$  is the residual and  $(u)$  are the spatial coordinates.



**Figure 2.** Example of Global regression of annual precipitation for Colombia, where we assume that the relationship among Mean Annual Precipitation (*MAP*), Digital Elevation Model (*DEM*) and Enhanced Vegetation Index (*EVI*) is spatially homogeneous. (a) represents the *MAP* from Tropical Rainfall Measuring Mission (*TRMM*), (b) represents the *MAP* explained by a global model of *DEM* and *EVI*, (c) shows the residual of the model. Note that (a) and (c) exhibit the same spatial patterns.

Instead of remaining the same everywhere, as in the case of the global linear model, the intercept and the slope of the *GWR* vary in terms of locations. In other words, the spatial non-stationarity of *GWR* implies that the same combination of *EVI* and *DEM* may refer to different annual precipitation values in different parts of the study region.

In order to be able to do it, a separate regression is run for each point of the spatial domain, using a spatial kernel that centers on a given point and weights observations subjected to a distance decay function, also called bandwidth, as shown in Figure 3.



**Figure 3.** Geographically Weighted Regression scheme. Geographical weighting is achieved by a fixed Gaussian kernel function with a given bandwidth moving across the spatial domain. Bandwidth determines the rate at which the weights decay around each cell, and reflects the degree of spatial variation: if the bandwidth gets larger, the model will tend to a global regression model.

Separate equations are built for every kernel in the spatial domain, incorporating the dependent and explanatory variables falling within the bandwidth of kernel. As the bandwidth gets larger the *GWR* model approaches the global model. Conversely, as the bandwidth decreases, the *GWR* will increasingly rely on observations closer to the regression point. Basically, the main assumption is that nearer observations have more influence in estimating the local set of regression coefficients than observations farther away.

Without going into much detail, in literature, many different approaches have been proposed to define the shape and extent of this spatial kernel [13]. The most popular for gridded dataset, just as in our case, is the fixed Gaussian kernel. The fixed kernel suits regular gridded datasets, whereas the adaptive kernel is suitable for sparse and irregular configurations. However, whichever weighting function is used, the result is highly sensitive to the bandwidth, whose choice is more important than the shape of the kernel [45].

Bandwidth optimization provides a way of choosing bandwidth that makes optimal trade-off between bias and variance [46]: the smaller the bandwidth, the larger the variance but the lower the bias; conversely, the larger the bandwidth, the larger the bias but the smaller the variance. The optimum kernel bandwidth for the *GWR* can be found by minimizing some model fit indicators, such as a leave-one-out cross-validation (*CV*) [47] or the Akaike Information Criterion (*AIC*) [48]. However, the latter is excluded since it significantly increases the computational burden of the fitting process.

In our study we employ a fixed Gaussian kernel and the *CV* criterion is used to determine the “optimal” value for the bandwidth for each country: in practice we employ the bandwidth associated to the lowest *CV*. However, a test study over two small countries such as French Guyana and Benin (representative of South America and West Africa, respectively), indicates that the results of bandwidth minimization obtained with *AIC* and *CV* method are absolutely identical.

In fact, there is a calculation limitation: we cannot always employ the *GWR* for each country as a whole because the computational burden of this implementation becomes a significant barrier. For this

reason, each country with more than 250,000 cells (1,250,000 km<sup>2</sup> with *CHIRP* dataset at 5 km resolution, e.g., Brazil or Argentina) is partitioned into smaller sub-regions for whom we apply the *GWR*. In practice we divide the spatial domain in blocks of adjacent cells, and we sequentially run the *GWR* over these blocks of contiguous pixels. This partition is subjectively-decided but it allows to avoid computation limitation and it cuts down calculation times.

In order to quantify the effectiveness of this partition, that may introduce some errors, a test case has been designed over the French Guyana with the *CHIRP* precipitation. This country has been chosen because it represents an optimal size (*i.e.*, 83.534 km<sup>2</sup>) to: (1) perform the *GWR* on the entire country, (2) perform the *GWR* on its partition into two sub-regions which are statistically significant, and (3) evaluate the difference of the modeled rainfall with the two different methods. Given the elongated shape of the French Guyana, the partition was carried out by dividing the country into two blocks along an imaginary parallel/horizontal line.

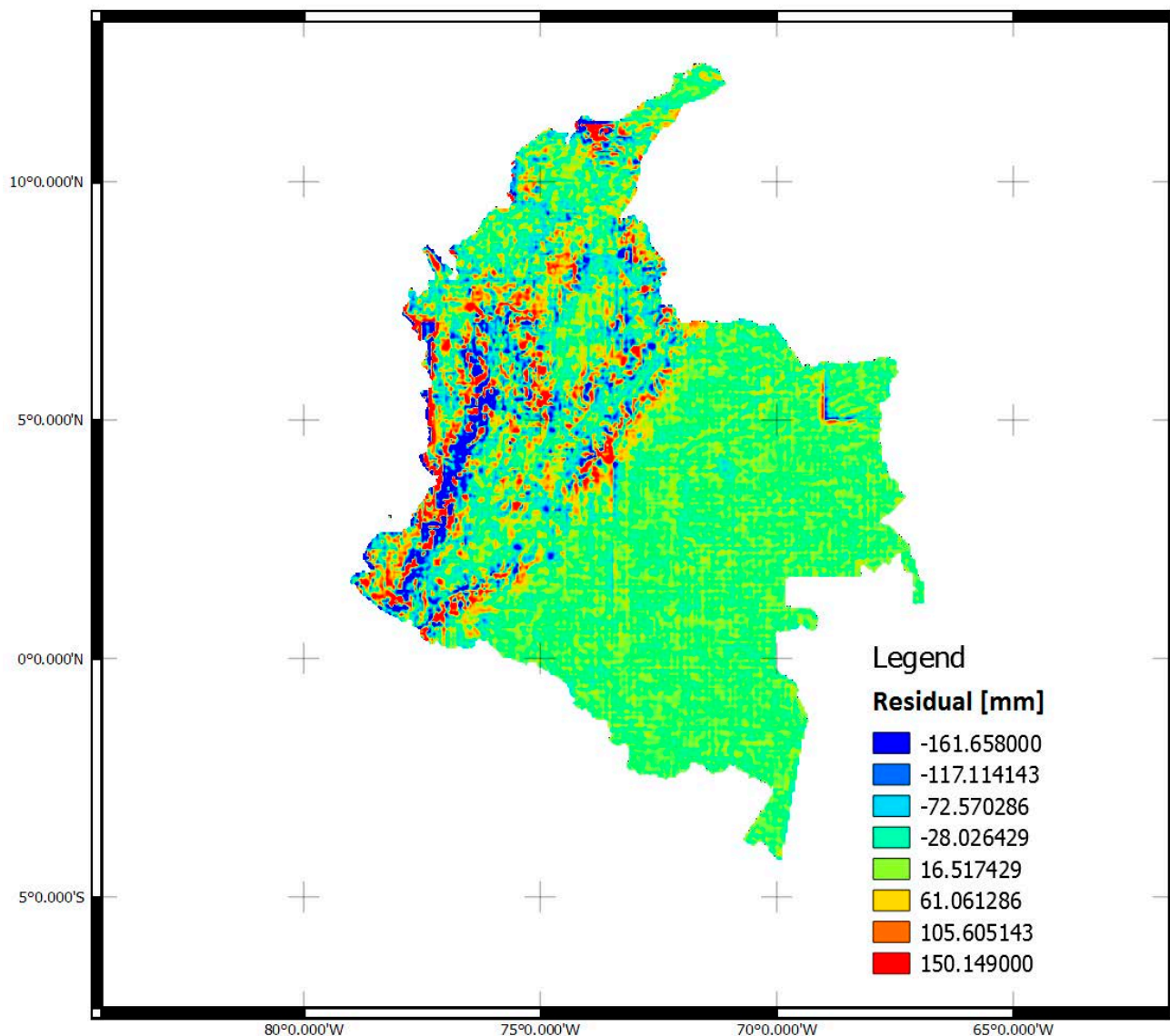
The differences of annual precipitation predicted with a *GWR* performed (1) on the entire country and (2) on the two subsets indicate that no significant biases are introduced: the absolute error is about 0.6 mm. A possible explanation is that the area of influence of the *GWR* is strictly limited to the neighborhood. Therefore the *GWR* is “local” and the net effect of increasing the size of the spatial domain lies only in computation times, while the difference pertaining to the results is negligible; and the benefits of the proposed methodology can be considered to be far beyond their potential drawbacks.

Clearly, *GWR* is a powerful spatial downscaling technique, but there are other methods that may fit certain situations better. Specifically, since information on rainfall, elevation and vegetation is available at all grid nodes (*i.e.*, all raster nodes or pixels in the maps), we could have used Kriging with External Drift [49] or Regression-Kriging [50] methods instead. Example of Regression-Kriging of precipitation across China is provided by Teng *et al.* [51]. However, Kriging techniques are computationally demanding and often cannot be used [52]. In our case, the computation of predictions and errors using Kriging for entire countries might need lots of memory and processing time and it is not feasible. Furthermore, *GWR* presents appealing and unique features such as the quantification of the impact of each of the independent variables of the model over space, in addition to feasible computation times. In other words, with the proposed methodology it is always possible to quantify the independent variable coefficients over space and to map the changes of their influence on the dependent variable.

In order to quantify the performance of the *GWR*, Figure 4 shows the residual of the *GWR* method, *i.e.*, the difference between remotely sensed and explained precipitation across Colombia using *CHIRP* dataset.

The residual of the model ranges from −160 to 150 mm, whereas the average value is equal to 3.5 mm. The highest values are generally located where annual precipitation is higher and where the orography is more complex, e.g., the Andean range. It is also possible to see a kind of tiled texture of the residual that depends on the bandwidth and on the number of independent variables used in the regression. In fact, if the regression for a given pixel uses only the elevation, while in the neighborhood also *EVI* is employed, the residual associated to the pixel will “stand out” against the background. Moran test excludes autocorrelation for the residual across the entire country but the

upper-right zone (*i.e.*, the square-shaped feature) and the slopes of Andean range (*i.e.* the bluish and the reddish areas in the central-left zone).

