



**HAL**  
open science

# A Simple Statistical Model of the Uncertainty Distribution for Daily Gridded Precipitation Multi-Platform Satellite Products

Rômulo a J Oliveira, Rémy Roca

► **To cite this version:**

Rômulo a J Oliveira, Rémy Roca. A Simple Statistical Model of the Uncertainty Distribution for Daily Gridded Precipitation Multi-Platform Satellite Products. *Remote Sensing*, 2022, 14, 10.3390/rs14153726 . hal-04235716

**HAL Id: hal-04235716**

**<https://cnrs.hal.science/hal-04235716>**

Submitted on 10 Oct 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Article

# A Simple Statistical Model of the Uncertainty Distribution for Daily Gridded Precipitation Multi-Platform Satellite Products

Rômulo A. J. Oliveira <sup>1,2,\*</sup> and Rémy Roca <sup>1</sup>

<sup>1</sup> Laboratoire d'Etudes en Géophysique et Océanographie Spatiales, Université de Toulouse III, CNRS, CNES, IRD, 31062 Toulouse, France; remy.roca@legos.obs-mip.fr

<sup>2</sup> Géosciences Environnement Toulouse, Université de Toulouse III, CNRS, CNES, IRD, 31062 Toulouse, France

\* Correspondence: juca.oliveira@legos.obs-mip.fr

**Abstract:** Multi-platform satellite-based precipitation gridded estimates are becoming widely available in support of climate monitoring and climate science. The characterization of the performances of these emerging Level-4 products is an active field of research. This study introduced a simple Gaussian mixture model (GMM) to characterize the distribution of uncertainty in these satellite products. The following three types of uncertainty were analyzed: constellation changes-induced uncertainties, sampling uncertainties and comparison with rain-gauges. The GMM was systematically compared with a single Gaussian approach and shown to perform well for the variety of uncertainties under consideration regardless of the precipitation levels. Additionally, GMM has also been demonstrated to be effective in evaluating the impact of Level-2 PMW rain estimates' detection threshold definition on the constellation changes-induced uncertainty characteristics at Level-4. This simple additive perspective opens future avenues for better understanding error propagation from Level-2 to Level-4.

**Keywords:** precipitation; satellite; uncertainty analysis; error model; Gaussian mixtures



**Citation:** Oliveira, R.A.J.; Roca, R. A Simple Statistical Model of the Uncertainty Distribution for Daily Gridded Precipitation Multi-Platform Satellite Products. *Remote Sens.* **2022**, *14*, 3726. <https://doi.org/10.3390/rs14153726>

Academic Editor: Silas Michaelides

Received: 20 June 2022

Accepted: 26 July 2022

Published: 3 August 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 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 (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Precipitation is at the heart of the global water and energy cycle that is strongly impacted by global warming [1]. This has prompted the scientific and operational community to develop new products to support climate monitoring and a better understanding of processes [2]. At the one degree one day scale, the emergence of a suite of new and renewed precipitation estimates from in situ, reanalysis and satellite [3] has been shown to provide observational constraints for detection and attribution analysis [4], climate model validation [5,6] and process analysis [7]. In particular, with the advent of the long time series of satellite precipitation-relevant measurements, multi-platforms satellite precipitation estimates have been shown to reveal overall good performances [8,9].

While at a high resolution, multi-platform precipitation products' uncertainty has actively been explored [10–14] at the daily accumulated scale of  $1^\circ \times 1^\circ$ , less is known in terms of structural errors of the products. Numerous atmospheric processes contribute for modulating the precipitation (detection and intensity) distribution in space and time. Assessing this using satellite, on an  $1^\circ$  /daily scale, would support research on global precipitation (e.g., on the investigation of global extremes) and facilitate better intercomparison and validation exercises with a common gridded database. A few validation studies have nevertheless been conducted in the tropics revealing some regionally varying uncertainties that are difficult to interpret in terms of algorithms' limitations [15]. Chambon et al. [16] conducted a thorough error propagation analysis of the TAPEER product at a  $1^\circ \times 1^\circ$  daily scale. Their study revealed the importance of the sampling uncertainty in the error budget, compared to the Level-2 (L2) uncertainty and satellite calibration sources. It also showed the non-linearity of the L2 to Level-4 (L4) bias propagation, emphasizing the specific nature of the  $1^\circ \times 1^\circ$  accumulated daily precipitation uncertainty. Uncertainty depends on the

assumed error model [17] and also relates to the sources of intrinsic errors that create it in the first place. In the case of multi-platforms satellite products, for instance, the changes in the microwave constellation configuration can impact performances and the structural errors of the final product [18,19]. For daily gridded products, sampling errors due to a finite number of observations can significantly alter the comparison to ground-based networks [20].

As an attempt to improve the existing framework for better understanding the uncertainty of the daily  $1^\circ \times 1^\circ$  scale, the ability of a simple decomposition of the uncertainty, under an additive error model assumption and as a sum of two Gaussian distributions was assessed. The following three types of uncertainties were considered here: (i) constellation changes-induced uncertainties, (ii) sampling uncertainties and the (iii) uncertainties related to comparisons with rain-gauge network data. Each of the uncertainties are illustrated with a multi-platform satellite precipitation product. Section 2 presents the data and details the algorithm used for the Gaussian mixture calculations. Section 3 shows the results of the fitting which are systematically contrasted with those of a single Gaussian approach. Finally, our findings are summarized and discussed in Section 4.

## 2. Data and Method

### 2.1. Uncertainty Distribution as a Function of the Precipitation Accumulation

In the study, we adopted an additive error model by which we assumed the uncertainty to be obtained as a difference between the estimated precipitation accumulation and a reference. An additive error model was preferred to a multiplicative one since, at the  $1^\circ \times 1^\circ$  one day scale of interest here, this approach was shown to be well suited to characterize various sources of errors [15,20,21]. The distribution of uncertainty as a function of the precipitation accumulation  $P$  was hence computed as the Relative Error ( $RE$ ) per precipitation accumulation bin:

$$RE(P_{SRC}) = \left( \frac{P_{SRC} - P_{REF}}{P_{REF}} \right) \times 100, \quad (1)$$

where  $SRC$  is one of the sources of uncertainty under consideration and  $REF$  is a reference. Since the precipitation distribution usually tends to emphasize lower precipitation accumulations due to their higher frequency of occurrence, the precipitation values were log-scale binned-distributed, in order to provide more weight to the moderate and large precipitation accumulation categories by increasing their samples, which would benefit the model fitting.

Generally, the relative error distributions are considerably high and largely spread at low precipitation accumulations, with a high occurrence of non-representative extreme values (outliers). As the accumulated precipitation increases, the RE decrease strongly (distribution tends to narrower) and the diminishing of outliers is also observed. Given that, we utilized the interquartile IQR range rule ( $Q_1/Q_3 \pm IQR \times 1.5$ ) of the RE distributions at each precipitation accumulation bin (outliers neglected). Subsequently, the precipitation binned RE distributions were normalized to the corresponding maximum RE value. Indeed, the RE distributions can vary according to the uncertainty source, as discussed in detail in the following.

### 2.2. Sources of Uncertainty and Datasets

#### 2.2.1. Constellation Changes-Induced Uncertainties

The time evolution of the microwave constellation induces inhomogeneity in the time series of constellation-based precipitation estimates, as in [18]. This can be thought of as uncertainty arising from the changes in the platform number and configuration that forms the constellation. Such uncertainty can readily be estimated using data-denial experiments whereby the change of the configuration can be easily simulated with the Global Interpolated RAINfall Estimation (GIRAFE, [22]) framework. GIRAFE is an algorithm that provides  $1^\circ$ /daily rain accumulation estimates based on the combination of multi-source satellite

observations/estimates (i.e., from geostationary and low earth orbit satellites), globally and has been available since 2000 for up to 55°N/S. GIRAFE evolved from the Tropical Amount of Precipitation with an Estimate of ERrors (TAPEER) approach [19,23], which is based on the Universally Adjusted GOES Precipitation Index (UAGPI) technique [24,25]. In the present GIRAFE version, the instantaneous rain rate retrievals (i.e., the L2 input database) were extracted from the Goddard Profiling Algorithm (GPROF2017, [26]) and the Precipitation Retrieval and Profiling Scheme (PRPS2019, [27]) PMW algorithms were obtained from the GPM V05 and V06 database, respectively. Oliveira et al. [28] provided a systematic analysis of this source of uncertainty over the last decades and here we illustrated the resulting uncertainty distributions by exploring only 2 configurations that are representative of the actual changes in the constellation setup. The GPM “golden era” constellation period was selected as a reference for comparison with the other two different configurations, namely (i) an idealized case for a single SSMI/S and (ii) a configuration used to exploit the impact in the absence of MT1.SAPHIR and GPM.GMI. Table 1 shows the PMW radiometers considered for each experimental configuration. Data-denial experiments were performed for June-July-August of 2014 (JJA2014) and December-January-February of 2014/5 (DJF2014/5). The study domain corresponds to a quasi-global zone between 55°N/S. The GIRAFE framework training volume of  $5^\circ \times 5^\circ \times 5$  days was adopted for both the intensity and detection parameters, which were assumed as 1.5 and 0.5 mm h<sup>-1</sup>, respectively.

**Table 1.** The adopted data denial experiments and their respective constellation configurations, based on the number of platforms available for each corresponding period.

