



Article

# A Novel Bias Correction Method for Soil Moisture and Ocean Salinity (SMOS) Soil Moisture: Retrieval Ensembles

Ju Hyoung Lee <sup>†</sup> and Jungho Im <sup>\*</sup>

Received: 27 September 2015; Accepted: 25 November 2015; Published: 2 December 2015

Academic Editors: Nicolas Baghdadi and Prasad S. Thenkabail

School of Urban and Environmental Engineering, Ulsan National Institute of Science and Technology (UNIST), UNIST-gil 50, Eonyang-eup, Ulju-gun, Ulsan 44919, Korea

<sup>\*</sup> Correspondence: [ersgis@unist.ac.kr](mailto:ersgis@unist.ac.kr); Tel: +82-52-217-2824; Fax: +82-52-217-2819

<sup>†</sup> Current address: Department of Environmental Engineering, Politecnico di Milano, Piazza Leonardo da Vinci 32, Milan 20133, Italy; [ju.lee@mail.com](mailto:ju.lee@mail.com)

**Abstract:** Bias correction is a very important pre-processing step in satellite data assimilation analysis, as data assimilation itself cannot circumvent satellite biases. We introduce a retrieval algorithm-specific and spatially heterogeneous Instantaneous Field of View (IFOV) bias correction method for Soil Moisture and Ocean Salinity (SMOS) soil moisture. To the best of our knowledge, this is the first paper to present the probabilistic presentation of SMOS soil moisture using retrieval ensembles. We illustrate that retrieval ensembles effectively mitigated the overestimation problem of SMOS soil moisture arising from brightness temperature errors over West Africa in a computationally efficient way (ensemble size: 12, no time-integration). In contrast, the existing method of Cumulative Distribution Function (CDF) matching considerably increased the SMOS biases, due to the limitations of relying on the imperfect reference data. From the validation at two semi-arid sites, Benin (moderately wet and vegetated area) and Niger (dry and sandy bare soils), it was shown that the SMOS errors arising from rain and vegetation attenuation were appropriately corrected by ensemble approaches. In Benin, the Root Mean Square Errors (RMSEs) decreased from  $0.1248 \text{ m}^3/\text{m}^3$  for CDF matching to  $0.0678 \text{ m}^3/\text{m}^3$  for the proposed ensemble approach. In Niger, the RMSEs decreased from  $0.14 \text{ m}^3/\text{m}^3$  for CDF matching to  $0.045 \text{ m}^3/\text{m}^3$  for the ensemble approach.

**Keywords:** bias correction; SMOS soil moisture data assimilation; brightness temperature (TB) ensembles; West Africa

## 1. Introduction

Data assimilation for merging satellite observations with model states is widely applied to provide the boundary conditions with high quality for Numerical Weather Prediction (NWP) model predictions, as well as to improve the spatial and temporal resolutions of satellite data [1,2]. This method is preferred to model forecasting because models always rely on various assumptions to solve related equations and require a calibration for various input parameters. Satellite data such as soil moisture also have limitations in that they are not available over extreme regions, such as highly vegetated, flooded, or frozen soils, and there are several retrieval errors, for example in the event of rain or the presence of Radio-Frequency Interference (RFI). Thus, the data assimilation that can systematically integrate observations with model states has been considered optimal.

However, some pre-processing tasks are needed for a successful data assimilation. A data assimilation system tends to shift model states towards observations, whose error distribution is assumed to be random [3]. However, in addition to random errors, satellite data are also

contaminated by several systematic errors. For instance, Gruhier *et al.* [4] reported that L-band passive microwave satellite-retrieved soil moisture data produced large errors, especially in dry and sandy soils such as those in West Africa. Under such conditions, data assimilation reproduces large errors while converging towards the erroneous satellite observations. In this process, the Quality Control (QC) science flag is not helpful [5], resulting in a failure of data assimilation. Therefore, it is essential to acquire high quality observations prior to data assimilation [6,7].