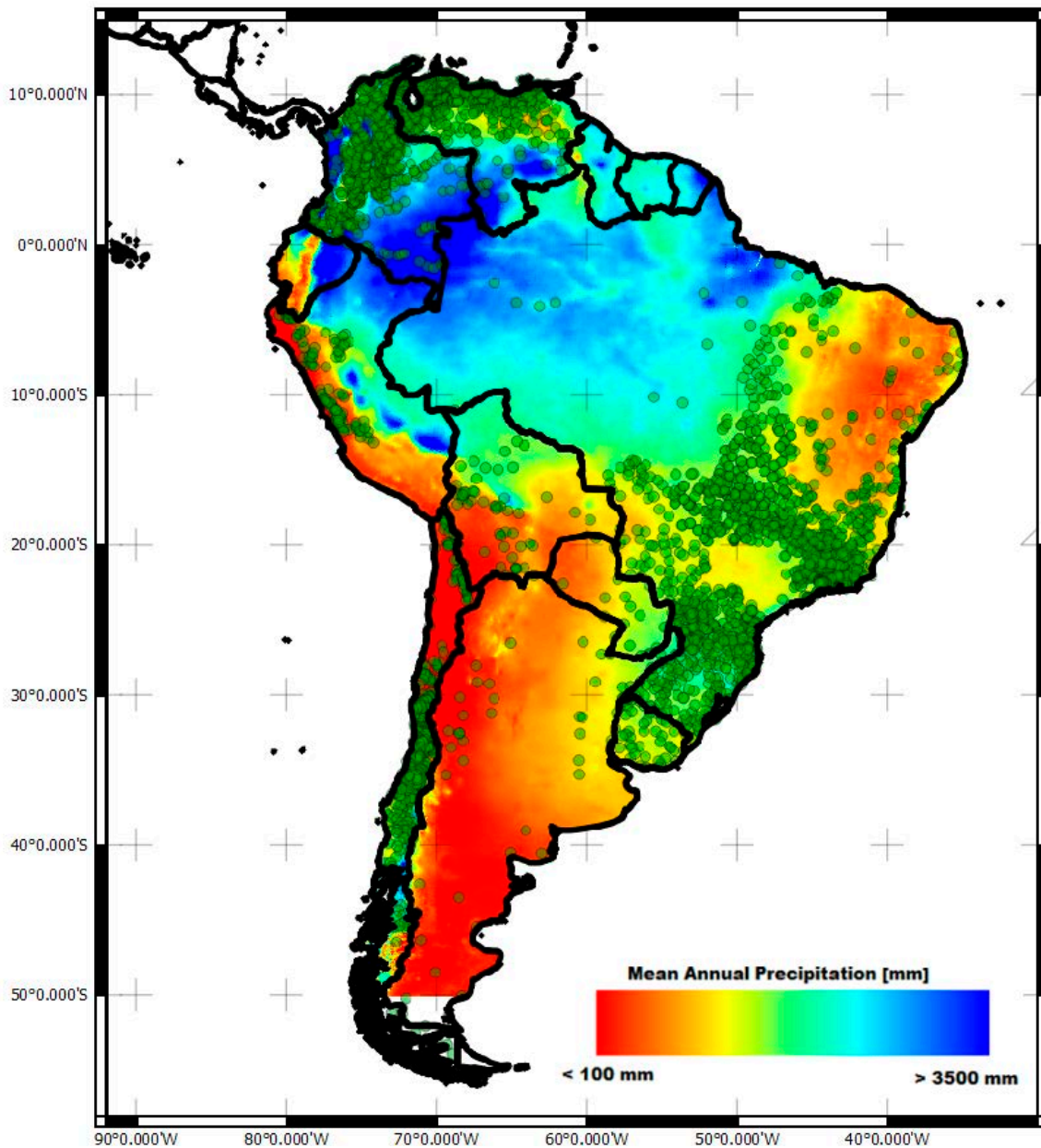


**Figure 4.** Residual of the Geographically Weighted Regression (*GWR*) method, *i.e.*, the difference between remotely sensed and explained precipitation across Colombia with *CHIRP* dataset.

### 3. Results

#### 3.1. South America

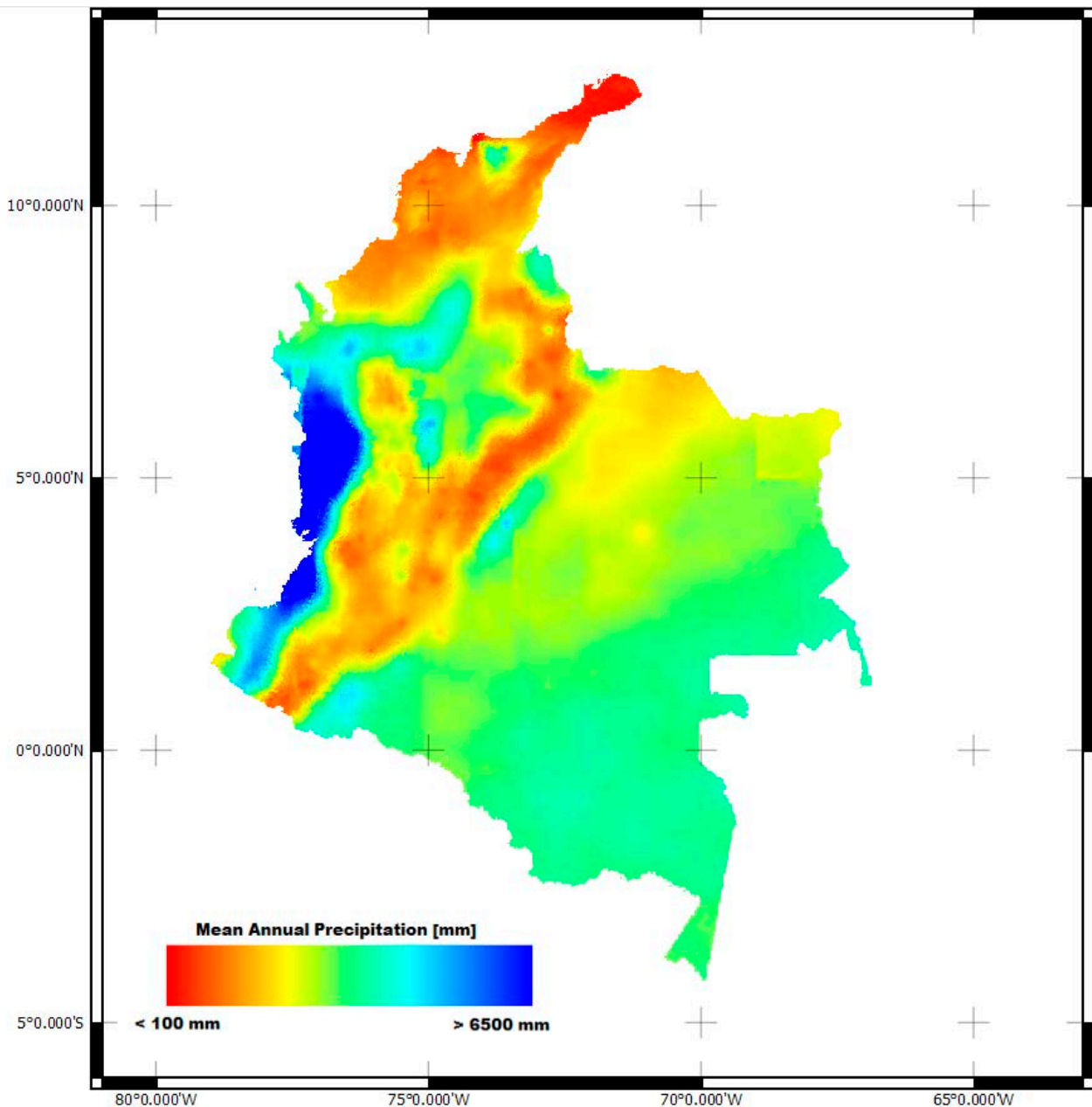
Figure 5 shows, by way of example, the output of the downscaling scheme across South America for *CHIRP* dataset. The number of available weather ground stations used for validation purpose, shown with the green dots is equal to 3353.



**Figure 5.** Mean Annual Precipitation across South America using the *CHIRP* dataset. Green points represent the rain gauge stations against with the high resolution *MAP* product has been validated.

Figure 6 shows the output of the downscaling scheme for *CHIRP* dataset across Colombia.



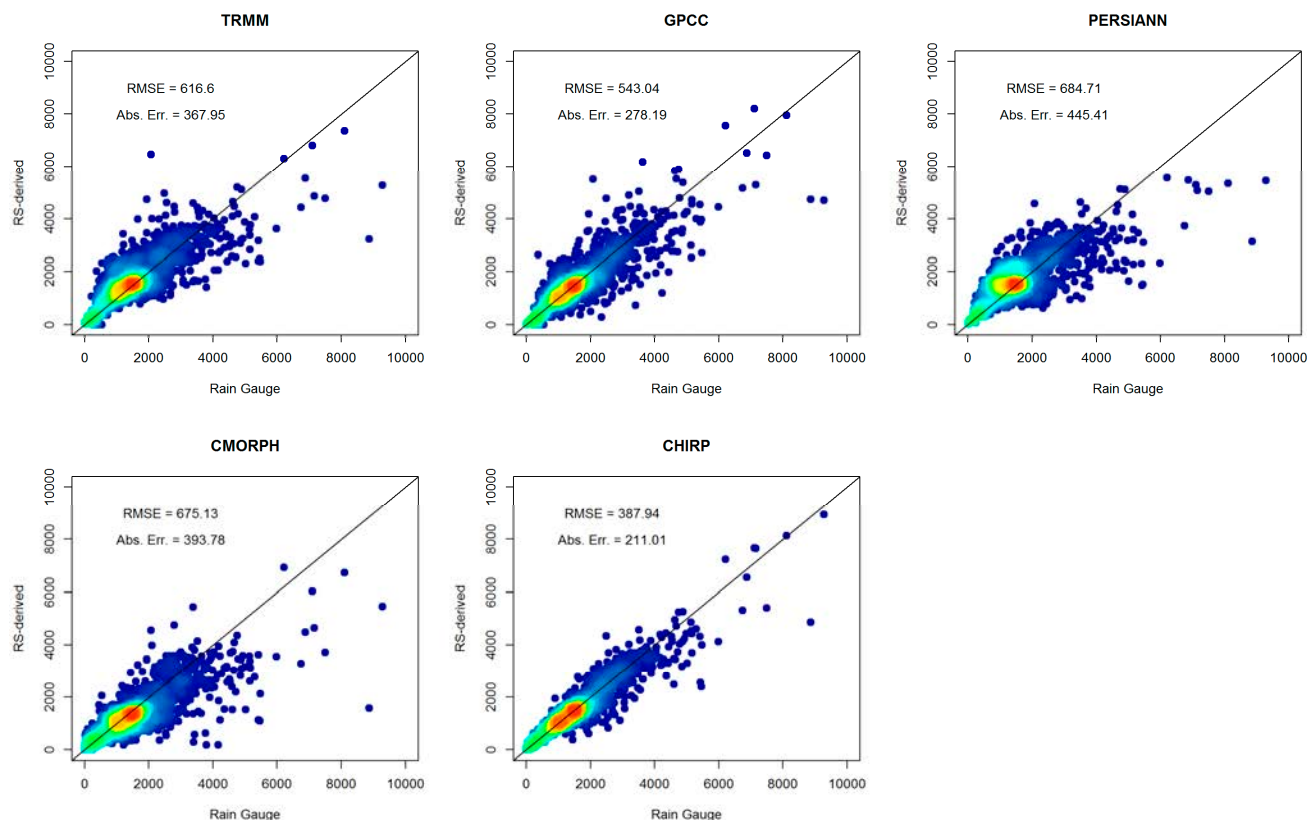


**Figure 6.** Mean Annual Precipitation across Colombia using the *CHIRP* dataset.

Extreme values of color bars at country level (*i.e.*, Figure 6) are different from those at continental level (*i.e.*, Figure 5) since the upper values depend on the statistical distribution (*i.e.*, percentiles) of the rainfall. It is possible to notice on the upper-right zone of Figure 6, a square-shaped region due to the same square-shaped feature in the coarse resolution *CHIRP* map. Also, note that the map of South America is cut at 50° S due to the spatial extent of the original *CHIRP* dataset.

Other than that, the map presents a smooth texture. Anyway, edges and high gradients depend on the original low-resolution dataset used in the downscaling procedure: if there is an abrupt change of precipitation, our method has to preserve this spatial pattern in the high resolution dataset.

Figure 7 shows the scatterplot of the difference between observed and downscaled precipitation across the entire South America.