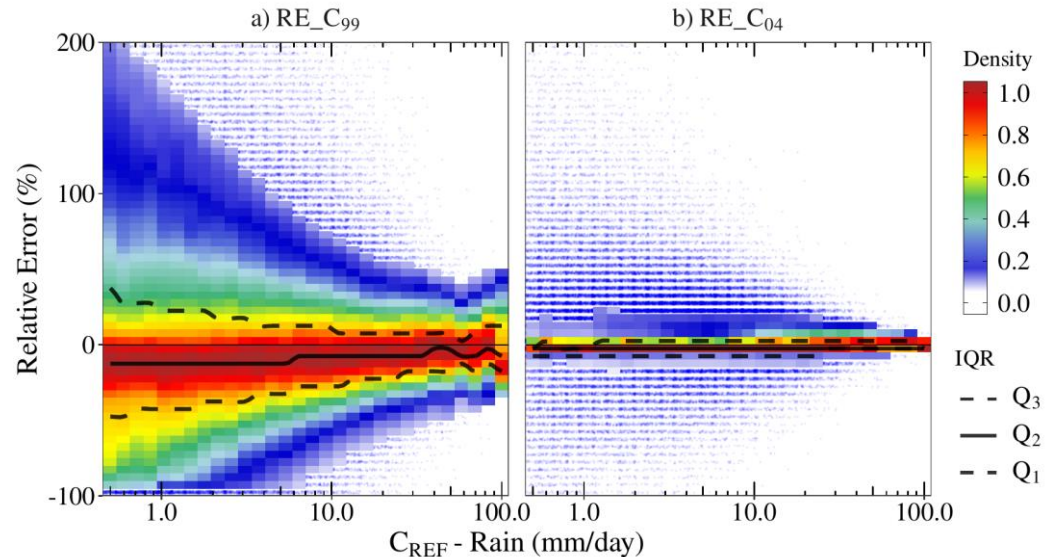
Experiment	Available Period	Plat. N°	Platforms
$C_{REF}$	4 March 2014–8 April 2015	12	SSMI/S(3), GCOMW1, TMI, GMI, SAPHIR *, MHS(4), ATMS
$C_{99}$	1 January 1988–31 December 1990	1	SSMI/S(1)
$C_{04}$	1 December 2006–29 February 2008	10	SSMI/S(3), GCOMW1, TMI, MHS(4), ATMS

C = Constellation configuration experiments. REF = Reference. \* Through PRPS2019 algorithm.

It is worth mentioning that the rain estimation framework used in this work (i.e., GIRAFE), for describing the uncertainty according to the constellation changes, was re-designed to accordingly geo-collocate the distinct platforms (imagers and sounders) and their different footprint sizes and shapes (retrieval resolution) that fall into an  $1^\circ \times 1^\circ$  grid. However, the uncertainty due to platform characteristics’ effect is still expected and remains due to the physical approaches at the sensor-level rain retrieval scheme. For instance, in the GPROF retrieval algorithm, which is commonly used as an input data source for L3 precipitation estimates (e.g., the IMERG V06 product), despite being able to consider multiple channels for providing the surface rain rate, its retrieval resolution is based on the ~19 GHz channel resolutions for most of the radiometers considered (i.e., imagers). Further, sounders (e.g., MHS, ATMS and SAPHIR), as cross-track sensors, have varying footprint sizes with the scan position and the adopted retrieval resolutions are taken at nadir by GPROF algorithm. Therefore, both the assumptions can lead to uncertainties at L2 rain retrievals, which can also be propagated to the L3 precipitation estimates. In this study, the footprint size artifact leading to uncertainty can be explained as part of the constellation configuration differences, however, its quantification was not explicitly explored. Nevertheless, further studies exploring the sensor footprint features as a component of uncertainty in daily precipitation estimates are recommended.

The relative error distribution shows strong variability depending firstly on the constellation configuration (Figure 1). As anticipated, the experiment closest to the reference configuration exhibited few relative differences and the distribution was narrow across the considered range of daily accumulation. For  $C_{04}$ , as an intermediate case, the distribution reached up to 50% uncertainty at maximum (i.e., IQR range rule). On the other hand, for the idealized one single SSMI/S configuration case ( $C_{99}$ ), the distribution was characterized by large errors (up to 200%) and a large spread, even at intense daily precip-

itation accumulations. The secondary cause of variability was the surface type. Indeed, the spread over the ocean region was much larger than over land for each configuration (not showed). Finally, for all set-ups, the relative errors' range decreased as the daily precipitation accumulation increased.



**Figure 1.** The uncertainty distribution due to constellation changes ( $y$ -axis, in %) as a function of the daily precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for two constellation configurations, during June–July–August of 2014, across the  $55^{\circ}\text{N/S}$  global zones and over-land surfaces. Panels show density scatter-plots (colored scale), superimposed by the IQR distribution (black lines), of the relative error (a,b) vs. the daily accumulated precipitation (log-scale distributed). Outliers are represented by blue dots.

The relative error distribution due to the constellation configuration was further illustrated by focusing on the RE\_C99 distributions that revealed the largest spread for small and a large daily accumulation cases (Figure 1a). The small accumulation distribution produced a truncated lower end and a large positive spread centered around a zero line. The IQR were negatively skewed and, despite the slightly positive tendency as precipitation increases, the higher densities were generally at about  $-15\% [\pm 5]$  of RE (systematic underestimation). The  $Q_1$ ,  $Q_2$  (median) and  $Q_3$  ranged from  $[-50\%, -20\%, +40\%]$  to  $[-30\%, -10\%, +20\%]$  for low-to-moderate precipitation accumulations ( $0.5$  to  $10 \text{ mm day}^{-1}$ ). In cases of larger precipitation amounts (e.g.,  $>10 \text{ mm day}^{-1}$ ), despite being negatively skewed, the distribution appeared more symmetrical and narrowed around the zeros. The overall RE distributions decreased, for instance, from approximately  $-80\%$  and  $+70\%$  for  $10 \text{ mm day}^{-1}$  to  $-45\%$  and  $+50\%$  at  $50 \text{ mm day}^{-1}$ . However, for precipitation accumulations greater than  $50 \text{ mm day}^{-1}$ , a slight increase towards positive RE (overestimation) was observed.  $Q_3$  increased by about  $15\%$ , from  $50$  to  $100 \text{ mm day}^{-1}$ . The uncertainty distribution at  $100 \text{ mm day}^{-1}$  expressed a slightly right-skewed distribution with median and density peaks at about  $-12\%$ . The closer the proximity in terms of constellation configuration (i.e., Figure 1b), the greater the chance that the uncertainty distribution acquires a unimodal and leptokurtic shape over the entire precipitation range.

In summary, these overall observed RE distribution characteristics, due to the constellation configuration, are closely linked to the nature of the PMW rain rates as input data that were also propagated up to the daily precipitation accumulations [18,20,29]. In fact, the overall impact that can strongly biasing L3 estimates can be explained through the better representation of precipitating yields spatially and temporally, being accordingly sampled by the PMW radiometers that contributed to the final daily amounts [30,31]. Therefore, assuming a multi-platform PMW rain retrieval database, variability in the overall rain distributions and its performances was expected which could also be attributed to the

retrieval algorithm [10,32–35]. For instance, according to Chambon et al. [21], the systematic uncertainties associated with PMW rain retrievals ranging between 2 and 10 mm h<sup>-1</sup> lead to an impact on the 1°/daily precipitation accumulations, resulting in non-Gaussian distributions. In this case, the constellation configuration-induced variability impacted the entire daily accumulated precipitation range. The impact is clearly noticeable through the modification of the RE distribution shape at each precipitation category, which can even be seen in the density spread or peak location.

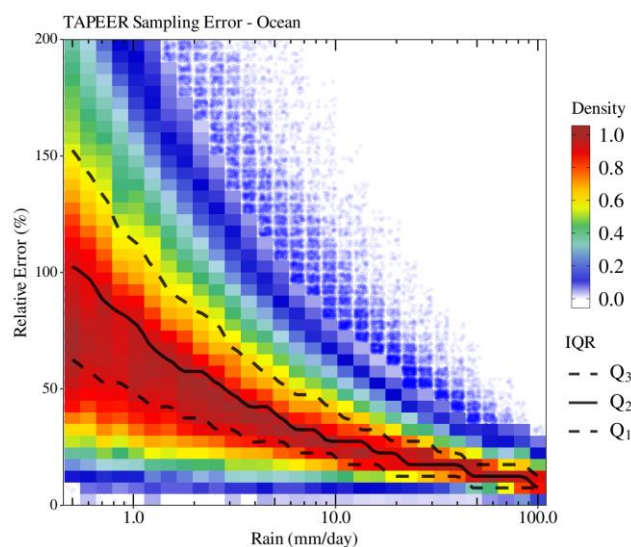
### 2.2.2. Sampling Uncertainties

The computation of daily 1° × 1° accumulation generates a simple uncertainty arising from the limited number of geostationary images used in the computation. In this context, geostationary images play a key role in the sampling uncertainty computation, for providing for instance the space and time rain/no-rain (detection) samples. The uncertainties due to sampling is akin to variance uncertainty of the mean and was developed under the Megha-Tropiques mission research efforts [36]. The associated error model  $S$ , is described at length in [20] and reads as follows:

$$S = \frac{\sigma}{\sqrt{\frac{A T}{d^2 \tau}}} \quad (2)$$

where  $\sigma$  is the space/time indicator variance of the instantaneous geostationary pixels in a degree (surface  $A$ ) during a day (duration  $T$ ), multiplied by a conditional precipitation rain rate (see [20] for details). The  $d$  and  $\tau$  scales are autocorrelation scales, in space and time, respectively, obtained from the calculations of space and time variograms. These are required to estimate the number of independent samples for the estimation of the variance uncertainty of the mean. The daily precipitation accumulations and the associated sampling uncertainties, from the TAPEER product [19,23], were considered. To assess this, we maintained a specific period of samples from the period June-July-August (JJA) of 2014 over the 30°S 30°N region, which is a product coverage zone (Tropics). Given that the uncertainty distributions did not differ considerably according to the surface type, as demonstrated by Roca et al. [22], we chose to focus on the ocean surface, which provides a larger sample size (larger number of grids).

Figure 2 shows the density scatterplot of the uncertainty distribution due to sampling, as a function of the daily precipitation intensity, for the JJA of 2014 over the tropics (30°N/S) over oceanic surfaces. In general, and as described by Chambon et al. [23], the relative sampling errors, which by definition are always positive, decrease as the rain accumulation intensity increases. In addition, the uncertainties have a large spread for lower precipitation amounts. The RE spread decreased markedly from 200% at 1 mm day<sup>-1</sup> to 80% at 10 mm day<sup>-1</sup>, reaching about 40% at precipitation amounts of greater than 50 mm day<sup>-1</sup>. Following that, the maximum RE density values were between 60% and 80% at lower precipitations (around mm day<sup>-1</sup>), decreasing to about 35% at 10 mm day<sup>-1</sup> and, subsequently, to 15% for large amounts (>80 mm day<sup>-1</sup>). Consequently, the same decreasing distribution, as a function of the precipitation, was also observed in the IQR, in which Q<sub>1</sub> (Q<sub>3</sub>) shifted downward from 65% (150%) to roughly 10% (20%) from low to large precipitation amounts, respectively. The IQR range diminished from 85% to approximately 10%. The median demonstrated a decreasing range of about 85% as precipitation intensity increased, varying from 100% (at 1 mm day<sup>-1</sup>) to roughly 15% (>80 mm day<sup>-1</sup>). Indeed, the overall error sampling distributions were preserved regardless of the surface type condition (land and ocean). The main surface type contrasts that can be accounted for are the larger RE spread at low-to-moderate (large) precipitation accumulations over ocean (land, not showed).