In this context, various satellite bias correction methods have been developed and applied to operational services for data assimilation [8,9]. Dee and Uppala [10] previously stated that the bias correction methods implemented at the National Centers for Environmental Prediction (NCEP) and in the European Centre for Medium-Range Weather Forecasts (ECMWF) data assimilation system are mainly intended to account for systematic errors in satellite observations. They modeled the attribution of satellite biases with a simplified linear predictor or bias parameter, which considers atmospheric conditions, instrument scanning bias, or viewing angle relative to the nadir. The main scientific challenge in applying this atmospheric (e.g., upper stratosphere or cloud detection) approach to soil moisture microwave sensors is how to specify the non-linear dynamics of satellite biases in the light of land surface conditions such as soil type, vegetation condition, or rainfall events.

Due to the complexity arising from non-linear error dynamics and land surface heterogeneity, in the field of soil moisture, the Cumulative Distribution Function (CDF) matching technique has been widely used for correcting soil moisture satellite biases [5,7,11]. Fundamentally, it shifts satellite-derived soil moisture data towards climatology obtained from a long record of reference data in order to minimize their difference. However, this approach is different in many aspects from the bias correction of Dee and Uppala [10] stated above. First, CDF matching considers the stationary errors based upon a long period (~10 years) of climatology reference data [11]. However, the data assimilation analysis is often applied to the initialization of a short-range NWP system or land surface model that just evolves for several days. In addition, 4% of Root Mean Square Error (RMSE) accuracy goals of SMOS or Soil Moisture Active Passive (SMAP) missions aim for IFOV satellite errors rather than the long-term trends of climatology [12]. Although retrieval errors can significantly alter even a single satellite product or seasonal dynamics, these systematic errors are not evaluated by CDF matching. Secondly, as it is difficult to acquire consistently high quality reference data of several years, it is unavoidable to transfer the errors inherent in reference data during the matching process. This cannot be resolved by replacing the reference data with other data sets, as there is no perfect data set to rely on (either other satellite sensors or model estimates). Thirdly, as CDF matching is based on the assumption that rescaling parameters are systematically independent of the type of sensors or the retrieval algorithms [13], it is neither specific to systematic errors in the particular version of a retrieval algorithm used nor sensitive to the systematic effects from an incidence angle, frequency, or calibration of sensors [14]. Thus, rescaled results for different remote sensors remain the same if not changing the reference data used for rescaling. Fourthly, a temporally dynamic variation of rescaling parameters was ignored in the implementation of CDF matching, although it was previously suggested that there is at least seasonal variability [13,15]. Taken together, CDF matching is independent of biases in the satellite sensor or retrieval algorithm, as the biases in CDF matching are fundamentally defined by comparative judgment between the reference data and satellite retrievals. Thus, it is argued that there is a need to operationally detect and mitigate the systematic and absolute IFOV errors of satellite retrievals, because the observational errors in operational data assimilation systems are currently set at a global constant value.

To overcome such problems and to operationally mitigate retrieval algorithm-specific and spatially heterogeneous IFOV biases of satellite retrievals, we suggest a novel scheme employing retrieval ensembles. An ensemble approach is one of the emerging techniques widely applied to data assimilation, weather prediction, and climate systems [16–18]. Several studies have successfully applied an ensemble approach for satellite retrieval error analysis. In order to improve the quality of satellite retrievals, Zhang *et al.* [19] generated retrieval ensembles of atmospheric profiles from

the Atmospheric InfraRed Sounder (AIRS) with the perturbation of the temperature eigenvectors, and determined retrieval errors from them. Hossain and Anagnostou [20] found the optimal value of rainfall data with satellite retrieval ensembles. They successfully exhibited the probabilistic representation of satellite-retrieved rainfall products by establishing the stochastic error structure of retrieval products. Olson *et al.* [21] propagated the random errors to measure rainfall errors from passive microwave radiometer observations, and successfully quantified the IFOV rainfall errors at low resolutions.