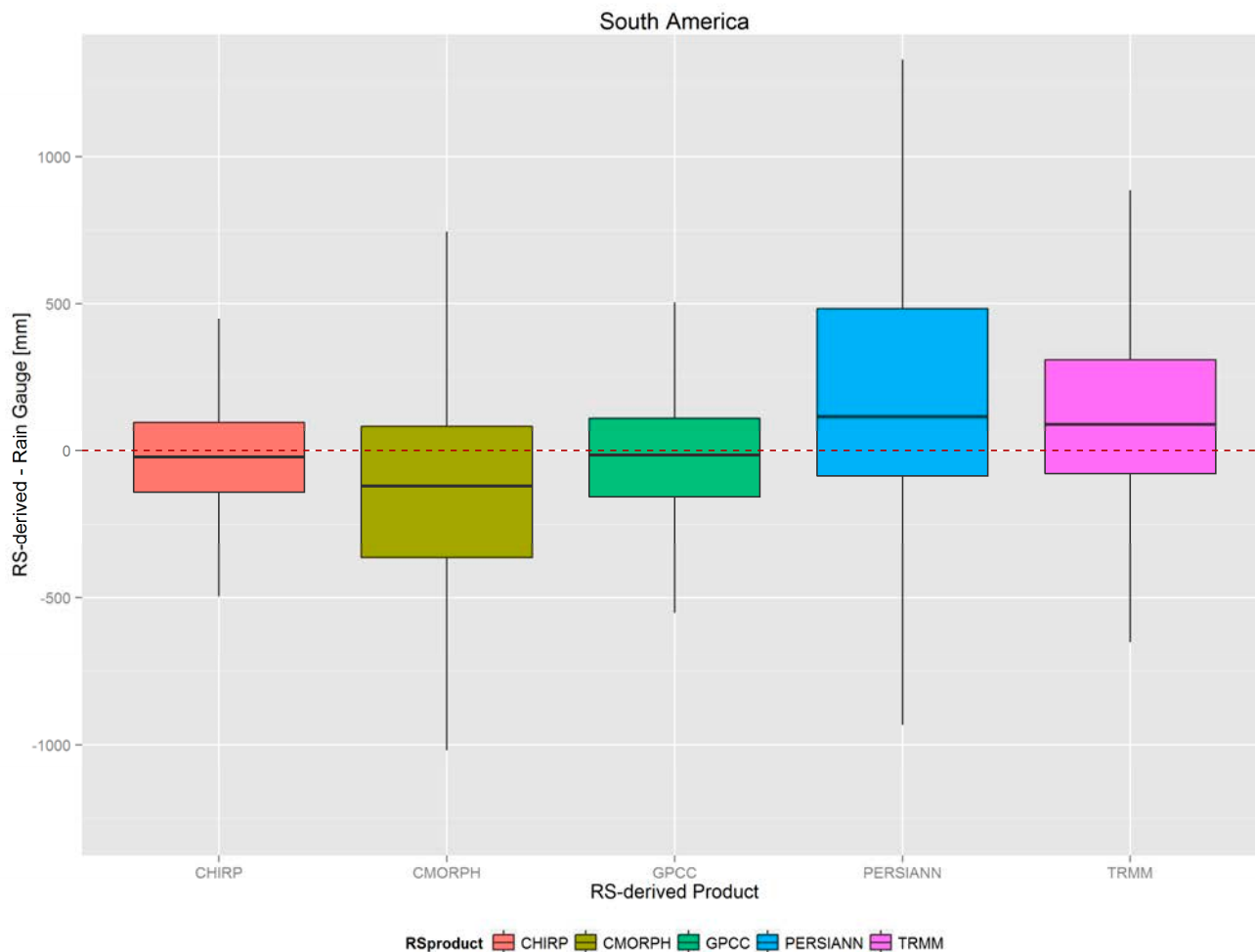


**Figure 7.** Scatterplot of Rain gauges *versus* Satellite-derived (RS-derived) precipitation for the entire South America. Reddish (bluish) areas in the plots indicate a higher (lower) density of sample points.

For each panel, the insets show the difference in annual precipitation for each satellite-derived product *versus* rain gauge measurements. A perfect matching would collapse all points on the 1:1 line. The color scale on each plot indicates the occurrence of the two distributions; the blue color corresponds to a low density of sample points whereas the red colour indicates the highest density. Each panel in this figure also reports the root mean squared error (*RMSE*), and the absolute error (*Abs. Err.*) between satellite-derived and *in situ* records. *CHIRP* exhibits the best correlation, whereas *CMORPH* and *PERSIANN* provide the highest *RMSE* and *Abs. Err.* The *RMSE* varies in the two cases from 388 mm (*CHIRP*) to 685 mm (*PERSIANN*). Less than outstanding results come from *TRMM* that positions itself between *PERSIANN* and *CHIRP*. *GPCC* shows good results, especially in terms of *Abs. Err.*

Box-and-Whisker plot of Figure 8 summarizes the difference between downscaled and observed precipitation across South America for different satellite-derived products, showing the statistical distribution of the difference.

Results confirm our previous findings: *CHIRP* and *GPCC* exhibit the best correlation. Good results come also from *TRMM* and *CMORPH*, whereas *PERSIANN* displays the highest disagreement. *CMORPH* tends to underestimate, while *PERSIANN* tends to overestimate precipitation.

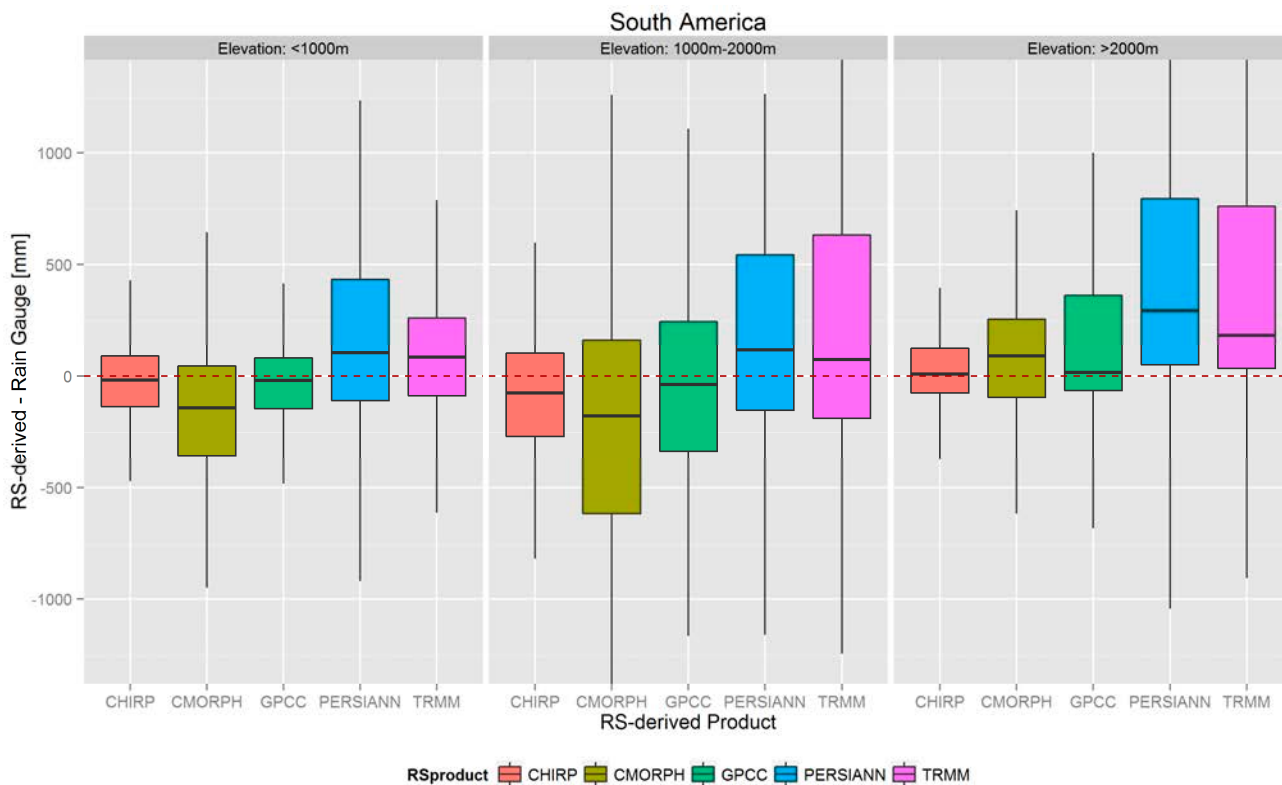


**Figure 8.** Box-and-Whiskerplot of the difference between downscaled and observed precipitation across the entire South America for different RS products.

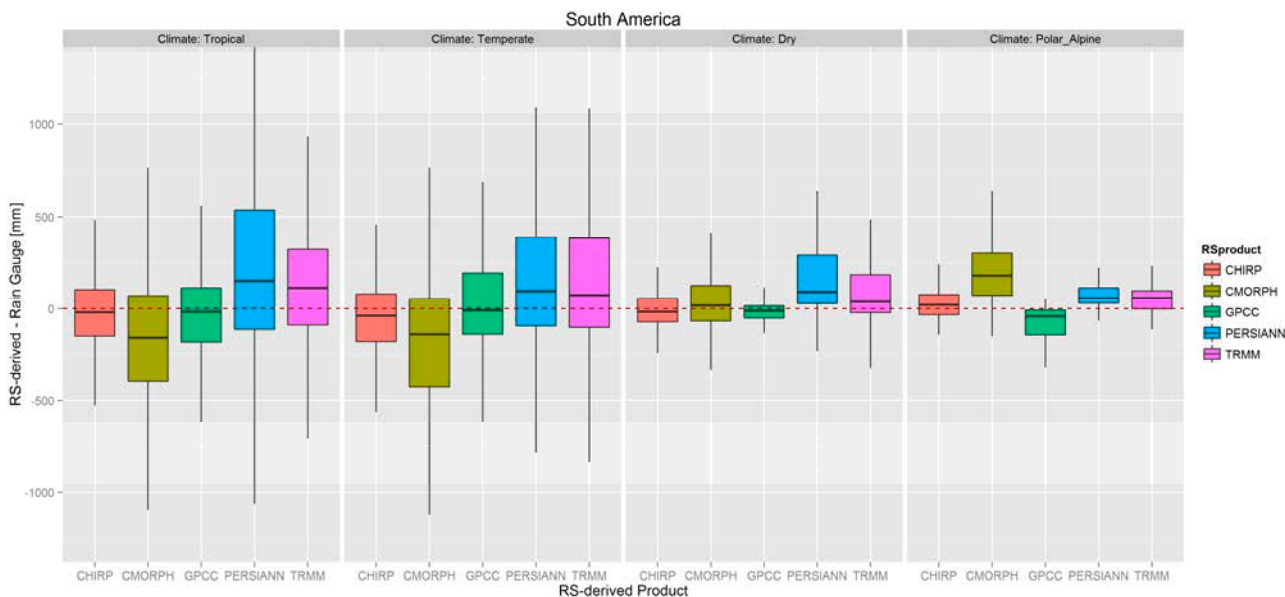
Figure 9 provides a summary of the difference between downscaled and observed precipitation with respect to different elevation classes (*i.e.*, the elevation of rain gauges).

Since most of rain gauges are located below 1000 m asl, good results for that class (especially for *CHIRP*, *CMORPH*, *TRMM* and *GPCC*) lend confidence to the satellite-derived products. Rain gauges located between 1000 and 2000 m asl present a large spread. Also the “above 2000 m” class presents notably bad statistics, with the exception of *CHIRP* and *CMORPH*. The interquartile range (*i.e.*, the “width” of the box) of this elevation class indicates a general overestimation of RS-derived rainfall. It is also possible to observe that *CHIRP* achieves better results than *GPCC* for the “above 2000 m” and the “1000–2000 m” classes.