**Figure 2.** The uncertainty distribution due to sampling ( $y$ -axis, in %) as a function of the daily precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for June–July–August of 2014, between  $30^\circ\text{N/S}$  and over oceanic surfaces from TAPEER product. Panel shows the density scatterplot (colored scale) superimposed by the log-binned IQR distributions (in black) of the relative error vs. the daily accumulated precipitation (log-scale distributed). Outliers are represented by blue dots.

The misrepresentation of the sampling errors, as a component of the error budget, can propagate to the final daily precipitation distributions [20]. This would impact both the detection and the quantification of a certain daily precipitation intensity category [23]. According to Roca et al. [22], at intense precipitation accumulations the sampling relative uncertainty varies from 15% to 20%. Accordingly, the final daily precipitation amounts can be better represented, benefiting from the large range in the PMW observations/sampling, including the sounder platforms (e.g., SAPHIR), which lead to better spatial and temporal detection of those precipitating events that contribute to the daily precipitation accumulations, e.g., over the tropical zones [19]. Furthermore, the appropriated representation of such sampling error distributions throughout a statistical model is not a trivial task, since the RE distribution can assume different density shapes depending on the precipitation intensity.

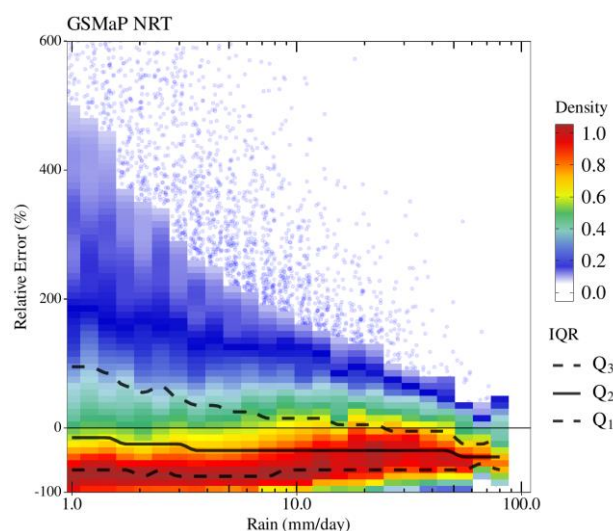
### 2.2.3. Uncertainties Obtained by Comparisons with Rain-Gauge Network

Beyond the intrinsic error sources identified above, the uncertainty of the accumulated daily integrated precipitation estimation is often assessed, in bulk, by conducting a comparison of rain-gauges networks. It is useful to estimate if the satellite product is characterized by a bias with respect to the in-situ measurements. To illustrate the simple modelling of this source of uncertainty, we used the JAXA's Global Satellite Mapping of Precipitation (GSMaP, [37]) near-real-time version 6 (NRT V6) product. The hourly  $0.1^\circ$  original data were averaged over  $1^\circ \times 1^\circ$  grid cells and accumulated over one day, from 12Z to 12Z to match the rain-gauge network reporting convention.

In this section, the analyses were carried out considering, as ground reference, the in-situ data from several daily accumulated precipitation observations (12Z–12Z) distributed over Brazil. After applying multiple and sequential quality control checks, also performed by the Centre for Weather Forecasting and Climatic Studies of Brazil (CPTEC/INPE) and the HYdro-geochemistry of the AMazonian Basin (HYBAM), daily accumulated gridded precipitation fields were created by computing the arithmetic average of all available stations within a pre-defined grid cell of  $1^\circ \times 1^\circ$  of spatial resolution. The approach considers the number of available gauges inside each grid cell (i.e., at least 5 rain gauges) as well as the standard error, in order to reduce the impact of the ground truth uncertainties [15]. Although these Brazilian gauge observations have long-term availability, the analyses here

focused on the period from December to March across 3 years (2014, 2015 and 2016)—which represents the rainy season in much of Brazil [38].

Figure 3 depicts the RE distribution of the GSMaP NRT satellite-based precipitation product as a function of the daily precipitation intensity. A large amount of low-to-moderate daily precipitation accumulations were observed, especially due to the substantial contribution of the RE overestimations that reached, for instance, up to +500% at 1 mm day<sup>-1</sup>. This dispersion tends to decrease considerably as the precipitation intensity increases. However, the RE density peaks were often located in the underestimation portion throughout the precipitation intensity range. From 1 to 10 mm day<sup>-1</sup>, the RE density peaks were mostly observed at about −90%, whereas for precipitation accumulations greater than 10 mm day<sup>-1</sup> the RE density peaks were located around −50%. This pattern was also followed by the negatively biased IQR distributions, in which Q<sub>3</sub> and Q<sub>2</sub> presented a constant decrease from +50% to −10% and from −10% to −55%, respectively. No significant changes were observed for the Q<sub>1</sub> component (values were constant at −75%, varying between ±5%). Yet, it is noteworthy that the RE density distribution changed from a non-normal distribution (highly positive skewed) at low-to-moderate precipitation intensities, to normal distributions at high precipitation intensities, both of which were negatively biased.



**Figure 3.** The relative difference distribution between satellite (GSMaP NRT) and ground-based data ( $y$ -axis, in %) as a function of the daily precipitation ( $x$ -axis, in mm day<sup>-1</sup>) over Brazil (land only) for December-January-February-March of three consecutive years (May 2014, June 2015 and July 2016). Panel shows the density scatterplot (colored scale) superimposed by the log-binned IQR distributions (in black) of the relative error vs. the daily accumulated precipitation (log-scale distributed). Outliers are represented by blue dots.

This overall uncertainty distribution scenario, for satellite-based daily accumulated precipitation products, can be explained by the impact and influence of different uncertainty sources [39]. According to Roca et al. [20], assuming the various uncertainty sources to be independent, the error budget can be defined according to sampling, algorithmic and calibration error terms. In addition, as described by Maggioni et al. [40], L3 satellite-based precipitation products can be affected by other errors that are linked, for instance, to the characteristics of the rain (i.e., intensity and occurrence), the climate and the environmental conditions, such as the seasonality, topography surface type, etc. Such performance distributions were summarized by Maggioni et al. [11] for the Tropical Rainfall Measuring Mission (TRMM) Era. Gosset et al. [15] highlighted the importance of the density of gauges and its qualities as a reference for satellite-based precipitation assessments, as it is crucial for the success of categorical and continuous verification. In Brazil, the L3 satellite-based precipitation uncertainties, especially in terms of relative bias, were demonstrated to be as-



sociated with distinct conditions, such as the surface type, precipitation regime, topography, gauge density, etc. [41–44].

### 2.3. The Gaussian Mixture Model

Gaussian mixture model (GMM) is a probabilistic function that tends to group multiple Gaussian distributions into a single distribution and can be defined as:

$$p(x_N \mid \alpha_M, \mu_M, \sigma_M) = \sum_{m=1}^M \alpha_m p \quad (3)$$

where  $x_N$  is the data sample,  $\alpha_M$  represents the weight of each component (mixture)  $m = \dots, 1, \dots, M$ . The components, also referred to as the modes, are represented as a normal distribution with parameters' mean  $\mu_M$  and standard deviation  $\sigma_M$  values.

GMM are widely applied for fitting as well as to identify subgroups that compose a given distribution. In meteorology, GMM has been considered, for instance, for forecast verification [45], polarimetric radar-based rainfall-rate estimation [46], predicting precipitation events [47], density estimation and the feature identification of Atmospheric Lagrangian Particle Dispersion Models [48]. Several studies have been conducted so far on the development and improvement of an evolutionary GMM algorithm that automatically fits the distribution by minimizing the given distribution error [49–51].

Here, the GMM method was implemented thanks to the evolutionary “Distribution Optimization” algorithm [51] which separates the  $M$  (modes) that composes the mixture of gaussians. The algorithm involves multiple and sequential steps (i.e., creation of GMMs' population, fitness calculation, filter, mutation, recombination and final check for GMM selection). Finally, the algorithm is complete after a fixed number of iterations by returning the GMM distribution with the highest fitness and its correspondent parameters (i.e.,  $\mu$ ,  $\sigma$ ,  $\alpha$  and Bayesian decision borders) which can be independently used as input parameters. The final algorithm output includes both the individual gaussian elements ( $M$ ) and the associated statistics responsible for describing the overall GMM distribution, which is derived from the sum of all individuals  $M$ . In order to find the GMM with the highest fitness, the Distribution Optimization algorithm considers, as a core, an adjustable error function  $X^2$  based on chi-square statistics and the probability density function (PDF). To minimize the associated distribution error, the fitness function considers multiple statistics, such as best root-mean-square (RMS), the standard deviation, the overlap error, etc. Thus, for each individual GMM element ( $M = 1, 2, \dots$ ), the  $\mu$ ,  $\sigma$ ,  $\alpha$ , the Bayesian decision borders and other associated goodness-of-fit statistics (e.g., Similarity Error, Overlap Error, Mixed Distribution Error) can be retrieved via the Distribution Optimization algorithm. The automated genetic machine-learning GMM “Distribution Optimization” algorithm is available in the R library “DistributionOptimization” (<https://cran.r-project.org/package=DistributionOptimization>, accessed on 1 February 2021) and a detailed description is provided by Lerch et al. [51].