The objectives of this study are to provide a rescaling function with retrieval ensembles and to mitigate SMOS errors over West Africa, as this region is exposed to several complicated retrieval errors such as a high vertical gradient of soil moisture during the West African Monsoon (WAM) period and the presence of vegetation, and as it is also difficult to simulate dry soils with models [22]. This paper is organized as follows. In Section 2, a brief overview of the SMOS radiative algorithm and the theoretical rationale of the ensemble approach is introduced along with the methods of CDF matching and ensemble generation skills. The rescaling results from CDF matching and retrieval ensembles are compared and discussed in Section 3. The discussion in Section 4 describes the operational merits, uncertainties, and limitations of the proposed approach. Conclusions are given in Section 5.

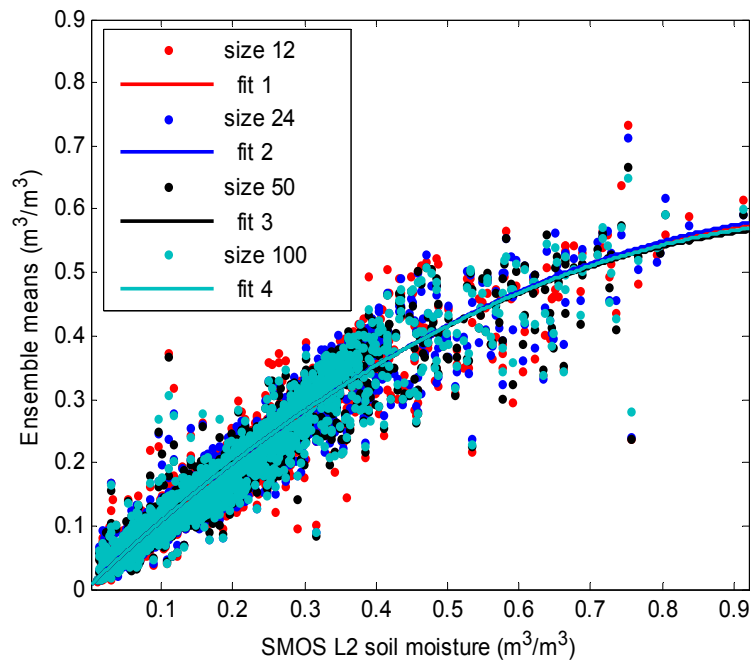
## 2. Data and Methods

The study domain situated at the sub-Saharan area (5–16° N and 10° W–10° E; see Supplementary Material Figure S1) was previously illustrated by Lee *et al.* [23]. There is a clear latitudinal gradient over this study domain. The Niger supersite (*i.e.*, 50 km by 40 km) in the Sahel region is very dry. The dry air mass from the North is transported across the Sahara. In contrast, the Benin supersite in the Guinean region is moderately wet. Vegetation types in Benin include forest, fallows, and crops [24]. The vegetated area in Benin is influenced by moist air from the Gulf of Guinea. The rainfall gauge network data in Benin showed 1289.09 mm during the experiment period, while it was 358.52 mm in Niger [23]. The Analyses Multidisciplinaires de la Mousson Africaine (AMMA) field campaign for this region was established in Benin and Niger for the purpose of calibration and validation of satellite data, as previously introduced by [25,26]. Detailed super-site locations for Benin and Niger are available at <http://www.amma-catch.org/>. The soil moisture field measurements in 2010 were upscaled by taking a linear spatial average of eight probe measurements deployed at a depth of 5 cm within a single pixel at 0.25 degrees. AMMA field campaign data are available at the international soil moisture network (<https://ismn.geo.tuwien.ac.at/>).