Figure 10 shows the difference between downscaled and observed precipitation with respect to different climate classes, using the Köppen climate classification [53].



**Figure 9.** Box-and-Whiskerplot of the difference between downscaled and observed precipitation for different elevation classes across the entire South America.

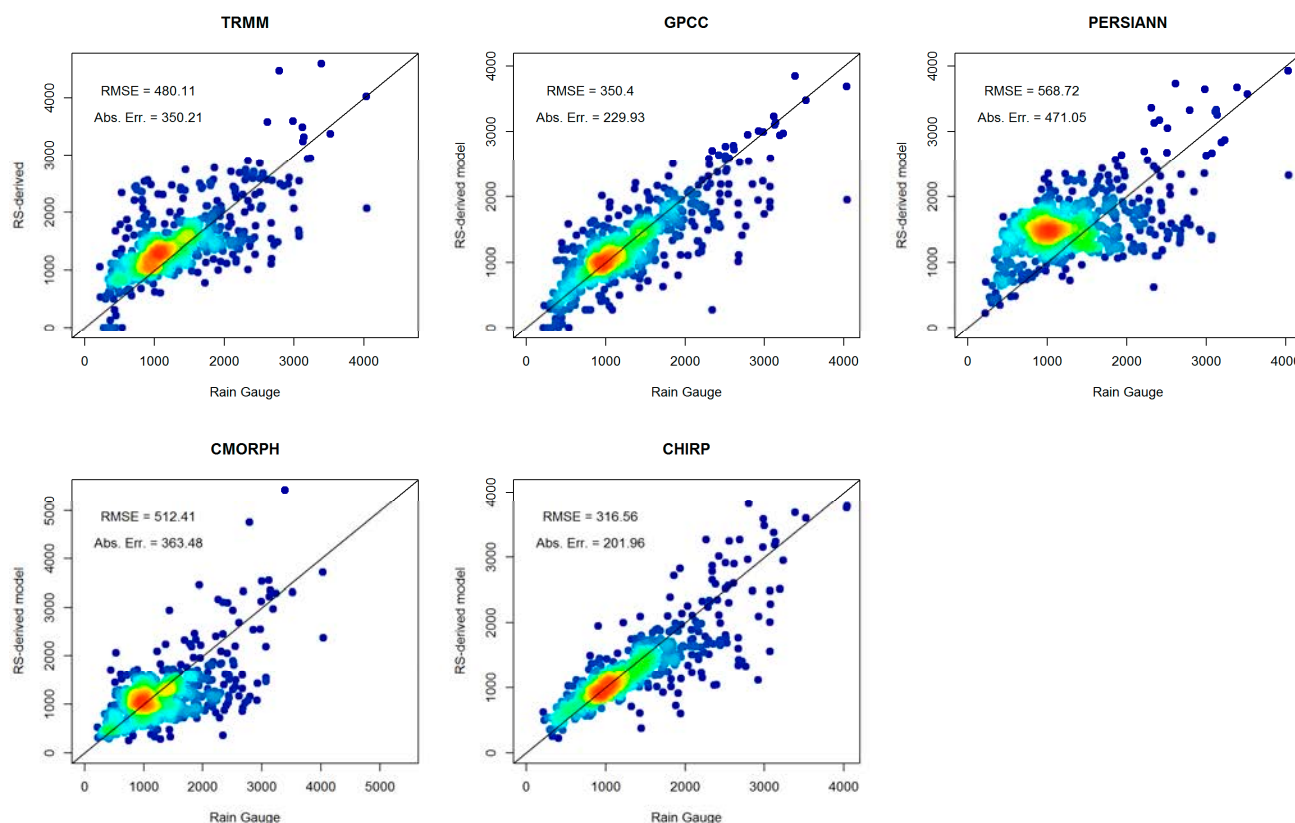


**Figure 10.** Box-and-Whiskerplot of the difference between downscaled and observed precipitation for different climate classes across the entire South America.

Here as well, *CHIRP* and *GPCC* exhibit the best correlation, whereas *PERSIANN* exhibits the highest difference. *CHIRP* achieves slightly better results than *GPCC* for Tropical and Polar/Alpine climate classes, the other way around for Dry climate. *CMORPH* shows notably good results for Dry climate, not outstanding results for Tropical and Temperate climates, and poor results for Polar/Alpine

climate (it does not contain the zero line inside the interquartile range). Both *PERSIANN* and *TRMM* present bad results and tend to underestimate rainfall. However, *TRMM* achieves better results than *PERSIANN* in terms of interquartile and extreme ranges, with the exception of the Temperate climate, and the zero line is always inside the interquartile range.

The scatterplot analysis performed at continental scale, has been carried out also at country scale: Figure 11 shows the scatterplot of the difference between observed and downscaled precipitation for Venezuela.



**Figure 11.** Scatterplot of Rain gauges vs. Satellite-derived precipitation for Venezuela.

*CHIRP* exhibits the best correlation, whereas *PERSIANN* displays the highest disagreement. The *RMSE* varies in the two cases from 317 mm (*CHIRP*) to 568 mm (*PERSIANN*).

Table 2 summarizes the main statistical indicators.

**Table 2.** Statistical indicators of Satellite-derived dataset versus rain gauge stations across nine validation countries. Numbers in parentheses in the first row refer to the number of rain gauge station employed for each country.

	<i>Argentina</i> (46)			<i>Bolivia</i> (52)			<i>Brazil</i> (1339)		
	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>
TRMM	162.1	95.4	15.0	378.8	251.0	36.2	188.1	130.2	0.3
GPCP	104.0	54.0	-1.9	219.3	149.6	13.5	145.4	103.5	-0.7
PERSIANN	255.5	168.9	44.2	505.0	347.1	45.8	221.3	155.5	0.8
CMORPH	134.4	89.0	12.5	499.0	239.2	9.2	248.9	195.4	-10.1
CHIRP	97.7	79.2	-12.9	330.7	150.9	-5.0	180.9	126.4	-2.2

Table 2. Cont.

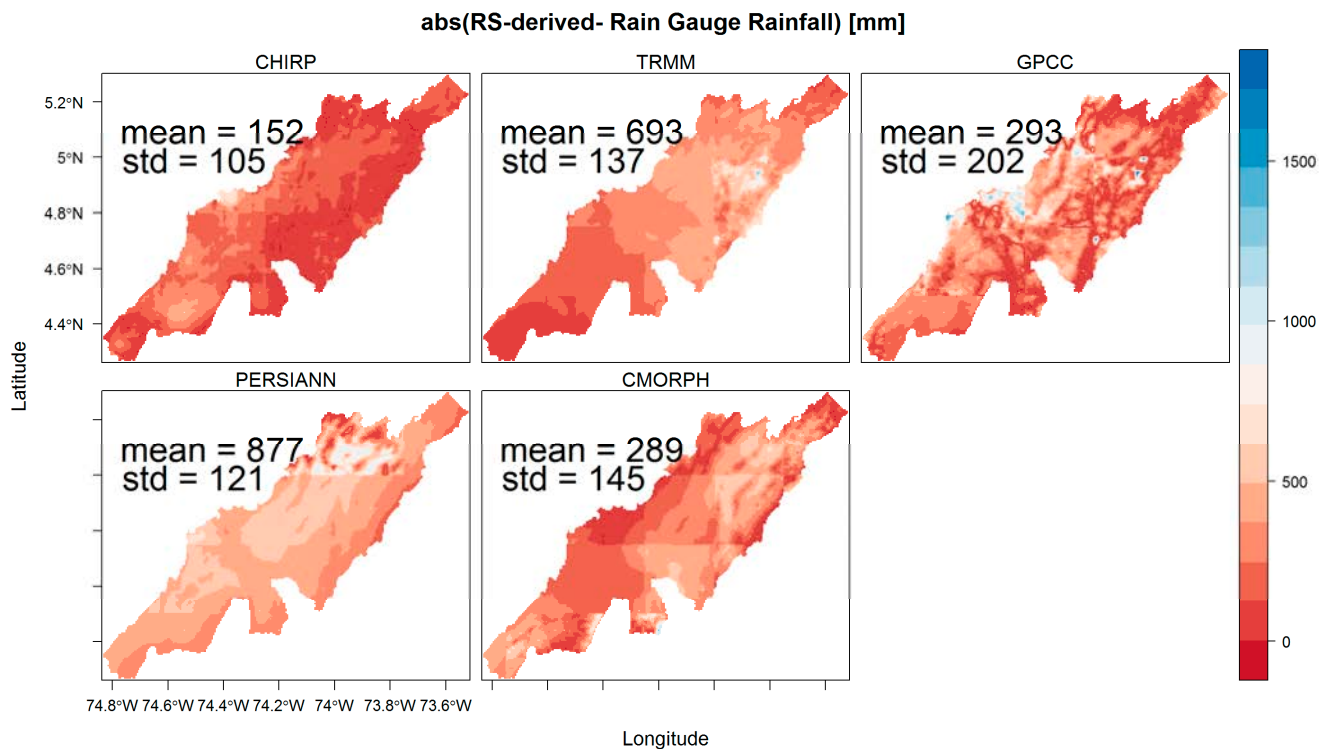
	Chile (344)			Colombia (683)			Paraguay (18)		
	RMSE	Abs. Err.	MPE	RMSE	Abs. Err.	MPE	RMSE	Abs. Err.	MPE
TRMM	357.9	193.5	25.1	955.6	639.0	26.1	105.4	89.2	3.8
GPCC	381.7	197.1	-4.8	871.7	502.6	5.5	60.0	44.3	-0.1
PERSIANN	523.1	284.7	30.0	1010.6	698.0	27.6	154.9	123.6	5.8
CMORPH	466.4	318.9	66.6	1034.7	623.4	-11.2	139.8	100.1	-4.3
CHIRP	279.8	160.4	13.8	570.0	311.9	1.2	86.7	67.0	-5.0
	Peru (89)			Uruguay (47)			Venezuela (743)		
	RMSE	Abs. Err.	MPE	RMSE	Abs. Err.	MPE	RMSE	Abs. Err.	MPE
TRMM	399.9	264.7	23.8	170.7	150.5	11.2	480.1	350.2	23.9
GPCC	354.9	244.5	16.8	103.2	75.8	-4.8	350.4	229.9	1.8
PERSIANN	440.5	353.7	72.1	166.6	146.5	11.6	568.7	471.0	37.1
CMORPH	542.2	418.3	-44.2	90.0	70.8	1.1	512.4	363.5	0.6
CHIRP	472.8	353.2	-29.7	105.1	86.4	4.3	316.6	202.0	1.1

Also, numbers in parentheses in the first row quantify the number of rain gauge stations employed for each country. It is possible to observe that Brazil, Colombia and Venezuela have dense weather networks.

Generally, results pertaining to *CHIRP* and *GPCC* are consistent with previous findings of Figure 7 and Figure 11. The highest agreement is reached for *CHIRP* and *GPCC* for all the test countries but Uruguay. *GPCC* outperforms *CHIRP* for Peru. *CMORPH* on one hand gets reasonably good results for Uruguay, on the other hand gets notably poor results for the other countries. Across Venezuela, a country with a high number of rain gauges, it is possible to observe that *CHIRP* exhibits the highest correlation, conversely, *CMORPH* the lowest one. Note that the worst results in terms of validation at continental scale comes from Colombia that exhibits the worst set of statistical indicators, increasing thus the *RMSE* and the absolute errors. This is probably due to: (1) the highest complexity, (2) the highest values (*i.e.*, up to 10,000 mm of annual precipitation), and (3) the spatial variability of rainfall fields across Colombia, also visible from the Köppen Geiger climate classification [53]. Conversely, Brazil, the country with the highest number of rain gauges, maintains reasonably good results with a *RMSE* around 200 mm and the lowest *MPE*.