For this study, the GMM distribution fit parameters (i.e.,  $\mu$ ,  $\sigma$ ,  $\alpha$  and Bayesian decision borders) were retrieved to reconstruct and assess the uncertainties using two Gaussians ( $M = 2$ ). The results were contrasted with a single gaussian model ( $M = 1$ ). The number of iterations adopted was 100 (sensitivity tests were performed on the number of iterations starting from 100, with  $M = 1, \dots, 5, 10$  gaussian modes, and revealed no significant gains on the goodness-of-fit results). The goodness-of-fit statistics, such as the Akaike Information Criterion (AIC, [52]), were also computed, in order to demonstrate the final GMM distribution performances across the precipitation bins. The residuals  $\varepsilon$  between the estimated uncertainty distributions from the models (GMM and Gaussian) and the actual uncertainty distribution values OBS were estimated as follows:

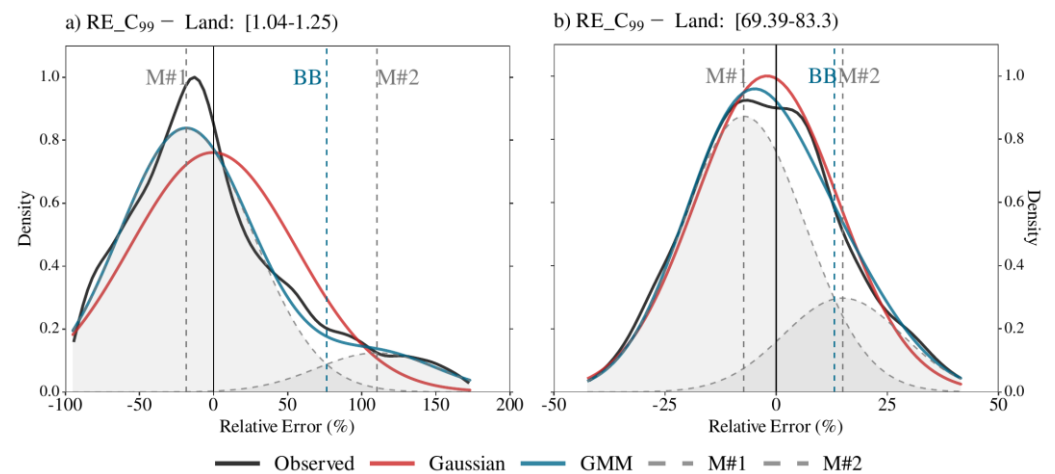
$$\varepsilon = Model_{[GMM, Gaussian]} - OBS \quad (4)$$

### 3. Satellite Precipitation Uncertainty Distribution Approximation

In this section, the three above-described satellite-based precipitation uncertainty distribution types, i.e., (i) constellation change-induced uncertainties; (ii) sampling uncertainty and (iii) comparison with rain-gauges network, were explored using the aforementioned GMM model. The performances of the GMM model were also contrasted with that of a simple Gaussian model as a function of the precipitation accumulation.

#### 3.1. The GMM Model Distributions and Performances

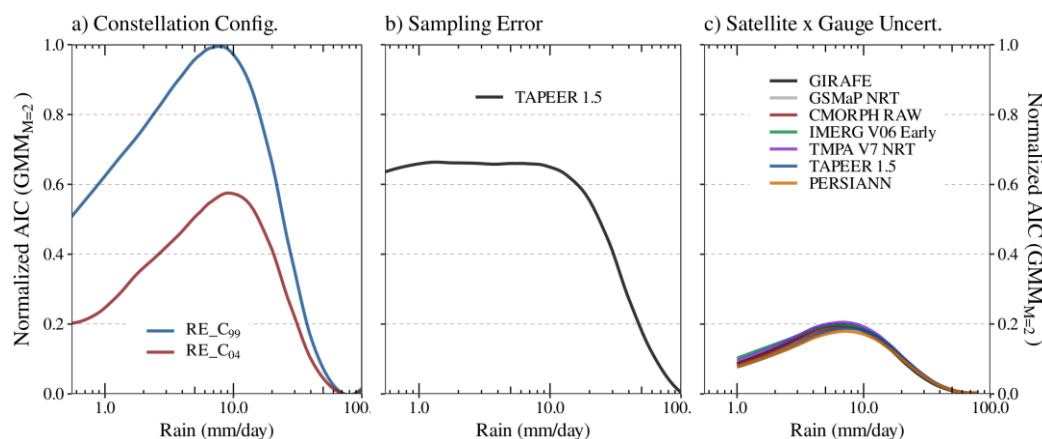
Figure 4 shows an example of the observed uncertainty distributions compared with the GMM and a simple Gaussian fit model for two daily precipitation categories/bins, i.e., at low  $1.04\text{--}1.25\text{ mm day}^{-1}$  (Figure 4a) and large  $69.39\text{--}83.3\text{ mm day}^{-1}$  (Figure 4b) accumulations, through the  $C_{99}$  and the  $C_{REF}$  experiments. As mentioned, with increasing rainfall intensity, the right-skewed uncertainty distribution shape began to acquire a more symmetrical distribution shape with the median centered at zero (less unbiased). For instance, in the low accumulation case, the RE distribution presented a negative median of about  $-9.8\%$ , while the median for precipitation intensity was  $-2.7\%$ .



**Figure 4.** An example of over-land global uncertainty distributions ( $x$ -axis, in %) during the JJA of 2014 for the daily precipitation bins (a)  $1.04\text{--}1.25\text{ mm day}^{-1}$  and (b)  $69.39\text{--}83.3\text{ mm day}^{-1}$ , resulted from the comparison between the  $C_{99}$  and the  $C_{REF}$  experiments. The GMM fit with  $M = 2$  (blue line) is contrasted with a simple Gaussian fit (red line) to represent the observed distribution (black line). Note that GMM “modes” (M#1 and M#2) are in grey filled area dashed curves. The perpendicular dashed lines indicate the Bayesian boundaries between the two GMM modes (in blue) and its correspondent mean values (in grey).

At low daily precipitation amounts (Figure 4a), the outperformance of the GMM model was clearly observed in representing both the RE under- and over-estimations compared to the Gaussian model. However, the models’ representations of the RE distributions at low to moderate daily precipitations were more problematic due its larger sample and the heterogeneity of observations that led to the highest densities (peak numbers and locations). This consequently resulted in larger residuals, especially at the RE underestimations where the maximum densities were found, as the main factor responsible for the precipitation underestimations at this precipitation classes. In contrast, the uncertainty overestimations (positive RE) were well fitted by the GMM model. Yet, despite still being positively skewed, the Gaussian model density peak was more shifted around zero, while the GMM and observed density peaks were approximately  $-25\%$ . At large precipitation accumulations (Figure 4b), both the Gaussian model and the GMM model could reproduce the unimodal RE distributions, resulting in significantly small residual magnitudes, especially at the density peak, and with only slight differences between each other.

Systematic performance, by comparing the GMM and Gaussian models, and as a function of the precipitation accumulation in addition to the contrast between the three types of uncertainty, was better illustrated in the AIC goodness-of-fit statistics (Figure 5). Overall, the GMM skills varied according to precipitation intensity, which was linked to the nature of the precipitation uncertainty source. This also included the constellation changing assumption that presented the largest sensitivity to the AIC results and stability between multiple satellite precipitation products. The superior goodness-of-fit statistics were found for the most part for large precipitation accumulations (substantially increasing greater than  $10 \text{ mm day}^{-1}$ ) regardless of the satellite error source. The maximum AIC results were observed in the range from 5 to  $10 \text{ mm day}^{-1}$ , except for the sampling error (Figure 5b), which also presented high and constant AIC values at precipitation accumulations of  $0.5\text{--}10 \text{ mm day}^{-1}$ . The more different the constellation configuration, the more the model is prone to overfitting (higher AICs) the uncertainty representation (Figure 5a). Figure 5c shows that the GMM proved to be well suited to represent the uncertainties of satellite products when compared to rain gauges, resulting in lower AIC values overall. To better illustrate that, we expanded the AIC performance evaluation to multiple NRT satellite-only precipitation products. In addition to the GIRAFE and the GSMaP NRT precipitation products, the Climate Prediction Center (CPC) morphing technique Version 1.0 RAW (CMORPH-RAW, [53]), the Integrated Multi-satellitE Retrievals for GPM V06—the Early/precipitationCal version (IMERG V06E, [54]), the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis Version 7 Real Time (TMPA-RT, [55]), the TAPEER Version 1.5 (TAPEER 1.5, [20,23]) and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN, [56,57]), were also assessed. Although the main distributions of uncertainty as a function of precipitation accumulation were retained between the products, more apparent AIC goodness-of-fit differences (second-order uncertainty) were detected for low-to-moderate precipitation accumulations. This can be attributed to various factors, including the algorithm itself as well as the source of observations (e.g., PMW rain rates, IR images and so on), which can be the subject of further investigations.

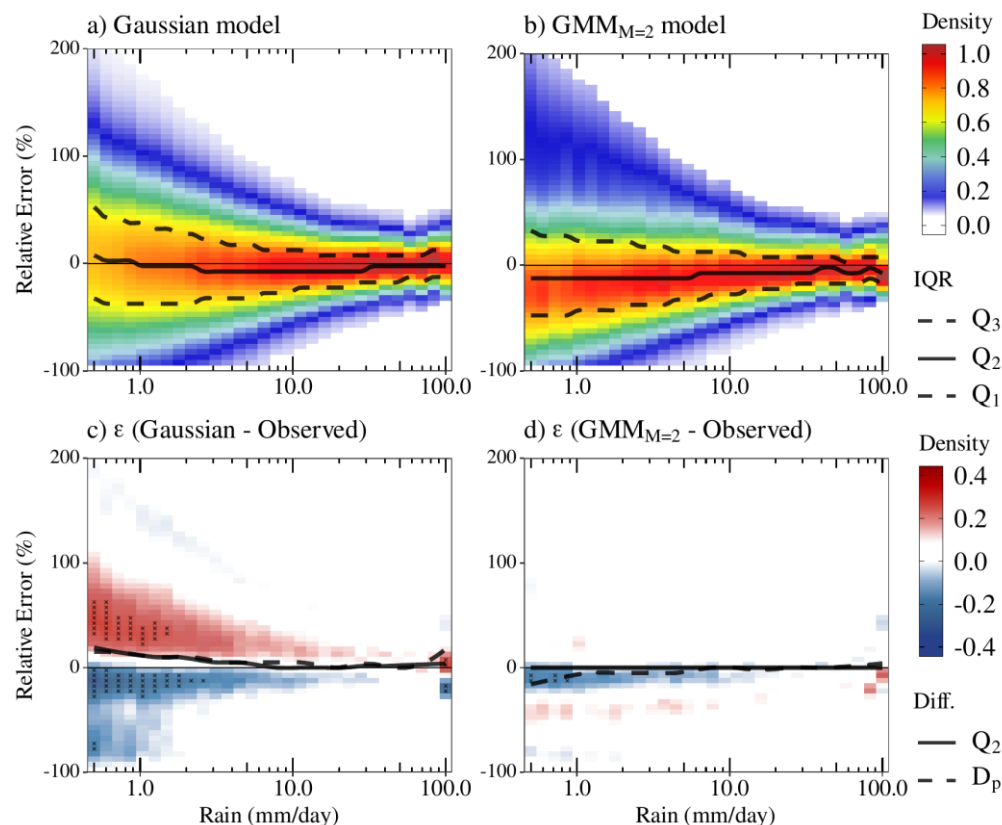


**Figure 5.** The GMM goodness-of-fit through the Akaike information criterion (AIC) statistics, as a function of the precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ). Comparisons between the three error sources considered: (a) due to constellation changes induced uncertainties, (b) due to sampling uncertainty and (c) through a comparison between multiple satellite-based with rain-gauges network. AIC results were normalized by the maximum value of all error sources.