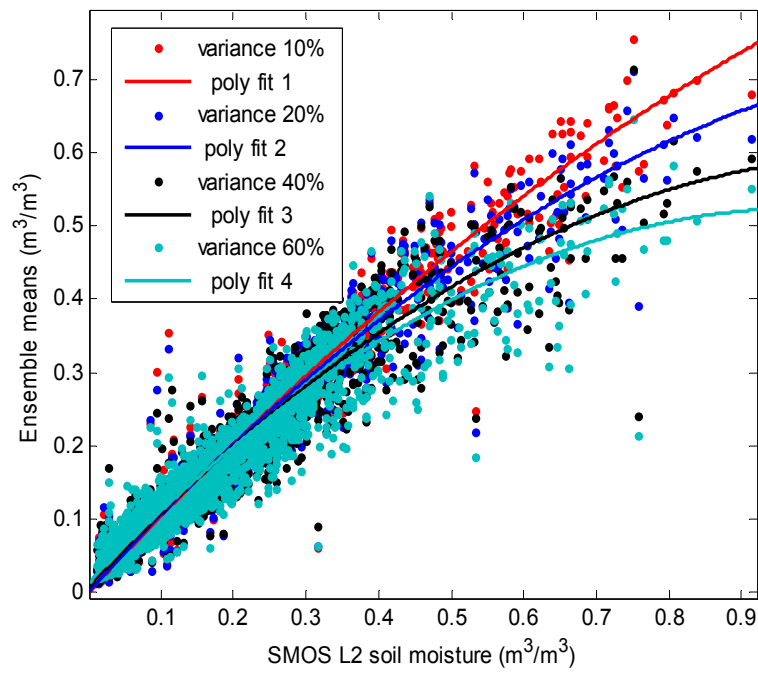
### 2.1. SMOS Retrieval Algorithm

A detailed theoretical description of the L-band Microwave Emission of the Biosphere (L-MEB) radiative transfer forward model was previously provided by Kerr *et al.* [27]. In this paper, we focus on the key elements of the radiative transfer model for the purpose of articulating SMOS retrieval error sources.

The SMOS retrieval algorithm employs a Bayesian approach that minimizes the difference between brightness temperature measurements and simulations to non-linearly retrieve geophysical parameters including soil moisture [28]. Through iterations, *a priori* default parameters are updated until the simulations are matched with the SMOS measurements. From a model library, it selects a retrieval model that represents the dominant landscape in a given node (refer to Figure 1 in Kerr *et al.* [27]). This dominant landscape is determined by the aggregation of sub-pixel land cover fractions from the ECOCLIMAP data set, a global database of land surface parameters at 1 km resolution. Two potential error sources are suggested for such a retrieval approach. The first potential error to consider is brightness temperature measurement errors in a cost function, which indirectly affects the parameters retrieved through iterations. The second potential error source is land misclassification, which may cause an incorrect selection of retrieval models.

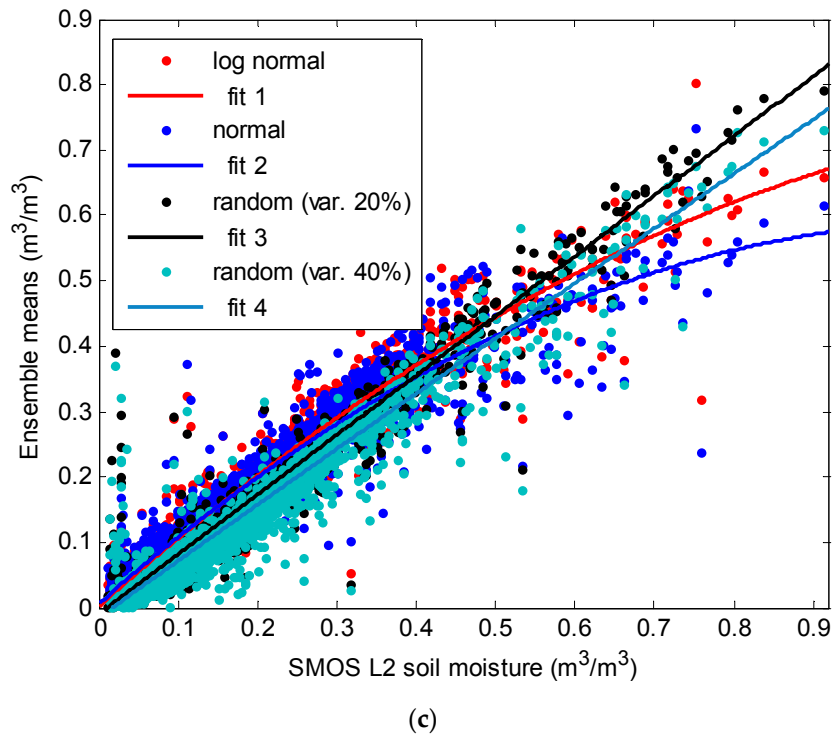


(a)



(b)

Figure 1. Cont.