Figure 12 provides a “local scale” example of the performance of the RS-derived rainfall products across the Bogotá river basin, where the capital city of Colombia is located.

Bogotá river basin has been chosen because of: (1) the complex orography (the altitude ranges from 50 m to 3000 m asl), (2) the presence of around 50 rain gauges with precipitation records from 1975 onward, and (3) the high spatial variability of annual precipitation, ranging from 500 to 1300 mm. Figure 12 shows the absolute value of the difference between satellite-derived product and observed rainfall. The latter has been interpolated with Ordinary Kriging technique. Annotations in the maps indicate the mean and the standard deviation of the absolute value of the difference.

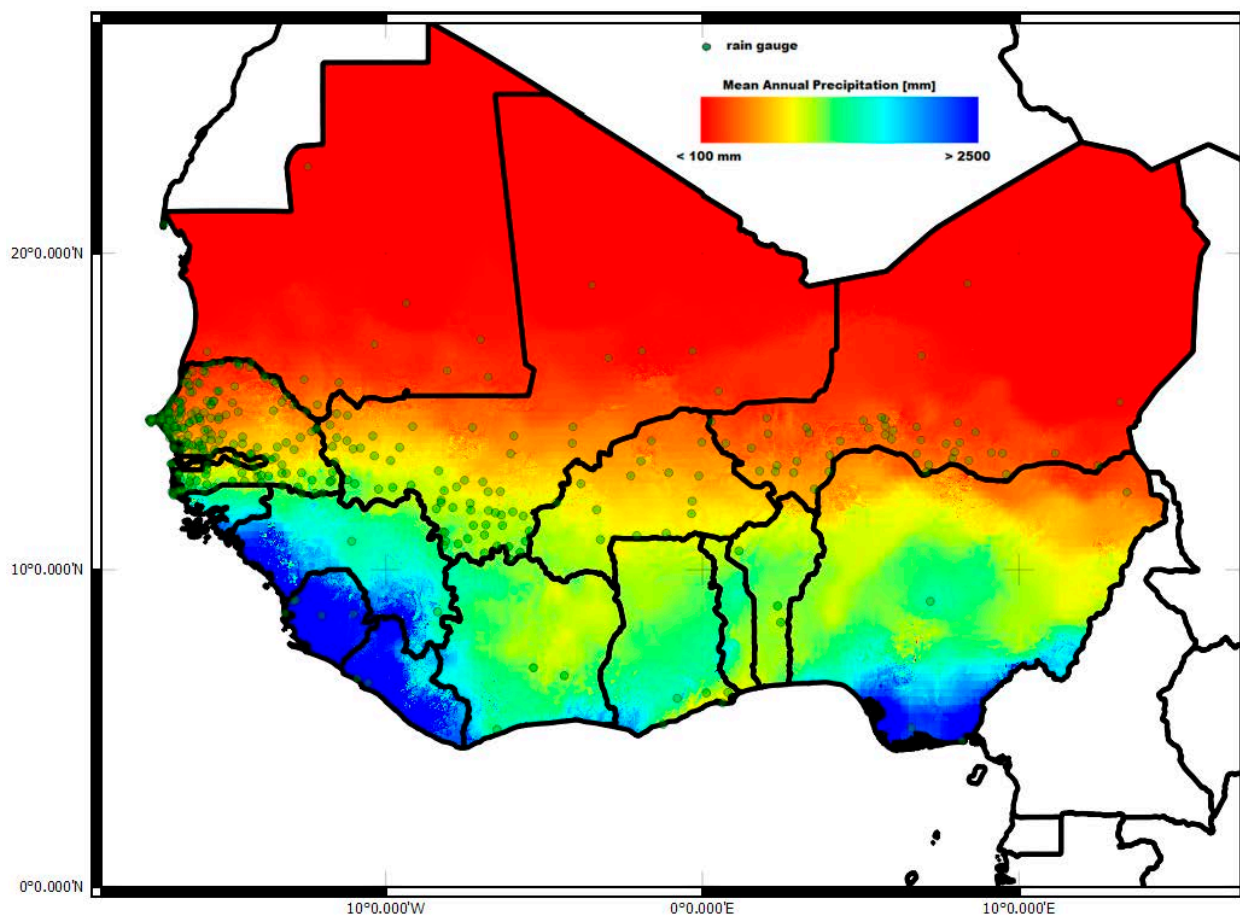


**Figure 12.** Example of difference between for the Bogotá river Basin in Colombia. Maps represent the absolute value of the difference between RS-derived product and observed rainfall, whereas the text annotations indicate average (*mean*) and standard deviation (*std*) values of the latter.

It is possible to observe isolated spots with a difference greater than 1500 mm, mainly located in the most elevated areas of the basin. These errors are probably due to the complex orography and the high spatial variability of rainfall. *CHIRP* exhibits the best performance, followed by *CMORPH* and *GPCC*. *CMORPH* confirms our previous findings, achieving good results for the “above 2000 m” elevation class. *TRMM* and *PERSIANN* show the highest divergence, with average errors up to 650 and 850 mm, respectively. Squared structures in the texture of the maps, mainly noticeable for *CHIRP*, *TRMM* and *CMORPH*, depend on the coarse resolution of the original rainfall products. Note that the interpolation technique used for the rain gauges (*i.e.*, Ordinary Kriging) does not affect the results: spatial patterns of the difference between satellite-derived and observed precipitation—and the corresponding statistics—do not depend on the interpolation method because of the high density of gauge stations within the basin.

### 3.2. West Africa

Figure 13 shows, by way of example, annual precipitation over West Africa using *GPCC* and *TRMM*, respectively.



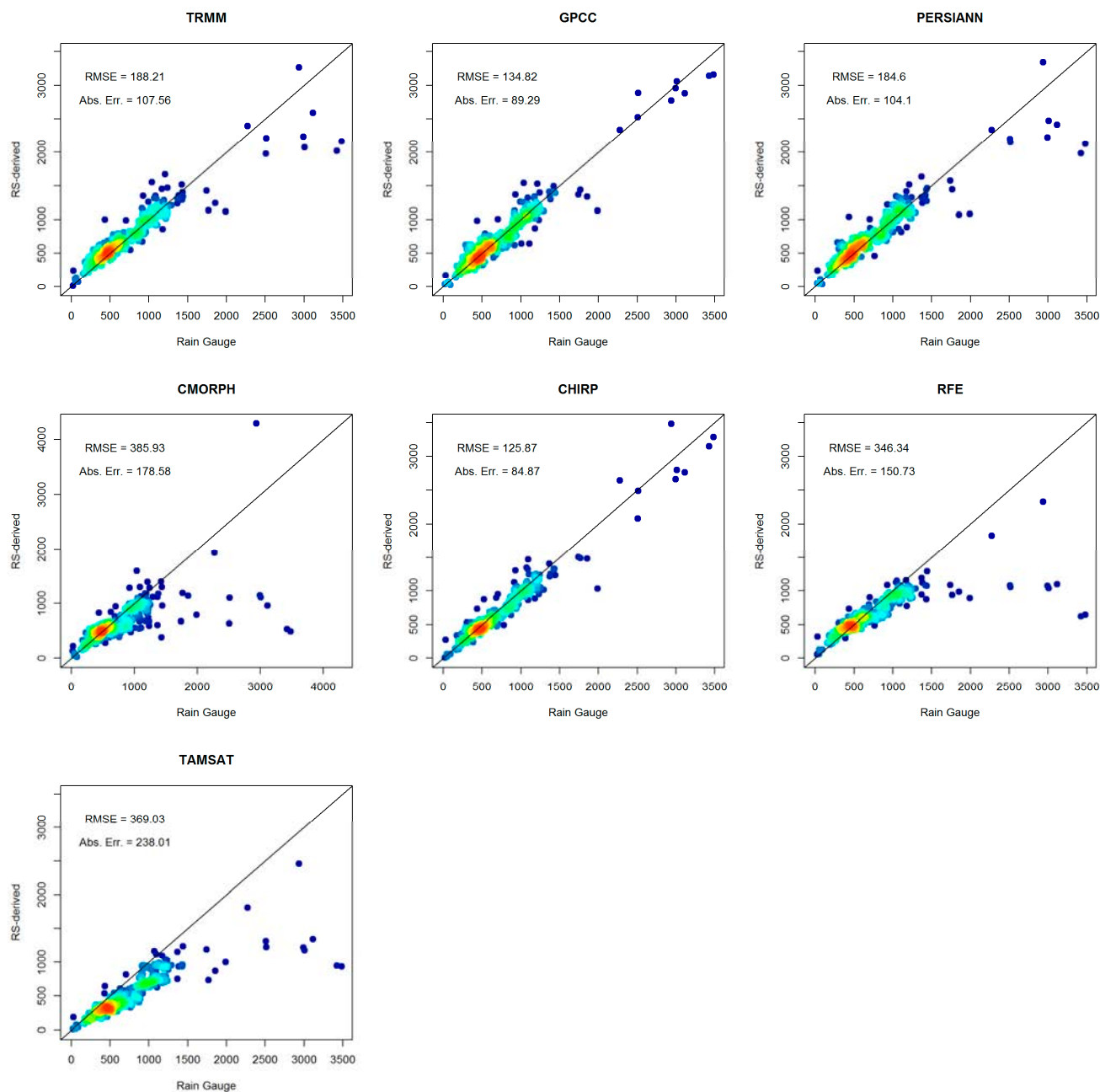
**Figure 13.** MAP of West Africa at 1 km using GPCP dataset. Green points represent the rain gauge stations against with the high resolution MAP product has been validated.

Clearly, the spatial patterns of MAP are different from each other, in the same way as the original low resolution datasets. In these figures, especially across the Sahel region, it is possible to notice a kind of tiled texture. The small squared structures are nothing more than areas where there is a statistical relationship between MAP and DEM and EVI, surrounded by areas where there is a relationship only among MAP and DEM (see points 4 and 5 in subsection 2.2). These irregularities in the texture of the high resolution maps are the drawbacks of the proposed methodology.

Whilst for South America we have several national weather service/networks, West Africa has one of the least developed weather network [54]. The number of available weather ground stations, shown with green dots in Figure 13 is equal to 344. Their spatial distribution is quite uneven, mainly concentrated across Senegal and rare across desert regions and the Gulf of Guinea. This is the reason why the country-level validation has been conducted only for the five countries with more than 10 rain gauge stations in order to have robust statistics.

As for South America, all the high resolution rainfall products are evaluated against the in-house national meteorological rainfall products. Figure 14 shows the scatterplot of the difference between observed and downscaled precipitation for the entire West Africa.