### 3.2. Constellation Changes-Induced Uncertainties

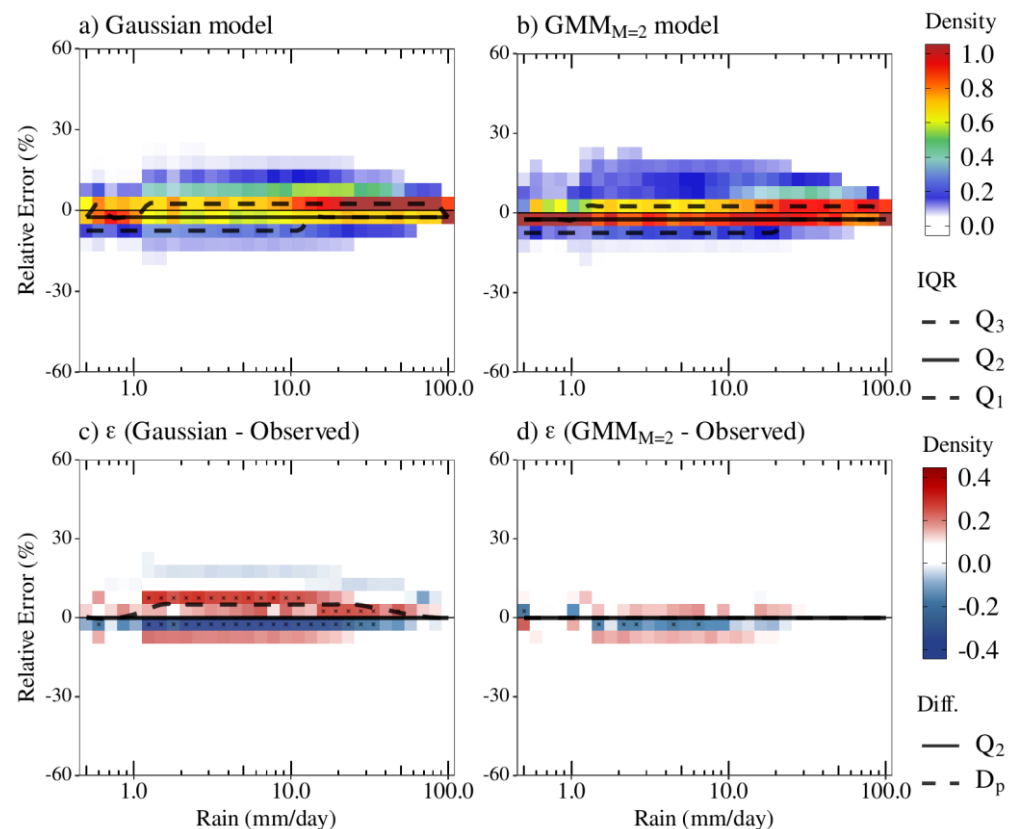
Systematic analysis of the accumulated precipitation from  $0.5$  to  $100 \text{ mm day}^{-1}$  confirmed the case study and revealed the better performance of the mixed model than for the single Gaussian model. The better representation of the observed RE density distributions provided by the GMM model is not related to the experimental configuration

(Figures 6 and 7), despite the better performance for the RE\_C04 experiment (seen the AIC goodness-of-fit comparisons in Figure 5). In the cases of more intense accumulation, even if gain is not as dramatic when using the mixture, the two Gaussian fit outperformed the single Gaussian approach.



**Figure 6.** Density scatterplot of the (a) Gaussian and the (b) GMM (with  $M = 2$ ) models and their respective residuals (c,d) in representing the relative error distributions ( $y$ -axis, in %) as a function of precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for the RE\_C99 experiment, during the JJA of 2014 and over land surfaces (between  $55^\circ\text{N/S}$  global zones). Black crosses indicate residual values above/below the  $\pm 0.2$  density values. Solid and dashed black lines in c-d correspond to the differences between the model and the observed median ( $Q_2$ ) and the density peaks ( $D_p$ ), respectively.

Overall, the largest residual concentrations were of up to  $10 \text{ mm day}^{-1}$  and were more prominent in the single Gaussian model. The Gaussian model largely overestimated (underestimated) the positive (negative) RE distributions, especially in the RE range from 25% to 60% (from  $-30\%$  to  $0\%$ ), which presented  $\epsilon$  of greater (lower) than  $\pm 0.2$ . The Gaussian model also overestimated both the maximum peak (mode) and median of the uncertainty density distributions in about 20% of low-to-moderate rain accumulations. In addition, the Gaussian model was not able to represent the uncertainty distributions at extremes precipitation accumulations (i.e.,  $\geq 100 \text{ mm day}^{-1}$ ). In this case, the Gaussian model uncertainty distributions were less biased (density peak centered roughly the zero—less positively skewed), overestimating the density peak of roughly 17% of uncertainties ranging from  $-5$  to  $15\%$ . This occurred differently than for the observed RE distributions (seen in Figure 1a), which had a slightly right-skewed distribution. To offset that, given the different density distribution shape compared to that which was observed, the Gaussian model also led to an underestimation of the negative uncertainties ranging between 15% and 30%.



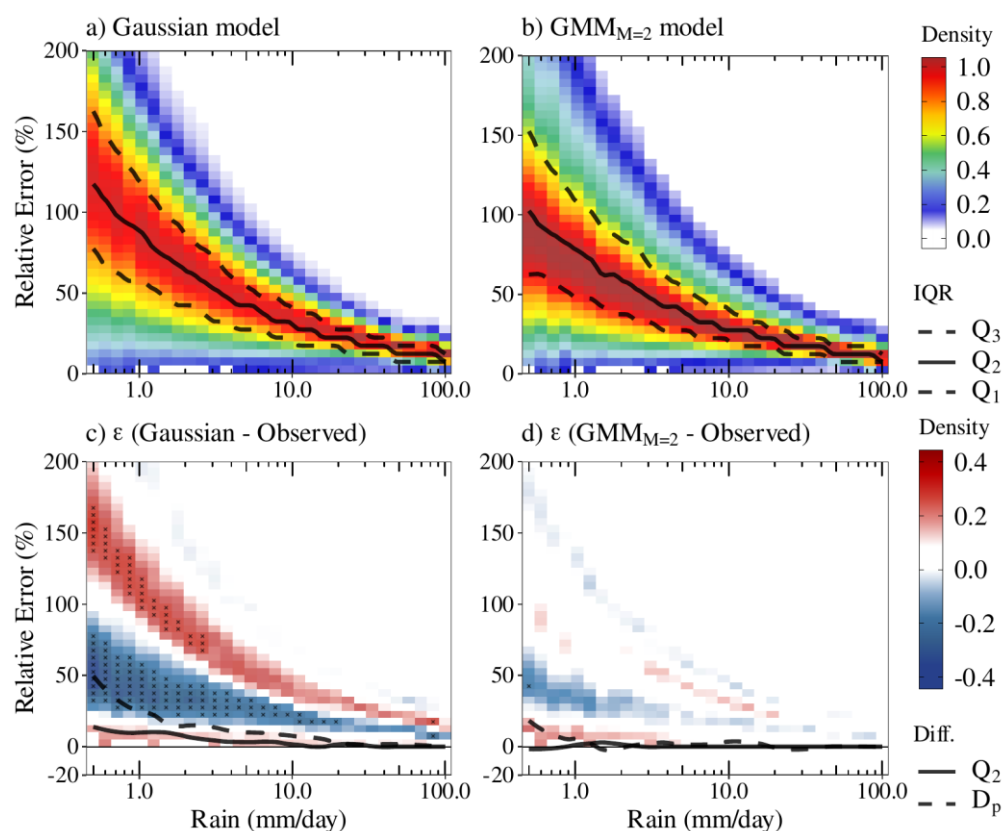
**Figure 7.** Density scatterplot of the (a) Gaussian and the (b) GMM (with  $M = 2$ ) models and their respective residuals (c,d) in representing the relative error distributions ( $y$ -axis, in %) as a function of the precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for the RE\_C<sub>04</sub> experiment, during the JJA of 2014 and over land surfaces (between 55°N/S global zones). Black crosses indicate residual values above/below the  $\pm 0.2$  density values. Solid and dashed black lines in c-d correspond to the differences between the model and the observed median and the density peaks, respectively.

On the other hand, the GMM model could better fit the entire precipitation accumulation range (Figure 6b,d). The outperformance of the GMM model is evidenced in multiple aspects (e.g., residual, density peak, median distribution) and indicates that the uncertainties remain, mainly, in representing the maximum density peaks at approximately up to  $10 \text{ mm day}^{-1}$  (in the range of  $-8\%$  RE on average). Above the precipitation accumulation of  $80 \text{ mm day}^{-1}$ , the GMM model slightly overestimated the density peak (less than 5%) surrounding the 0 and  $-20\%$  RE range. The maximum underestimations (lower than  $-0.2$ ), were found to be of lower than  $1 \text{ mm day}^{-1}$  (in the  $-16\%$  RE range). In general, no considerable (greater than 0.2) RE overestimations were found using the GMM model for both the RE\_C<sub>99</sub> and RE\_C<sub>04</sub> (Figure 7) experiments. The GMM model also stood out by being well suited to more unimodal and leptokurtic distributions over the entire precipitation range (i.e., the RE\_C<sub>04</sub> case), despite slightly underestimating the density peaks. The Gaussian model poorly reproduced the uncertainties across the entire range of precipitation, underestimating the observed distribution peaks at about  $-5\%$  RE. The RE\_C<sub>04</sub> Gaussian distributions were negatively skewed rather than positively skewed.

### 3.3. Sampling Uncertainties

In the same manner, Figure 8 depicts the performance differences between a single Gaussian model and the mixed Gaussians model, but for the sampling uncertainties (over-ocean) scenario. Although both the Gaussian and GMM models were able to reproduce the general observed sampling error distributions (seen in Figure 2), considerable performance differences were observed, especially in capturing the distribution maximum densities.