**Figure 1.** Different rescaling curves: (a) by size (at the same variance of 40%, same normal distribution); (b) by variance (at the same ensemble size of 24, the same normal distribution); (c) by error distribution (at the same ensemble size of 12).

## 2.2. Retrieval Ensembles

### 2.2.1. Theoretical Rationale

The ensemble approach is based upon a statistical Monte Carlo Method (MCM). When a MCM is applied to the measurements, the ensemble mean indicates the actual measurement, and the ensemble variance represents the measurement errors [29]. Similarly, the SMOS soil moisture retrievals are considered as the first-guess observations containing errors. For SMOS observational errors, the SMOS soil moisture retrieval ensembles are propagated from a random perturbation of the SMOS Level 1c (L1c) brightness temperature products. In more detail, the SMOS soil moisture ensembles allowing for brightness temperature errors are generated as follows:

$$X = R(x^t + \varepsilon) \tag{1}$$

where  $X$  is the SMOS soil moisture ensemble,  $R$  is the SMOS soil moisture retrieval processor (e.g., Level 2 Prototype Processor, L2PP),  $x^t$  is the SMOS brightness temperature L1c product, and  $\varepsilon$  is the brightness temperature error from a random perturbation (*i.e.*, the square root of variance of random number). The SMOS soil moisture ensemble mean is obtained by multiple realizations of a retrieval algorithm with stochastic brightness temperature ensembles that have a mean of brightness temperature measurements as follows:

$$x = \frac{1}{m} \sum_{i=1}^{i=m} X_i \tag{2}$$

where  $x$  is the ensemble mean,  $m$  is the ensemble size, and  $X_i$  is the  $i$ th ensemble member of SMOS soil moisture retrievals. The retrieval ensemble mean  $x$  in Equation (2) is hypothetically considered as the best estimate free from retrieval errors, based upon the assumption of MCM [16,29].

For the generation of SMOS soil moisture retrieval ensembles, brightness temperature was selected as the main error source  $x^t$  of SMOS soil moisture retrievals in Equation (1). The rationale is based upon Lee *et al.* [14], who suggested that retrieval ensembles vary depending on the error attribution, although the ensemble size remains the same. They previously found that the soil moisture ensemble mean generated by a brightness temperature perturbation scheme provided the best estimate, after examining different ensemble means obtained from various perturbation schemes (e.g., vegetation, landscape misclassification, and soil roughness) of SMOS retrievals. That was based upon the observations that the brightness temperature ensembles appropriately decreased their spread (*i.e.*, uncertainty) with respect to the improved parts of the newly updated SMOS retrieval algorithm. In other words, when replacing a model with one performing better over dry soils, the ensemble spread decreased over dry soils, unlike other perturbation schemes. In addition, they showed that the spread and variance of ensembles specifically from a brightness temperature perturbation scheme were the highest when compared to the other perturbation schemes of geophysical parameters. It was because the L-band is less sensitive to geophysical parameters such as soil roughness or vegetation when compared to active microwave sensors at higher frequencies [30].

### 2.2.2. Experimental Setup for Ensemble Generation

We examined the SMOS soil moisture ensembles during Day of Year (DoY) 186–188 in 2010. For R in Equation (1), the L2PP (ver. 5.51) retrieving soil moisture from brightness temperature was used. For brightness temperature measurement inputs, Level 1C products (L1OP processor ver. 5.04, at a linearly interpolated incidence angle of 42.5 degrees, ascending mode) were used. For providing an error term of  $\varepsilon$  in Equation (1), various error distributions and perturbation schemes were explored. Specifically, the error distributions examined include normal, random, and lognormal functions with different variances at 10%, 20%, 30%, 40%, and 60% and various ensemble sizes (*i.e.*,  $m$  in Equation (2)) at 12, 24, 50, and 100. Such ranges of variances were chosen for including the SMOS brightness temperature uncertainties at 20–40 K shown in the literature [1,14,31–33]. This magnitude of brightness temperature variances takes into account the effects of calibration errors, a high vertical gradient condition of soil moisture or the water film arising from WAM rainfall events over West Africa, RFI, soil dryness, soil texture, and vegetation attenuation. Ensemble size was referred to in previous studies [2,8,14].