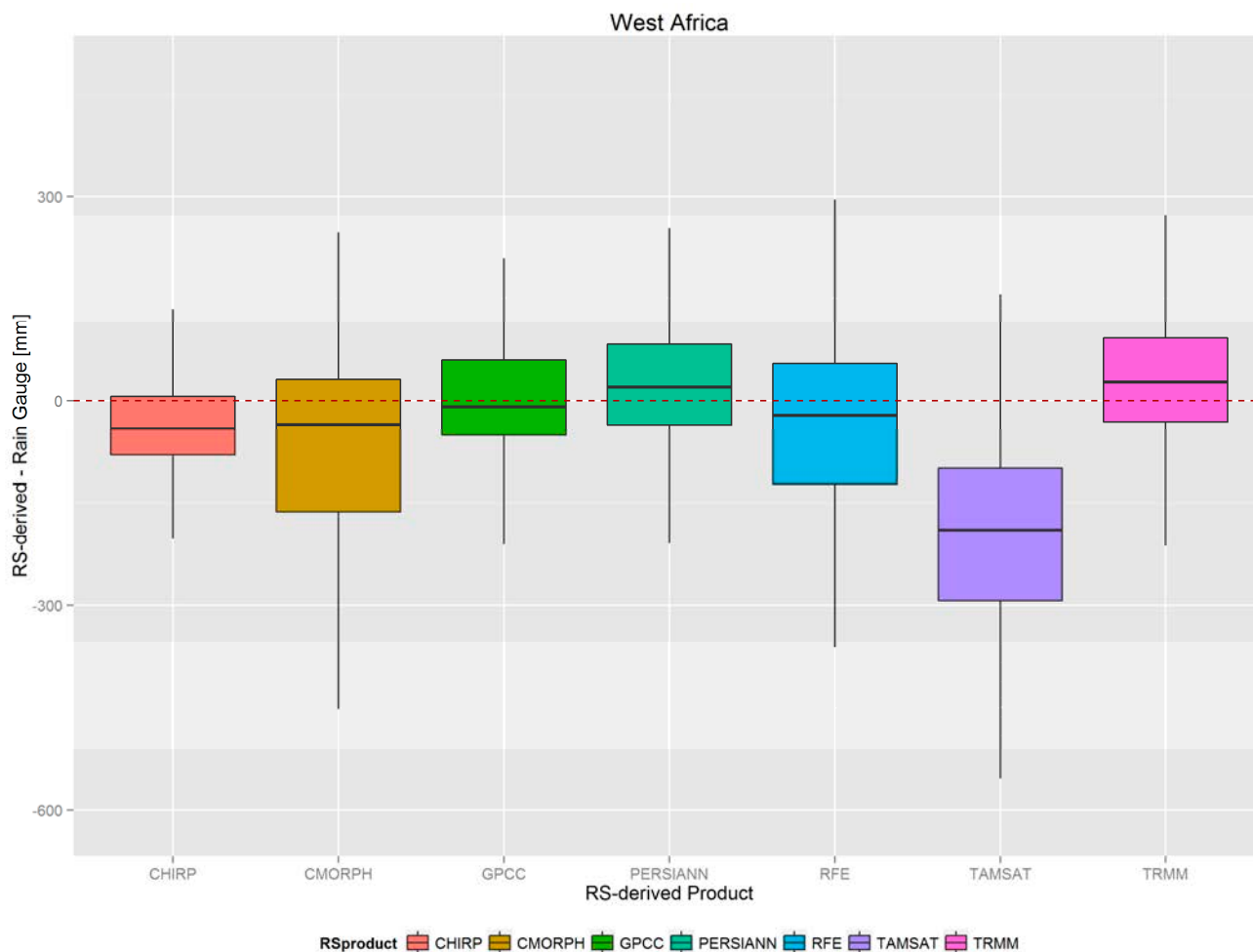




**Figure 14.** Scatterplot of Rain gauges *versus* Satellite-derived precipitation for West Africa.

Figure 14 shows that the best performance statistics refer to *CHIRP*, whereas *GPCC*, as for South America, presents comparable performance statistics. *PERSIANN* goes from the poor performance across South America to good performance across West Africa. *TRMM* gets reasonably good results; conversely *RFE*, *TAMSAT* and *CMORPH* have among the highest errors. *RMSE* values range from 135 mm to 386 mm, whereas absolute error ranges from 85 mm to 238 mm for *GPCC* and *CMORPH*, respectively. It is interesting that density plots exhibit a trace of bimodal distribution, noticeable by the reddish and greenish clusters around 500 and 1200 mm of annual precipitation.

Just as for South America, Box-and-Whisker plot of Figure 15 summarizes the difference between downscaled and observed precipitation across West Africa for different RS products.

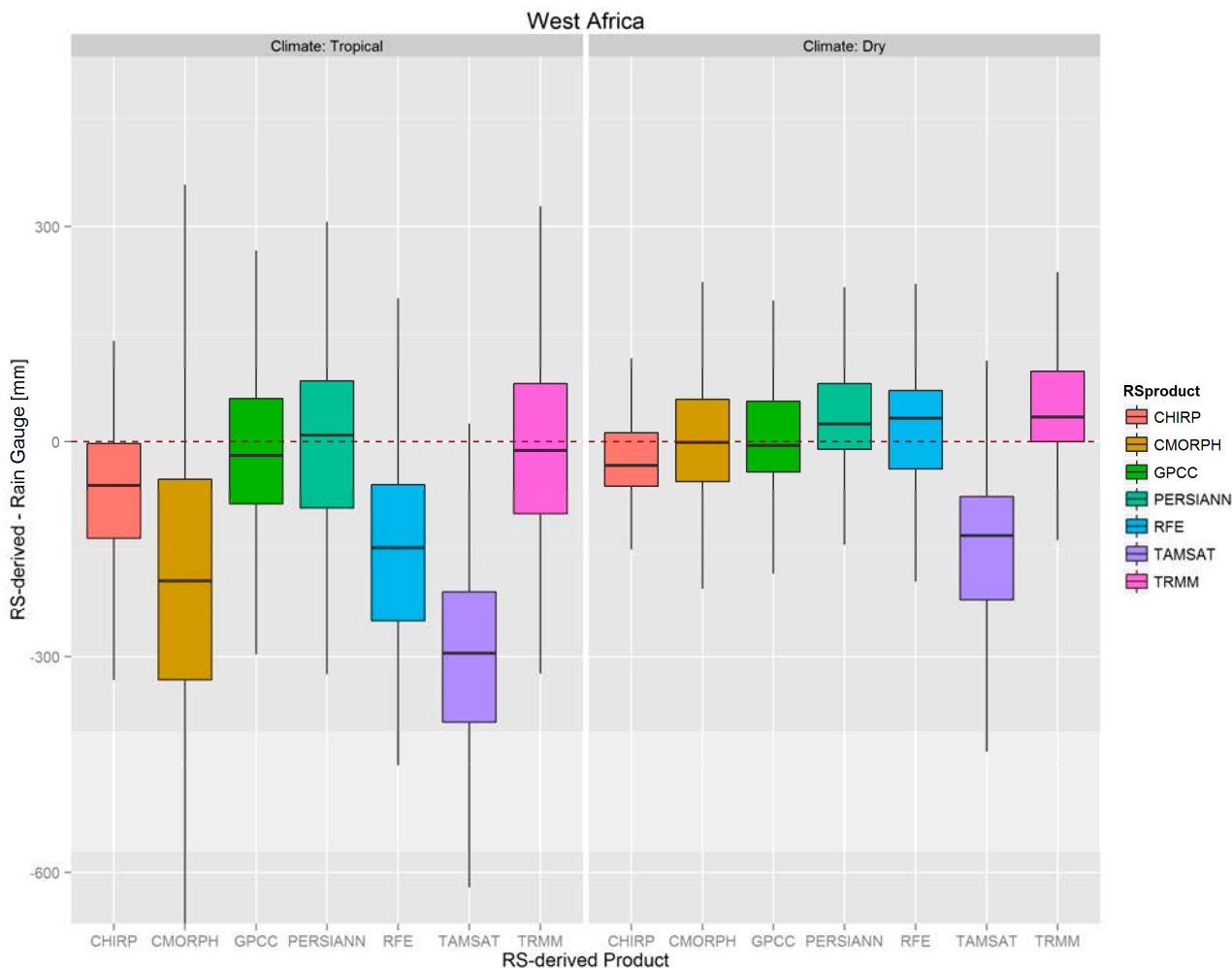


**Figure 15.** Box-and-Whiskerplot of the difference between downscaled and observed precipitation across West Africa for different RS products.

Results support our previous findings. *GPCC*'s median is closest to zero, however, *CHIRP* has an interquartile and an extreme range narrower than those pertaining to *GPCC*. *PERSIANN* and *TRMM* show comparable performances. *RFE* and *CMORPH* present notably poor results, whereas *TAMSAT* displays the highest disagreement. Also, note that the interquartile range of the latter does not contain the zero.

The elevation-dependent analysis has not been carried out because of: (1) the reduced size of the validation dataset, and (2) the distribution of the elevation. In fact, around 60% of the rain gauges across West Africa are located below 100 m asl.

Figure 16 provides a summary of the difference between downscaled and observed precipitation with respect to the two principal climate classes of the rain gauge network available across West Africa, namely Tropical and Dry climates.



**Figure 16.** Box-and-Whiskerplot of the difference between downscaled and observed precipitation across West Africa for different climate classes.

Box-and-Whiskerplot analysis shows that precipitation across Dry zones is reasonably well reproduced for all the satellite products but *TAMSAT* and *TRMM*, whereas across Tropical zones is well reproduced only for *GPCC*, *PERSIANN* and *TRMM*. *GPCC* presents the best ensemble statistics, *RFE* and *CMORPH* shows good results for Dry climate and *TAMSAT* shows notably bad statistics. Again, for both Tropical and Dry climates, *GPCC*'s median is closest to zero, even if *CHIRP* exhibits an interquartile and an extreme range narrower than those pertaining to *GPCC*.

Just as for South America, Table 3 summarizes the main statistical indicators used for the evaluation of the rainfall products.

For Senegal, that includes around 50% of the rain gauge stations of West Africa *CHIRP* is the best dataset, followed by *PERSIANN* and *GPCC*, while *TAMSAT* and *CMORPH* present the highest discrepancies. Also, *CMORPH* exhibits contrasting results in *RMSE*, *Abs. Err.* and *MPE*: the percentage error is fairly low, whereas both *RMSE* and *Abs. Err* are not the lowest ones.

Generally, results are consistent with previous findings of Figure 14 but for Mauritania where *PERSIANN* outperforms both *CHIRP* and *GPCC*. Also, Table 3 suggests that *CMORPH* presents a dual behavior: on one hand it gets reasonably good results for Burkina Faso and Niger, on the other hand it gets notably poor results for the other countries.

**Table 3.** Statistical indicators of Satellite-derived dataset *versus* rain gauge stations across nine validation countries. Numbers in parentheses in the first row refer to the number of rain gauge station employed for each country.