The positive skewness of the sampling error distributions, especially at low-to-moderate precipitation rates (Figure 2), was not properly captured by considering a single Gaussian model. The main impact was observed on the density peak location, which consequently contributed to large positive and negative residuals ( $>0.2$ ). An alternation between under- and over-estimations, indicating a considerable skewness, was observed for the Gaussian model. Although it tended to decrease as a function of the precipitation accumulation, this feature remained throughout the entire precipitation range. Such distribution had a strong impact at precipitation accumulations of up to  $20 \text{ mm day}^{-1}$ , for which the uncertainty distributions presented an overestimation for both the density peak (from 50% to 10%) and the median (between 7% and 15%).

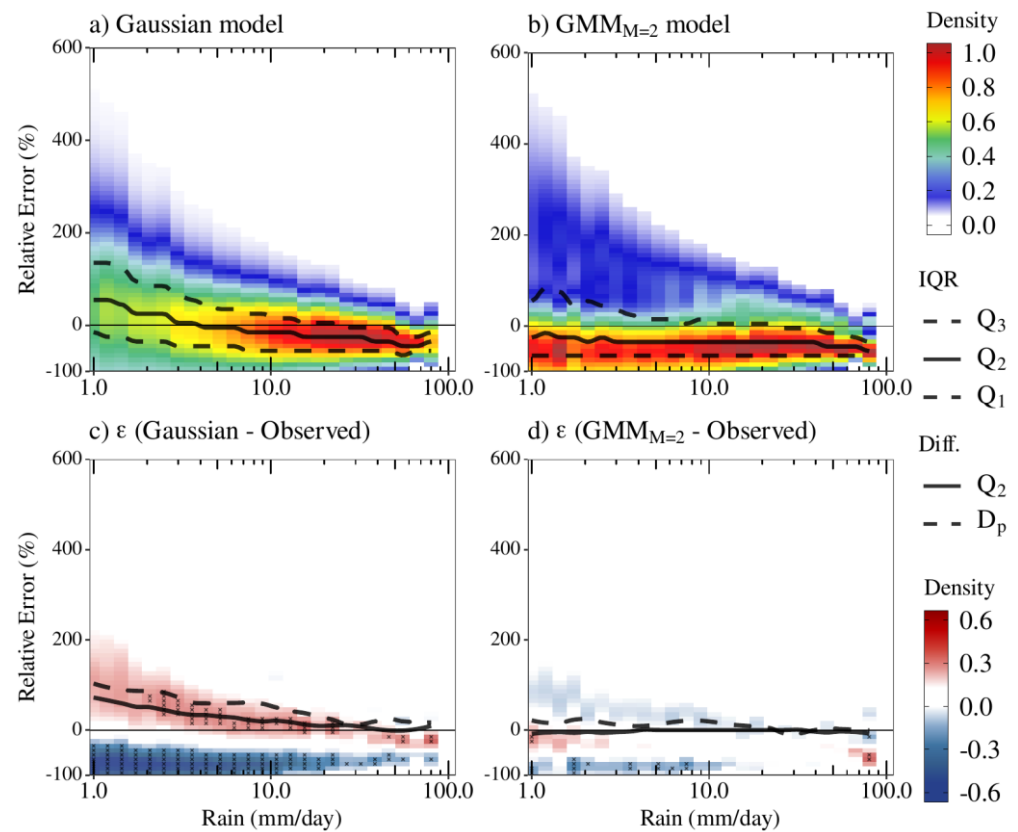


**Figure 8.** Density scatterplot of the (a) Gaussian and the (b) GMM (with  $M = 2$ ) models and its respective residuals (c,d) in representing the relative error distributions ( $y$ -axis, in %) as a function of the precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for the error sampling (by TAPEER1.5 product), during the JJA of 2014 and over ocean surfaces (between  $35^{\circ}\text{N/S}$  global zones). Black crosses mean the residual values above/below the  $\pm 0.2$  density values. Solid and dashed black lines in c-d correspond to the differences between the model and the observed median ( $Q_2$ ) and the density peaks ( $D_p$ ), respectively.

In short, the GMM model could better approximate the sampling uncertainties, being consistent for the most part of the precipitation range. GMM slightly overestimated the density peak at precipitation and at accumulations of lower than  $1 \text{ mm day}^{-1}$ . The outperformances were clearly evident for precipitation amounts of greater than  $20 \text{ mm day}^{-1}$ , where the RE distributions became more symmetrical (less skewed). The GMM model could well represent the sampling uncertainties, providing less residuals and correctly matching the medians and density peaks. The uncertainty underestimations at about  $+50\%$ , for precipitation of up to  $5 \text{ mm day}^{-1}$ , are worth noting. No considerable residuals ( $>\pm 0.2$ ) were observed over much of the precipitation range.

### 3.4. Comparison with Rain-Gauges Network

The ability to approximate the observed satellite-gauge uncertainty (relative bias) distributions using a single versus a mixture of two Gaussians is presented in Figure 9. As mentioned in Section 2.2.3, the main feature of the observed uncertainty distributions is the systematic underestimations found throughout the entire precipitation range (seen in Figure 3). Overall, the GMM model is clearly a more suitable choice compared to the single Gaussian model.



**Figure 9.** Density scatterplot of the (a) Gaussian and the (b) GMM (with  $M = 2$ ) models and its respective residuals (c,d) in representing the relative error distributions ( $y$ -axis, in %) as a function of precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) for the relative bias from the GSMaP NRT V6 precipitation product compared with the ground-based over Brazil (land-only) during the period of DJFM of 2014/5/6. Black crosses indicate residual values above/below the  $\pm 0.2$  density values. Solid and dashed black lines in c-d correspond to the differences between the model and the observed median ( $Q_2$ ) and the density peaks ( $D_p$ ), respectively.

The single Gaussian model showed several weaknesses, especially in representing the observed underestimation peaks—the highest densities of which were found at  $-30\%$  RE (Figure 3). Although there was a tendency for the concentration of residues to decrease as precipitation increased, the pattern of uncertainty was persistent throughout the range of precipitation accumulation. In fact, the Gaussian model tended to increase the weaknesses (less high-skewed) distributions, over- (under-)estimating the positive (negatives) uncertainties and varying as a function of the precipitation accumulations. In addition, an upward shift of the RE medians and density peaks distributions of about  $+60\%$  and  $+80\%$ , respectively, was observed at precipitation accumulations of up to  $10 \text{ mm day}^{-1}$ .

In contrast, the GMM model properly captured the general uncertainty distribution in all precipitation bins (from  $1$  to  $80 \text{ mm day}^{-1}$ ), especially the large positive weakness shapes of low-to-moderate precipitation amounts (e.g.,  $1$ – $10 \text{ mm day}^{-1}$ ). At high precipitation amounts, GMM had an opposite residual performance with a more positively skewed

distribution than the Gaussian model (less right-skewed). The density peak as well as the median were correctly located at the negative RE distribution portion. Nonetheless, although presenting an overall lower residual feature with nearly constant and unbiased median differences, the GMM model slightly shifted the RE density peaks upward (overestimated) in roughly +7% of precipitation of up to 20 mm day<sup>-1</sup>. According to Tian et al. [17], such error modeling weaknesses at large precipitation amounts can be attributed to the model's difficulty in capturing the nonlinear behavior due to satellite data clustering saturation.

#### 4. Discussions and Conclusions

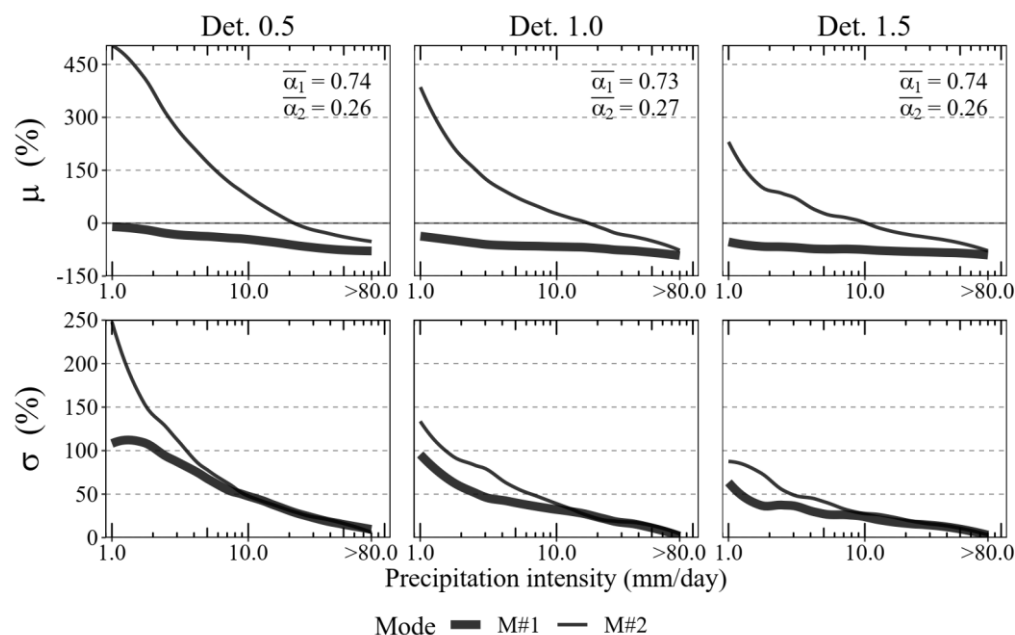
A simple statistical model was introduced to characterize the distribution of uncertainty in gridded multi-platform satellite precipitation products. This Gaussian Mixture Model was shown to outperform a more simplistic single Gaussian model under a variety of ranges of daily accumulated precipitation. It also performed well across a variety of error sources, spanning the classic comparison to gauge networks, to the sampling uncertainty including more recently expressed, constellation configuration-based errors.

Formulating such a simple model provides an additional understanding of using, for instance, a Look Up Table. Firstly, as exemplified in our companion paper on the changes of the constellation and the time-series, it is straightforward to formulate a stochastic perspective on the uncertainty. Oliveira et al. [28] indeed used the Gaussian mixture and a bootstrap technique to estimate the uncertainty distribution over the last two decades and to highlight the relatively stronger impact of moderate-to-high rain accumulation of the evolution of the constellation configuration. Secondly, the suitability of the Gaussian mixture to characterize the uncertainty distribution may provide insights into the processes from which the uncertainty arises. Indeed, our results show that two additive processes can explain the full uncertainty distribution. To further illustrate this perspective, a comparison with the Brazil surface network (Section 3.2) was performed for the GIRAFE product using different detection thresholds in the process of retrieval (Figure 10). The mean of the first mode of the mixture did not reveal considerable sensitivity to this important algorithm assumption. Its variability across the accumulation amount range remained similar for each threshold, despite its slightly decreasing tendencies as a function of the detection threshold. On the other hand, the variance of the mode was strongly impacted by the selection of the threshold, particularly during low-to-moderate accumulation (<10 mm day<sup>-1</sup>). For M#2, the situation was similar with the impact being stronger in the range of low-to-moderate accumulation for both the mean and the variance. This suggests that for M#1 and #2, the impact of the detection step is to add random uncertainty and does not generate much bias at larger rain accumulation (>10 mm day<sup>-1</sup>).