For producing rescaling functions, the different ensemble means generated from different perturbation schemes stated above corresponded to the original SMOS soil moisture end products. The fitting curves followed the second-order polynomial fitting because it showed high goodness-of-fit statistics above 0.9 as compared to other types of fit tested (e.g., Gaussian, exponential, power, rational, and sin functions).

### 2.3. CDF Matching

CDF matching [7] was used as a comparative run. A year of ERA interim reanalysis data (soil volumetric water layer 1, 0–7 cm) was employed for the reference climatology because it is widely used by soil moisture communities [34,35]. The percentiles of 5, 10, 30, 50, 70, 90, 95, and 100 were identified to define eight segments to assume a linear fit between the satellite and reference data [36].

## 3. Results and Discussion

### 3.1. Rescaling Functions

Figure 1 shows several ensemble groups altered by ensemble size, variance, and error distributions. Each data point has a one-to-one correspondence over the entire study domain between the retrieval ensemble mean and original SMOS end product. The number of pixels was approximately 4500 with no time-integration. Figure 1a shows that there is not much change in the final estimation of the ensemble mean, while the ensemble size dramatically changes from 12 to 100.



This implies that ensemble size is unlikely to be a control factor in retrieval ensemble generation. Instead of ensemble size, other factors such as error variance (Figure 1b) and error distribution (Figure 1c) showed a higher sensitivity. From these results and Lee *et al.* [14] who examined various error sources as discussed in Section 2.2.1, it was found that it is possible to modulate the ensemble mean at a reduced ensemble size in a computationally efficient manner.

Three rescaling functions to convert the original SMOS soil moisture product to the retrieval ensemble mean were selected from 12 cases of Figure 1, and are shown in Figure 2a. The selection was to compare the effects of the ensemble size, the variance, and different error distribution. In more detail, each ensemble group was generated by a normal distribution with a variance of 40% and an ensemble size of 100 (indicated by “Ens. (1)” below), by a normal distribution with a variance of 60% and an ensemble size of 24 (indicated by “Ens. (2)” below), and by a random distribution with a variance of 40% and an ensemble size of 12 (indicated by “Ens. (3)” below), respectively. Three rescaling functions are shown below:

$$\text{Ens. (1) : Rescaled soil moisture} = -0.4913 \times \text{SM}^2 + 1.0654 \times \text{SM} + 0.0054 \quad (3)$$

$$\text{Ens. (2) : Rescaled soil moisture} = -0.5233 \times \text{SM}^2 + 1.0364 \times \text{SM} + 0.0114 \quad (4)$$

$$\text{Ens. (3) : Rescaled soil moisture} = -0.0172 \times \text{SM}^2 + 0.8640 \times \text{SM} - 0.0157 \quad (5)$$

where SM indicates the SMOS soil moisture end product in  $\text{m}^3/\text{m}^3$ . R-square values for assessing the goodness-of-fit statistics into rescaling functions suggested were 0.9407, 0.9162, and 0.9327 for Equations (3)–(5), respectively. As this rescaling is a function of SMOS soil moisture (see Figure 2a), the degree of rescaling varies by the soil wetness. In other words, the correction degree of SMOS overestimation is significant, while a modulation of dry soils is marginal. Such a flexible adjustment is required because the SMOS error structure is related to soil texture and dryness. For example, a recent version of SMOS L2PP (version 5.51) performs well for dry and sandy soils [27], suggesting low uncertainty in dry soils. In contrast, wet soils in vegetated sites are still exposed to vegetation attenuation errors. Thus, there is no need to excessively correct dry soils. However, this does not imply that retrieval ensembles have no capability to modulate the underestimation.