	<i>Burkina Faso (15)</i>			<i>Mali (65)</i>			<i>Mauritania (12)</i>		
	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>
TRMM	37.8	34.5	−1.2	79.2	63.0	2.9	43.3	32.2	2.4
GPCC	51.5	40.2	−5.4	83.9	58.7	−3.4	30.8	23.9	−4.9
PERSIANN	42.6	35.3	−1.7	82.1	63.1	1.9	27.4	21.6	2.6
CMORPH	60.7	47.7	−3.4	197.0	127.9	−10.8	91.9	63.8	20.3
CHIRP	69.6	68.1	−8.8	69.0	54.3	−4.7	45.8	42.0	−24.7
RFE	63.2	53.9	4.5	135.7	104.0	−1.8	60.9	49.0	29.2
TAMSAT	224.8	211.5	−26.4	302.5	278.8	−31.8	110.2	97.6	−43.1
	<i>Niger (44)</i>			<i>Senegal (180)</i>					
	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>	<i>RMSE</i>	<i>Abs. Err.</i>	<i>MPE</i>			
TRMM	42.1	33.0	11.2	165.2	118.7	20.6			
GPCC	43.1	34.3	−4.0	158.6	113.4	12.6			
PERSIANN	42.5	31.4	6.0	165.7	116.8	17.7			
CMORPH	63.3	47.6	3.1	200.5	139.3	3.3			
CHIRP	51.4	43.4	−10.1	137.2	92.7	5.6			
RFE	68.0	59.4	18.2	166.9	111.5	5.4			
TAMSAT	85.7	74.5	−16.9	243.8	198.4	−27.6			

#### 4. Discussion

Obviously the scheme presented here must be considered as a first-order attempt towards rainfall downscaling across regions with sparse gauge data, and consequently has a number of limitations that prevent it from reproducing precipitation patterns completely truthfully.

Firstly, as pointed out in the Materials section, the temporal consistency among satellite-derived rainfall, *EVI* and rain gauge time series is a critical aspect. We always consider the longest time series of precipitation from both satellite and rain gauge—even if the corresponding timespans are not consistent with each other—in order to have a more reliable estimate of the *MAP*. On one hand, this assumption might be an oversimplification of the real situation and generate errors. On the other hand, if we use a consistent time window (e.g., 2000–2012) between *EVI* and satellite-derived precipitation: (1) *GPCC* is excluded, since there is no temporal overlap with the *EVI* dataset, and (2) we lose the temporal consistency with the validation dataset. Note that the validation dataset is a climatology that mainly consists of 30-year precipitation dataset.

Secondly, the validation of the high resolution precipitation product with ground-based measurements—the core element of our analysis [55]—is fraught with difficulties. Even if ground observations are available, they are limited to few points in space and time, and they often need a thorough quality control. Thereby, neither high-resolution satellite-derived nor rain gauge data can be considered as the “*truth*”, and caution must be paid when handling these datasets. This is particularly true in the case of West Africa, where *in situ* data have not been made extensively enough to validate properly the spatial patterns of *MAP*. In other words, the volume of validation data over West Africa,

especially in northern Mali, northern Niger, Liberia, and across the Gulf of Guinea is limited. Thereof, low values of *RMSE* could be misleading.

Thirdly, statistical indicators used in the validation phase such as *RMSE*, are scale-dependent and are sensitive to the magnitude of the rainfall regime. This implies that low *RMSE* across West Africa is also due to the fact that rainfall is generally lower than in South America. Conversely, across Colombia, the high values of *RMSE* are associated to the tropical climate and the displacement of the intertropical convergence zone with wet weather through the year (up to 5000 mm each year).

These limitations do not allow us to select directly the “most suitable” rainfall dataset, and require particular attentions and supplementary investigations.

That being said, our downscaling scheme has the capability to capture the complex and spatially varying relationship between precipitation, elevation and vegetation index, and as such can serve to quantify *MAP* across regions with sparse gauge network with an adequate spatial resolution for advanced hydrological analysis. Moreover, statistics presented in the Results section give an indication of the performance of the downscaled rainfall product. Country-level, elevation- and climate-class statistic complement the accuracy assessment of the downscaled data. These analyses, in spite of the scarce rain gauge network across South America and West Africa, provide a general indication to users on which dataset is better to use across South America or West Africa.

Also, note that the aim of this study is to provide high resolution precipitation maps, shown to be useful to increase the effectiveness of water resource management, rather than propose a new methodology. The originality of this study lies in the high resolution dataset at continental scale, not in the methodology. However, the proposed methodology offers some elements of novelty compared to the models that have inspired it (*i.e.*, Chen *et al.* [15]). Firstly, with the proposed methodology there is the possibility to select the best regression at the option to exclude *EVI* from the regression. Secondly, we have a “global point of view”—*i.e.*, we run our downscaling scheme over two continents—while Chen *et al.* [15] only addresses local scale needs. Thirdly, we use the *EVI* instead of the *NDVI*, and this accurate index is more advantageous to the regression. In fact, *EVI* is able to capture the relationship between precipitation and vegetation greenness being influenced to a lesser degree by soil or atmospheric effects.

Applications of these datasets are manifold, to name only one, the long-term average water balance. Maps of annual precipitation, along with evapotranspiration data, could be useful to quantify the Budyko curve for mean annual water balance [56] at continental level.

## 5. Conclusions

Although precise precipitation information can be derived from ground “truth”, that is, rain gauge measurements, few rain gauges are generally available in developing countries: spatial distribution of rainfall presents a considerable challenge for these regions which often have sparse gauge data and a highly variable climate. In principle, satellites are accurate and versatile instruments to evaluate the precipitation instead, but their coarse spatial resolution (*i.e.*, often in the region of 20 km) do not allow a direct use without undertaking appropriate downscaling.

Robust assessments of precipitation distribution at high spatial resolution for these regions could provide essential information for hydrological modeling, water resource planning and management, and ecosystem services, amongst others.

This research outlines a simple and portable approach to downscale at 1 km spatial resolution Mean Annual Precipitation (*MAP*) using different satellite-based products, using the scheme proposed by Chen *et al.* [15]. The results are specific to South America and West Africa, but the method is designed to operate worldwide.

The proposed methodology includes: (1) the Geographically Weighted Regression (*GWR*) among satellite-derived annual precipitation, elevation (*DEM*) and a Vegetation Index (*EVI*) at coarse spatial resolution, (2) the interpolation at 1 km resolution of the residual of the model—*i.e.*, the amount of satellite-derived precipitation that could not be explained by the model- and (3) the application of the *GWR* to the dataset of *DEM* and *EVI* at the resolution of 1 km, adding up the high-resolution residual. A pixel-wise flagging strategy documents each step of the procedure, providing us with the statistics of the regression and information about whether both *EVI* and *DEM* or only *DEM* are employed in the regression.

Validation indicates that high resolution *MAP* from *CHIRP* and *GPCC* present the best ensemble of performance statistics for South America. *CHIRP* exhibits the best correlation for the elevation class above 1000 m. Also, *CHIRP* achieves better results than *GPCC* across Temperate, Tropical, Polar/Alpine climates, the other way around across Dry climate. *GPCC* and *CHIRP* outperform the other products also across West Africa where *CHIRP* achieves better results than *GPCC* across Dry climate zones, the other way around across Tropical climate.

The high resolution annual precipitation across South America using *CHIRP* dataset presents a *RMSE* of 388 mm and an absolute error of 211 mm if compared to ground stations. For West Africa, the *RMSE* between high resolution annual precipitation from *CHIRP* and ground measurements is equal to 126 mm, whereas the absolute error is equal to 85 mm.

High resolution datasets of precipitation across South America and West Africa are crucial. Applications of this analysis are manifold: to name only few, *MAP* can be used to: (1) develop scenarios for land and river management, (2) manage the “water-energy-food” nexus [57], and (3) develop frequency analysis of rainfall extremes using jointly *MAP* values and *L-moments* [58].

The major advantage of this approach presented lies in its numerical simplicity. However, it should be noted that numerous assumptions lie beneath this methodology, and caution must be paid when handling the validation results because of the limited volume of ground measurements.

High resolution datasets are openly available and are provided as “Supplementary Files and Materials” alongside this article. High resolution datasets are also available in AquaKnow [59], a platform managed by the Joint Research Centre of the European Commission, dedicated to technical and scientific knowledge for the sustainable development of the water sector.

The next steps of this work include: (1) the implementation of a downscaling scheme at monthly time step, exploiting better the vegetation response to precipitation; and (2) the computation of the Budyko curve for mean annual and monthly water balance.

## Acknowledgments

The authors gratefully acknowledge support from JRC. This work was supported by EUROCLIMA and RALCEA projects, funded by European Commission EuropeAid Co-operation Office. The authors thank the providers of the dataset needed to perform this research. Rain gauge data for South America have been provided by: Jorge Núñez Cobo, Centro del Agua para Zonas Áridas y Semiáridas de América Latina y el Caribe (CAZALAC), Chile; Claudia Patricia Romero Hernández, Universidad Nacional, and Claudia Yaneth Contreras Trujillo, Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), Colombia; Arthur Lucas Bernardo, Laboratorio de Análise e Processamento de Imagens de Satélite (LAPIS), Brasil; Roberto Eduardo Zanvettor, Universidad Nacional de Córdoba (CREAN), Argentina; Franklin Javier Paredes Trejo and Freddy José La Cruz Alvaro, Centro de Investigaciones Hidrológicas y Ambientales (CIHAM-UC), Venezuela; Julia Acuña Azata, Servicio Nacional de Meteorología e Hidrología (SENAMHI), Perú. Rain gauge data for West Africa have been provided by: Sampson Oduro-Kwarteng, Kwame Nkrumah, University of Science and Technology, Ghana; Jacob Ehiorobo, University of Benin, Benin; Dogara Bashir and Martin Obada Eduvie, National Water Resources Institute, Nigeria; Alioune Kane, Awa Niang Fall, Université Cheikh Anta Diop de Dakar; Senegal. EVI 1 km resolution annual composites have been created using the method proposed by Tomislav Hengl at ISRIC (*i.e.*, the World Soil Information Foundation). The rainfall datasets of South America and West Africa at 1 km spatial resolution are available as “Supplementary Files” or by contacting the first author at [guido.ceccherini@ext.jrc.ec.europa.eu](mailto:guido.ceccherini@ext.jrc.ec.europa.eu). The authors acknowledge Alberto Pistocchi for his help for editing the paper. The authors acknowledge Lorena Hojas Gascon and William Becker for their invaluable help for editing the paper and the three anonymous reviewers for their fruitful comments.

## Author Contributions

All authors contributed extensively to the work presented in this paper. Guido Ceccherini developed the concept, implemented the downscaling scheme and wrote the manuscript. The other authors provided in-situ observation data, technical support, and conceptual advice and edited the manuscript. All authors discussed the results and implications on the manuscript at all stages.

## Conflicts of Interest

The authors declare no conflict of interest.

## References

1. Hosking, J.R.M.; Wallis, J.R. *Regional Frequency Analysis: An Approach Based on L-Moments*; Cambridge University Press: Cambridge, UK, 2005.
2. Schaefer, M.G. Regional analyses of precipitation annual maxima in Washington State. *Water Resour. Res.* **1990**, *26*, 119–131.
3. Tapiador, F.J.; Turk, F.J.; Petersen, W.; Hou, A.Y.; García-Ortega, E.; Machado, L.A.T.; Angelis, C.F.; Salio, P.; Kidd, C.; Huffman, G.J.; de Castro, M. Global precipitation measurement: Methods, datasets and applications. *Atmos. Res.* **2012**, *104–105*, 70–97.

4. Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The Global Precipitation Measurement Mission. *Bull. Am. Meteorol. Soc.* **2013**, *95*, 701–722.
5. Guan, H.; Wilson, J.L.; Xie, H. A cluster-optimizing regression-based approach for precipitation spatial downscaling in mountainous terrain. *J. Hydrol.* **2009**, *375*, 578–588.
6. Rebora, N.; Ferraris, L. The structure of convective rain cells at mid-latitudes. *Adv. Geosci.* **2006**, *7*, 31–35.
7. Sorooshian, S.; AghaKouchak, A.; Arkin, P.; Eylander, J.; Foufoula-Georgiou, E.; Harmon, R.; Hendrickx, J. M.H.; Imam, B.; Kuligowski, R.; Skahill, B.; Skofronick-Jackson, G. Advancing the remote sensing of precipitation. *Bull. Am. Meteorol. Soc.* **2011**, *92*, 1271–1272.
8. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213.
9. Immerzeel, W.W.; Rutten, M.M.; Droogers, P. Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sens. Environ.* **2009**, *113*, 362–370.
10. Hunink, J.E.; Immerzeel, W.W.; Droogers, P.A. High-resolution Precipitation 2-step mapping Procedure (HiP2P): Development and application to a tropical mountainous area. *Remote Sens. Environ.* **2014**, *140*, 179–188.
11. Jia, S.; Zhu, W.; Lv, A.; Yan, T. A statistical spatial downscaling algorithm of TRMM precipitation based on NDVI and DEM in the Qaidam Basin of China. *Remote Sens. Environ.* **2011**, *115*, 3069–3079.
12. Park, N.-W. Spatial Downscaling of TRMM precipitation using geostatistics and fine scale environmental variables. *Adv. Meteorol.* **2013**, *2013*, doi:10.1155/2013/237126.
13. Fotheringham, A.S.; Brunson, C.; Charlton, M. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, 1st ed.; John Wiley & Sons Inc.: Chichester, England, UK and Hoboken, NJ, USA, 2002.
14. Foody, G.M. Geographical weighting as a further refinement to regression modelling: An example focused on the NDVI-rainfall relationship. *Remote Sens. Environ.* **2003**, *88*, 283–293.
15. Chen, F.; Liu, Y.; Liu, Q.; Li, X. Spatial downscaling of TRMM 3B43 precipitation considering spatial heterogeneity. *Int. J. Remote Sens.* **2014**, *35*, 3074–3093.
16. Brocca, L.; Ciabatta, L.; Massari, C.; Moramarco, T.; Hahn, S.; Hasenauer, S.; Kidd, R.; Dorigo, W.; Wagner, W.; Levizzani, V. Soil as a natural rain gauge: Estimating global rainfall from satellite soil moisture data. *J. Geophys. Res.: Atmos.* **2014**, *119*, doi:10.1002/2014JD021489.
17. Wood, A.W.; Leung, L.R.; Sridhar, V.; Lettenmaier, D.P. Hydrologic Implications of dynamical and statistical approaches to downscaling climate model outputs. *Clim. Chang.* **2004**, *62*, 189–216.
18. Becker, A.; Finger, P.; Meyer-Christoffer, A.; Rudolf, B.; Schamm, K.; Schneider, U.; Ziese, M. A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present. *Earth Syst. Sci. Data* **2013**, *5*, 71–99.
19. GPCC. Available online: [ftp://ftp.dwd.de/pub/data/gpcc/html/gpcc\\_normals\\_v2011\\_doi\\_download.html](ftp://ftp.dwd.de/pub/data/gpcc/html/gpcc_normals_v2011_doi_download.html) (accessed on 20 August 2014).



20. 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. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **2007**, *8*, 38–55.
21. TRMM 3B43. Available online: <http://trmm.gsfc.nasa.gov/3b43.html> (accessed 20 August 2014).
22. Ashouri, H.; Hsu, K.-L.; Sorooshian, S.; Braithwaite, D.K.; Knapp, K.R.; Cecil, L.D.; Nelson, B.R.; Prat, O.P. PERSIANN-CDR: Daily precipitation climate data record from multi-satellite observations for hydrological and climate studies. *Bull. Am. Meteorol. Soc.* **2014**, *96*, doi:10.1175/BAMS-D-13-00068.1.
23. PERSIANN CDR. Available online: <http://www.ncdc.noaa.gov/cdr/operationalcdrs.html> (accessed on 20 August 2014).
24. CMORPH. Available online: [ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH\\_V1.0/](ftp://ftp.cpc.ncep.noaa.gov/precip/CMORPH_V1.0/) (accessed on 20 August 2014).
25. Joyce, R.J.; Janowiak, J.E.; Arkin, P.A.; Xie, P. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* **2004**, *5*, 487–503.
26. CHIRP. Available online: <http://chg.geog.ucsb.edu/data/chirps/> (accessed on 20 August 2014).
27. Funk, C.C.; Peterson, P.J.; Landsfeld, M.F.; Pedreros, D.H.; Verdin, J.P.; Rowland, J.D.; Romero, B.E.; Husak, G.J.; Michaelsen, J.C.; Verdin, A.P. *A Quasi-Global Precipitation Time Series for Drought Monitoring*; U.S. Geological Survey Data Series 832; USGS: Sioux Falls, SD, USA, 2014; p. 4.
28. Herman, A.; Kumar, V.B.; Arkin, P.A.; Kousky, J.V. Objectively determined 10-day African rainfall estimates created for famine early warning systems. *Int. J. Remote Sens.* **1997**, *18*, 2147–2159.
29. Xie, P.; Xiong, A.-Y. A conceptual model for constructing high-resolution gauge-satellite merged precipitation analyses. *J. Geophys. Res.: Atmos.* **2011**, *116*, doi:10.1029/2011JD016118.
30. RFE. Available online: <http://www.cpc.ncep.noaa.gov/products/fews/rfe.shtml/> (accessed on 20 August 2014).
31. Maidment, R.I.; Grimes, D.; Allan, R.P.; Tarnavsky, E.; Stringer, M.; Hewison, T.; Roebeling, R.; Black, E. The 30 year TAMSAT African Rainfall Climatology And Time series (TARCAT) data set. *J. Geophys. Res.: Atmos.* **2014**, *119*, doi:10.1002/2014JD021927.
32. TAMSAT. Available online: <http://www.met.reading.ac.uk/~tamsat/data/> (accessed on 20 August 2014).
33. Thompson, S.E.; Harman, C.J.; Troch, P.A.; Brooks, P.D.; Sivapalan, M. Spatial scale dependence of ecohydrologically mediated water balance partitioning: A synthesis framework for catchment ecohydrology. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2010WR009998.
34. Pinty, B.; Verstraete, M.M. GEMI: A non-linear index to monitor global vegetation from satellites. *Vegetatio* **1992**, *101*, 15–20.
35. Bradley, A.V.; Gerard, F.F.; Barbier, N.; Weedon, G.P.; Anderson, L.O.; Huntingford, C.; Aragão, L.E.O.C.; Zelazowski, P.; Arai, E. Relationships between phenology, radiation and precipitation in the Amazon region. *Glob. Change Biol.* **2011**, *17*, 2245–2260.

36. Sjöström, M.; Ardö, J.; Arneth, A.; Boulain, N.; Cappelaere, B.; Eklundh, L.; de Grandcourt, A.; Kutsch, W.L.; Merbold, L.; Nouvellon, Y.; *et al.* Exploring the potential of MODIS EVI for modeling gross primary production across African ecosystems. *Remote Sens. Environ.* **2011**, *115*, 1081–1089.
37. EVI. Available online: [https://lpdaac.usgs.gov/products/modis\\_products\\_table/mod13q1](https://lpdaac.usgs.gov/products/modis_products_table/mod13q1) (accessed on 20 August 2014).
38. Zhu, Z.; Bi, J.; Pan, Y.; Ganguly, S.; Anav, A.; Xu, L.; Samanta, A.; Piao, S.; Nemani, R.R.; Myneni, R.B. Global data sets of vegetation Leaf Area Index (LAI)3g and Fraction of Photosynthetically Active Radiation (FPAR)3g derived from Global Inventory Modeling and Mapping Studies (GIMMS) Normalized Difference Vegetation Index (NDVI3g) for the period 1981 to 2011. *Remote Sens.* **2013**, *5*, 927–948.
39. Pekel, J.-F.; Vancutsem, C.; Bastin, L.; Clerici, M.; Vanbogaert, E.; Bartholomé, E.; Defourny, P. A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. *Remote Sens. Environ.* **2014**, *140*, 704–716.
40. SWBD. Available online: [http://dds.cr.usgs.gov/srtm/version2\\_1/SWBD/](http://dds.cr.usgs.gov/srtm/version2_1/SWBD/) (accessed on 20 August 2014).
41. HydroSHEDS. Available online: <http://hydrosheds.cr.usgs.gov/index.php> (accessed on 20 August 2014).
42. Farr, T.G.; Rosen, P.A.; Caro, E.; Crippen, R.; Duren, R.; Hensley, S.; Kobrick, M.; Paller, M.; Rodriguez, E.; Roth, L.; *et al.* The shuttle radar topography mission. *Rev. Geophys.* **2007**, *45*, doi:10.1029/2005RG000183.
43. Lehner, B.; Verdin, K.; Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos. Trans. Am. Geophys. Union* **2008**, *89*, 93–94.
44. EUROCLIMA Project. Available online: <http://www.euroclima.org/en/euroclima> (accessed on 12 December 2014).
45. Gollini, I.; Lu, B.; Charlton, M.; Brunsdon, C.; Harris, P. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. Available online: <http://arxiv.org/pdf/1306.0413.pdf> (accessed on 17 March 2014).
46. Guo, L.; Ma, Z.; Zhang, L. Comparison of bandwidth selection in application of geographically weighted regression: A case study. *Can. J. For. Res.* **2008**, *38*, 2526–2534.
47. Bowman, A.W. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **1984**, *71*, 353–360.
48. Hurvich, C.M.; Simonoff, J.S.; Tsai, C.-L. Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **1998**, *60*, 271–293.
49. Hengl, T.; Heuvelink, G.B.M.; Stein, A. A generic framework for spatial prediction of soil variables based on regression-kriging. *Geoderma* **2004**, *120*, 75–93.
50. Hengl, T.; Heuvelink, G.B.M.; Rossiter, D.G. About regression-kriging: From equations to case studies. *Comput. Geosci.* **2007**, *33*, 1301–1315.
51. Teng, H.; Shi, Z.; Ma, Z.; Li, Y. Estimating spatially downscaled rainfall by regression kriging using TRMM precipitation and elevation in Zhejiang Province, southeast China. *Int. J. Remote Sens.* **2014**, *35*, 7775–7794.

52. Meng, Q. Regression kriging *versus* geographically weighted regression for spatial interpolation. *Int. J. Adv. Remote Sens. GIS* **2014**, *3*, 606–615.
53. Kottek, M.; Grieser, J.; Beck, C.; Rudolf, B.; Rubel, F. World map of the Köppen-Geiger climate classification updated. *Meteorol. Z.* **2006**, *15*, 259–263.
54. Overeem, A.; Leijnse, H.; Uijlenhoet, R. Country-wide rainfall maps from cellular communication networks. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 2741–2745.
55. Qin, Y.; Chen, Z.; Shen, Y.; Zhang, S.; Shi, R. Evaluation of satellite rainfall estimates over the Chinese Mainland. *Remote Sens.* **2014**, *6*, 11649–11672.
56. Zhang, L.; Potter, N.; Hickel, K.; Zhang, Y.; Shao, Q. Water balance modeling over variable time scales based on the Budyko framework—Model development and testing. *J. Hydrol.* **2008**, *360*, 117–131.
57. Finley, J.W.; Seiber, J.N. The nexus of food, energy, and water. *J. Agric. Food Chem.* **2014**, *62*, 6255–6262.
58. Di Baldassarre, G.; Castellarin, A.; Brath, A. Relationships between statistics of rainfall extremes and mean annual precipitation: An application for design-storm estimation in northern central Italy. *Hydrol. Earth Syst. Sci.* **2006**, *10*, 589–601.
59. AQUAKNOW.NET. Available online: <http://www.aquaknow.net/en/home> (accessed on 20 August 2014).

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).