Such preliminary analysis emphasizes that improving the L2 precipitation detection will likely decrease the random uncertainty of the multi-platform product when compared to the rain gauge network. Further analysis should explore the sensitivity of the mixture parameters to other algorithms' intrinsic parameters, such as the propagation of the bias in the L2 estimate in the mixture parameters. The sampling uncertainty is likely to also benefit from the mixture analysis although more work is required to assess this.

The current analysis was restricted to a given additive error model framework, but such a simple model could also be used to explore alternatives such as a multiplicative error framework or rainfall-runoff hydrological simulations [58].





**Figure 10.** Sensitivity of GMM (with  $M = 2$ ) mean ( $\mu$ , upper panels) and standard deviation ( $\sigma$ , bottom panels) parameters as a function of the precipitation ( $x$ -axis, in  $\text{mm day}^{-1}$ ) and according to different PMW detection thresholds of 0.5, 1.0, 1.5  $\text{mm h}^{-1}$  (from left to right panels), from the GIRAFE precipitation product. The weight ( $\alpha$ ) parameters, for both the M#1 and M#2, are shown in the top right of the figure (upper panels) and represent their respective averages as a function of precipitation.

**Author Contributions:** R.A.J.O. and R.R. designed the study and wrote the manuscript. R.A.J.O. conducted the analysis, software and the visualization. R.R. supported writing the original draft preparation and contributed to discussions and revisions. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study benefited from the IPSL mesocenter ESPRI facility, which is supported by CNRS, Sorbonne Université, UVSQ, CNES, Ecole Polytechnique and national research infrastructures Climeri-France and DATA TERRA. This study was performed by the EUMETSAT Satellite Application Facility on Climate Monitoring, and we acknowledge the financial support from the EUMETSAT member states through CM SAF.

**Data Availability Statement:** The TAPEER data are available at <https://www.icare.univ-lille.fr/megha-tropiques/products/> (accessed on 1 February 2021). The GSMaP Near Real Time V6 Precipitation data were acquired from <https://sharaku.eorc.jaxa.jp/GSMaP/> (accessed on 1 February 2021). The GPM IMERG Early Precipitation L3 Half Hourly V06 data were obtained from the Goddard Earth Sciences Data and Information Services Center (GES DISC; <https://doi.org/10.5067/GPM/IMERG/3B-HH-E/06> (accessed on 1 February 2021)). PERSIANN data were obtained from Center for Hydrometeorology and Remote Sensing at the University of California, Irvine (CHRS, UCI, <ftp://persiann.eng.uci.edu/CHRSdata/PERSIANN> (accessed on 1 February 2021)). TRMM (TMPA) Rainfall Estimate L3 3 hourly V7 data were obtained from GES DISC (<https://doi.org/10.5067/TRMM/TMPA/3H/7> (accessed on 1 February 2021)). The CMORPH V1.0 RAW data are available at <ftp://ftp.cpc.ncep.noaa.gov/precip/> (accessed on 1 February 2021). All the figures have been created using the R software package (version 3.6.3 for Linux; <http://CRAN.R-project.org/>). R Development Core Team. R: A Language and Environment for Statistical Computing. (2008). R package “DistributionOptimization” algorithm is available through the <https://cran.r-project.org/package=DistributionOptimization> (accessed on 1 February 2021).

**Acknowledgments:** We thank S. Cloché for her support with the handling of these various datasets. This study benefited from the IPSL mesocenter ESPRI facility, which is supported by CNRS, UPMC, Labex L-IPSL, CNES and Ecole Polytechnique. The authors acknowledge the EUMETSAT support under the CMSAF program.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Stephens, G.L.; Slingo, J.M.; Rignot, E.; Reager, J.T.; Hakuba, M.Z.; Durack, P.J.; Worden, J.; Rocca, R. Earth's water reservoirs in a changing climate. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2020**, *476*, 20190458. [[CrossRef](#)] [[PubMed](#)]
2. Levizzani, V.; Kidd, C.; Aonashi, K.; Bennartz, R.; Ferraro, R.R.; Huffman, G.; Roca, R.; Turk, F.J.; Wang, N. The activities of the international precipitation working group. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 3–15. [[CrossRef](#)] [[PubMed](#)]
3. Roca, R.; Alexander, L.V.; Potter, G.; Bador, M.; Jucá, R.; Contractor, S.; Bosilovich, M.G.; Cloché, S. FROGS: A daily  $1 \times 1$  gridded precipitation database of rain gauge, satellite and reanalysis products. *Earth Syst. Sci. Data* **2019**, *11*, 1017–1035. [[CrossRef](#)]
4. Madakumbura, G.D.; Thackeray, C.W.; Norris, J.; Goldenson, N.; Hall, A. Anthropogenic influence on extreme precipitation over global land areas seen in multiple observational datasets. *Nat. Commun.* **2021**, *12*, 3944. [[CrossRef](#)] [[PubMed](#)]
5. Tapiador, F.; Navarro, A.; Levizzani, V.; García-Ortega, E.; Huffman, G.; Kidd, C.; Kucera, P.; Kummerow, C.; Masunaga, H.; Petersen, W.; et al. Global precipitation measurements for validating climate models. *Atmos. Res.* **2017**, *197*, 1–20. [[CrossRef](#)]
6. Martinez-Villalobos, C.; Neelin, J.D. Climate models capture key features of extreme precipitation probabilities across regions. *Environ. Res. Lett.* **2021**, *16*, 024017. [[CrossRef](#)]
7. Roca, R. Estimation of extreme daily precipitation thermodynamic scaling using gridded satellite precipitation products over tropical land. *Environ. Res. Lett.* **2019**, *14*, 095009. [[CrossRef](#)]
8. Alexander, L.V.; Bador, M.; Roca, R.; Contractor, S.; Donat, M.G.; Nguyen, P.L. Intercomparison of annual precipitation indices and extremes over global land areas from in situ, space-based and reanalysis products. *Environ. Res. Lett.* **2020**, *15*, 055002. [[CrossRef](#)]
9. Roca, R.; Haddad, Z.S.; Akimoto, F.F.; Alexander, L.; Behrangi, A.; Huffman, G.; Kato, S.; Kidd, C.; Kirstetter, P.-E.; Kubota, T.; et al. *The Joint IPWG/GEWEX Precipitation Assessment*; Roca, R., Haddad, Z.S., Eds.; World Climate Research Programme: Geneva, Switzerland, 2021; p. 125.
10. Elsaesser, G.S.; Kummerow, C.D. The Sensitivity of Rainfall Estimation to Error Assumptions in a Bayesian Passive Microwave Retrieval Algorithm. *J. Appl. Meteorol. Clim.* **2015**, *54*, 408–422. [[CrossRef](#)]
11. Maggioni, V.; Meyers, P.C.; Robinson, M.D. A Review of Merged High-Resolution Satellite Precipitation Product Accuracy during the Tropical Rainfall Measuring Mission (TRMM) Era. *J. Hydrometeorol.* **2016**, *17*, 1101–1117. [[CrossRef](#)]
12. Haddad, Z.S.; Turk, F.J.; Utsumi, N.; Kirstetter, P. *Assessment of the Sub-Daily Global Satellite Precipitation Products*; Roca, R., Haddad, Z.S., Eds.; World Climate Research Programme (WCRP): Geneva, Switzerland, 2021; pp. 1–22.
13. Maggioni, V.; Sapiiano, M.R.P.; Adler, R.F.; Tian, Y.; Huffman, G.J. An Error Model for Uncertainty Quantification in High-Time-Resolution Precipitation Products. *J. Hydrometeorol.* **2014**, *15*, 1274–1292. [[CrossRef](#)]
14. Oliveira, R.; Maggioni, V.; Vila, D.; Porcaccia, L. Using Satellite Error Modeling to Improve GPM-Level 3 Rainfall Estimates over the Central Amazon Region. *Remote Sens.* **2018**, *10*, 336. [[CrossRef](#)]
15. Gosset, M.; Alcoba, M.; Roca, R.; Cloché, S.; Urbani, G. Evaluation of TAPEER daily estimates and other GPM-era products against dense gauge networks in West Africa, analysing ground reference uncertainty. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 255–269. [[CrossRef](#)]
16. Chambon, P.; Jobard, I.; Roca, R.; Viltard, N. An investigation of the error budget of tropical rainfall accumulation derived from merged passive microwave and infrared satellite measurements. *Q. J. R. Meteorol. Soc.* **2012**, *139*, 879–893. [[CrossRef](#)]
17. Tian, Y.; Huffman, G.J.; Adler, R.F.; Tang, L.; Sapiiano, M.; Maggioni, V.; Wu, H. Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.* **2013**, *40*, 2060–2065. [[CrossRef](#)]
18. Chambon, P.; Roca, R.; Jobard, I.; Capderou, M. The Sensitivity of Tropical Rainfall Estimation from Satellite to the Configuration of the Microwave Imager Constellation. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 996–1000. [[CrossRef](#)]
19. Roca, R.; Taburet, N.; Lorant, E.; Chambon, P.; Alcoba, M.; Brogniez, H.; Cloché, S.; Dufour, C.; Gosset, M.; Guilloteau, C. Quantifying the contribution of the Megha-Tropiques mission to the estimation of daily accumulated rainfall in the Tropics. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 49–63. [[CrossRef](#)]
20. Roca, R.; Chambon, P.; Jobard, I.; Kirstetter, P.-E.; Gosset, M.; Bergès, J.C. Comparing Satellite and Surface Rainfall Products over West Africa at Meteorologically Relevant Scales during the AMMA Campaign Using Error Estimates. *J. Appl. Meteorol. Clim.* **2010**, *49*, 715–731. [[CrossRef](#)]
21. AghaKouchak, A.; Mehran, A.; Norouzi, H.; Behrangi, A. Systematic and random error components in satellite precipitation data sets. *Geophys. Res. Lett.* **2012**, *39*. [[CrossRef](#)]
22. Roca, R.; Guérou, A.; Oliveira, R.A.J.; Chambon, P.; Gosset, M.; Cloché, S.; Schröder, M. Merging the Infrared Fleet and the Microwave Constellation for Tropical Hydrometeorology (TAPEER) and Global Climate Monitoring (GIRAFE) Applications. In *Satellite Precipitation Measurement. In Advances in Global Change Research*; Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., Turk, F., Eds.; Springer: Cham, Switzerland, 2020; Volume 67, pp. 429–450. [[CrossRef](#)]
23. Chambon, P.; Roca, R.; Jobard, I.; Aublanc, J. The TAPEER-BRAIN product: Algorithm theoretical basis document, level 4. *Megha-Tropiques Technol. Memo* **2012**, *4*, 13.
24. Xu, L.; Gao, X.; Sorooshian, S.; Arkin, P.; Imam, B. A Microwave Infrared Threshold Technique to Improve the GOES Precipitation Index. *J. Appl. Meteorol.* **1999**, *38*, 569–579. [[CrossRef](#)]

25. Kidd, C.; Kniveton, D.R.; Todd, M.C.; Bellerby, T.J. Satellite Rainfall Estimation Using Combined Passive Microwave and Infrared Algorithms. *J. Hydrometeorol.* **2003**, *4*, 1088–1104. [[CrossRef](#)]
26. Kummerow, C.D.; Randel, D.L.; Kulie, M.; Wang, N.-Y.; Ferraro, R.; Munchak, S.J.; Petkovic, V. The Evolution of the Goddard Profiling Algorithm to a Fully Parametric Scheme. *J. Atmospheric Ocean. Technol.* **2015**, *32*, 2265–2280. [[CrossRef](#)]
27. Kidd, C.; Matsui, T.; Ringerud, S. Precipitation Retrievals from Passive Microwave Cross-Track Sensors: The Precipitation Retrieval and Profiling Scheme. *Remote Sens.* **2021**, *13*, 947. [[CrossRef](#)]
28. Oliveira, R.A.J.; Roca, R.; Finkensieper, S.; Cloché, S.; Schröder, M. A time-dependent error model for satellite constellation-based daily precipitation estimates. *Atmos. Res.* **2022**, *accepted*.
29. Oliveira, R.A.J.; Gosset, M.; Roca, R.; Kidd, C. Impact of Level-2 satellite rainfall retrievals' characteristics on daily accumulated rainfall estimates: A sensitivity analysis based on the TAPEER framework. **2022**, *in preparation*.
30. Guilloteau, C.; Foufoula-Georgiou, E.; Kummerow, C.D. Global Multiscale Evaluation of Satellite Passive Microwave Retrieval of Precipitation during the TRMM and GPM Eras: Effective Resolution and Regional Diagnostics for Future Algorithm Development. *J. Hydrometeorol.* **2017**, *18*, 3051–3070. [[CrossRef](#)]
31. Guilloteau, C.; Foufoula-Georgiou, E.; Kirstetter, P.; Tan, J.; Huffman, G.J. How Well Do Multisatellite Products Capture the Space-Time Dynamics of Precipitation? Part I: Five Products Assessed via a Wavenumber-Frequency Decomposition. *J. Hydrometeorol.* **2021**, *22*, 2805–2823.
32. Kidd, C.; Tan, J.; Kirstetter, P.-E.; Petersen, W.A. Validation of the Version 05 Level 2 precipitation products from the GPM Core Observatory and constellation satellite sensors. *Q. J. R. Meteorol. Soc.* **2018**, *144*, 313–328. [[CrossRef](#)]
33. Kidd, C.; Huffman, G.; Maggioni, V.; Chambon, P.; Oki, R. The Global Satellite Precipitation Constellation: Current Status and Future Requirements. *Bull. Am. Meteorol. Soc.* **2021**, *102*, E1844–E1861. [[CrossRef](#)]
34. Tan, J.; Petersen, W.A.; Kirchengast, G.; Goodrich, D.C.; Wolff, D.B. Evaluation of Global Precipitation Measurement Rainfall Estimates against Three Dense Gauge Networks. *J. Hydrometeorol.* **2018**, *19*, 517–532. [[CrossRef](#)]
35. You, Y.; Petkovic, V.; Tan, J.; Kroodsmas, R.; Berg, W.; Kidd, C.; Peters-Lidard, C. Evaluation of V05 Precipitation Estimates from GPM Constellation Radiometers Using KuPR as the Reference. *J. Hydrometeorol.* **2020**, *21*, 705–728. [[CrossRef](#)]
36. Roca, R.; Brogniez, H.; Chambon, P.; Chomette, O.; Cloché, S.; Gosset, M.E.; Mahfouf, J.-F.; Raberanto, P.; Viltard, N. The Megha-Tropiques mission: A review after three years in orbit. *Front. Earth Sci.* **2015**, *3*, 17. [[CrossRef](#)]
37. Kubota, T.; Aonashi, K.; Ushio, T.; Shige, S.; Takayabu, Y.N.; Kachi, M.; Arai, Y.; Tashima, T.; Kawamoto, N.; Mega, T.; et al. Global Satellite Mapping of Precipitation (GSMaP) Products in the GPM Era. In *Satellite Pre-Cipitation Measurement*; Levizzani, V., Kidd, C., Kirschbaum, D., Kummerow, C., Nakamura, K., Turk, F.J., Eds.; Springer: Cham, Switzerland, 2020; pp. 355–373.
38. Reboita, M.S.; Gan, M.A.; Da Rocha, R.P.; Ambrizzi, T. Regimes de precipitação na América do Sul: Uma revisão bibliográfica. *Rev. Bras. Meteorol.* **2010**, *25*, 185–204. [[CrossRef](#)]
39. Ebert, E.E.; Janowiak, J.E.; Kidd, C. Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. *Bull. Am. Meteorol. Soc.* **2007**, *88*, 47–64. [[CrossRef](#)]
40. Maggioni, V.; Massari, C.; Kidd, C. Errors and uncertainties associated with quasiglobal satellite precipitation products. In *Precipitation Science*; Michaelides, S., Ed.; Elsevier: Amsterdam, The Netherlands, 2022; Volume 1, pp. 377–390. [[CrossRef](#)]
41. Oliveira, R.; Maggioni, V.; Vila, D.; Morales, C. Characteristics and Diurnal Cycle of GPM Rainfall Estimates over the Central Amazon Region. *Remote Sens.* **2016**, *8*, 544. [[CrossRef](#)]
42. Rozante, J.R.; Vila, D.A.; Chiquetto, J.B.; Fernandes, A.D.A.; Alvim, D.S. Evaluation of TRMM/GPM Blended Daily Products over Brazil. *Remote Sens.* **2018**, *10*, 882. [[CrossRef](#)]
43. Gadelha, A.N.; Coelho, V.H.R.; Xavier, A.C.; Barbosa, L.R.; Melo, D.C.; Xuan, Y.; Huffman, G.J.; Petersen, W.A.; Almeida, C.D.N. Grid box-level evaluation of IMERG over Brazil at various space and time scales. *Atmos. Res.* **2018**, *218*, 231–244. [[CrossRef](#)]
44. Satgé, F.; Pillot, B.; Roig, H.; Bonnet, M.-P. Are gridded precipitation datasets a good option for streamflow simulation across the Juruá river basin, Amazon? *J. Hydrol.* **2021**, *602*, 126773. [[CrossRef](#)]
45. Lakshmanan, V.; Kain, J.S. A Gaussian Mixture Model Approach to Forecast Verification. *Weather Forecast.* **2010**, *25*, 908–920. [[CrossRef](#)]
46. Li, Z.; Zhang, Y.; Giangrande, S.E. Rainfall-Rate Estimation Using Gaussian Mixture Parameter Estimator: Training and Validation. *J. Atmos. Ocean. Technol.* **2012**, *29*, 731–744. [[CrossRef](#)]
47. Ling, H.; Zhu, K. Predicting Precipitation Events Using Gaussian Mixture Model. *J. Data Anal. Inf. Process.* **2017**, *5*, 131–139. [[CrossRef](#)]
48. Crawford, A. The Use of Gaussian Mixture Models with Atmospheric Lagrangian Particle Dispersion Models for Density Estimation and Feature Identification. *Atmosphere* **2020**, *11*, 1369. [[CrossRef](#)]
49. Scrucca, L. GA: A Package for Genetic Algorithms in R. *J. Stat. Softw.* **2013**, *53*, 1–37. [[CrossRef](#)]
50. Ultsch, A.; Thrun, M.C.; Hansen-Goos, O.; Lötsch, J. Identification of Molecular Fingerprints in Human Heat Pain Thresholds by Use of an Interactive Mixture Model R Toolbox (AdaptGauss). *Int. J. Mol. Sci.* **2015**, *16*, 25897–25911. [[CrossRef](#)] [[PubMed](#)]
51. Lerch, F.; Ultsch, A.; Lötsch, J. Distribution Optimization: An evolutionary algorithm to separate Gaussian mixtures. *Sci. Rep.* **2020**, *10*, 648. [[CrossRef](#)]
52. Akaike, H. A new look at the statistical model identification. *IEEE Trans. Autom. Control* **1974**, *19*, 716–723. [[CrossRef](#)]
53. Joyce, R.; 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. [[CrossRef](#)]

54. Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.; Joyce, R.; Xie, P.; Yoo, S.H. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). In *NASA Algorithm Theoretical Basis Document, Version 06*; NASA: Washington, DC, USA, 2019; p. 38. Available online: [https://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V06.pdf](https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf) (accessed on 1 February 2021).
55. 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. [[CrossRef](#)]
56. Hsu, K.; Gao, X.; Sorooshian, S.; Gupta, H.V. Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteorol.* **1997**, *36*, 1176–1190. [[CrossRef](#)]
57. Sorooshian, S.; Hsu, K.; Gao, X.; Gupta, H.V.; Imam, B.; Braithwaite, D. Evolution of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteorol. Soc.* **2000**, *81*, 2035–2046. [[CrossRef](#)]
58. de Paiva, R.C.D.; Buarque, D.C.; Collischonn, W.; Bonnet, M.-P.; Frappart, F.; Calmant, S.; Bulhões Mendes, C.A. Large-scale hydrologic and hydrodynamic modeling of the Amazon River basin. *Water Resour. Res.* **2013**, *49*, 1226–1243. [[CrossRef](#)]