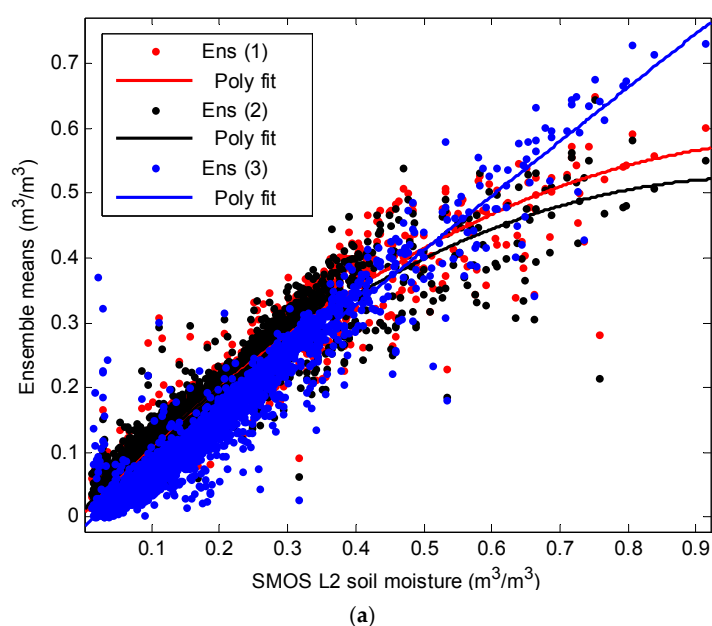
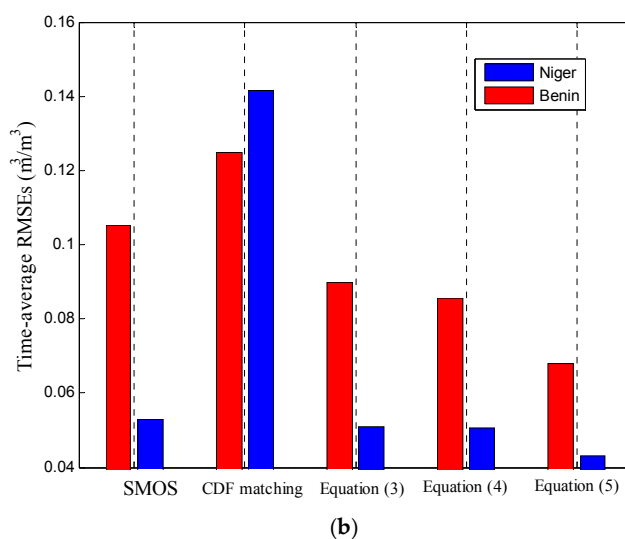


Figure 2. Cont.



**Figure 2.** (a) Rescaling functions (polynomial fit) of retrieval ensembles; (b) time-average RMSEs ( $\text{m}^3/\text{m}^3$ ) of the original SMOS data, CDF matching, Equations (3)–(5).

The SMOS overestimation is shown in Figure 2a, where SMOS soil moisture exceeded the hydrologically sensible levels of saturation points to exhibit spuriously high values up to  $0.9 \text{ m}^3/\text{m}^3$  on the x-axis. In Figure 2b, the time-average of RMSEs from DoPIY 150 to 240 (time range also shown in Figure 3a) in 2010 shows that such overestimation was mitigated by retrieval ensembles. As compared to Niger (dry and sandy bare soils), the SMOS overestimation was especially large in Benin (moderately wet and vegetated areas), as illustrated by the first column of Figure 2b. This is attributed to vegetation attenuation, considering the spatial distribution of the leaf area index (LAI) (see Supplementary Material Figure S1). At the second column of Figure 2b, CDF matching increased the RMSEs of SMOS data at both sites. However, all the ensemble rescalings at the third to fifth columns of Figure 2b appropriately diminished the original SMOS overestimation. In particular, it was noted that Equation (5) made a reasonable correction with a smaller ensemble size of 12 than Equation (3) at 100 and Equation (4) at 24, which implies that a large ensemble size is not a requirement for improving performance.

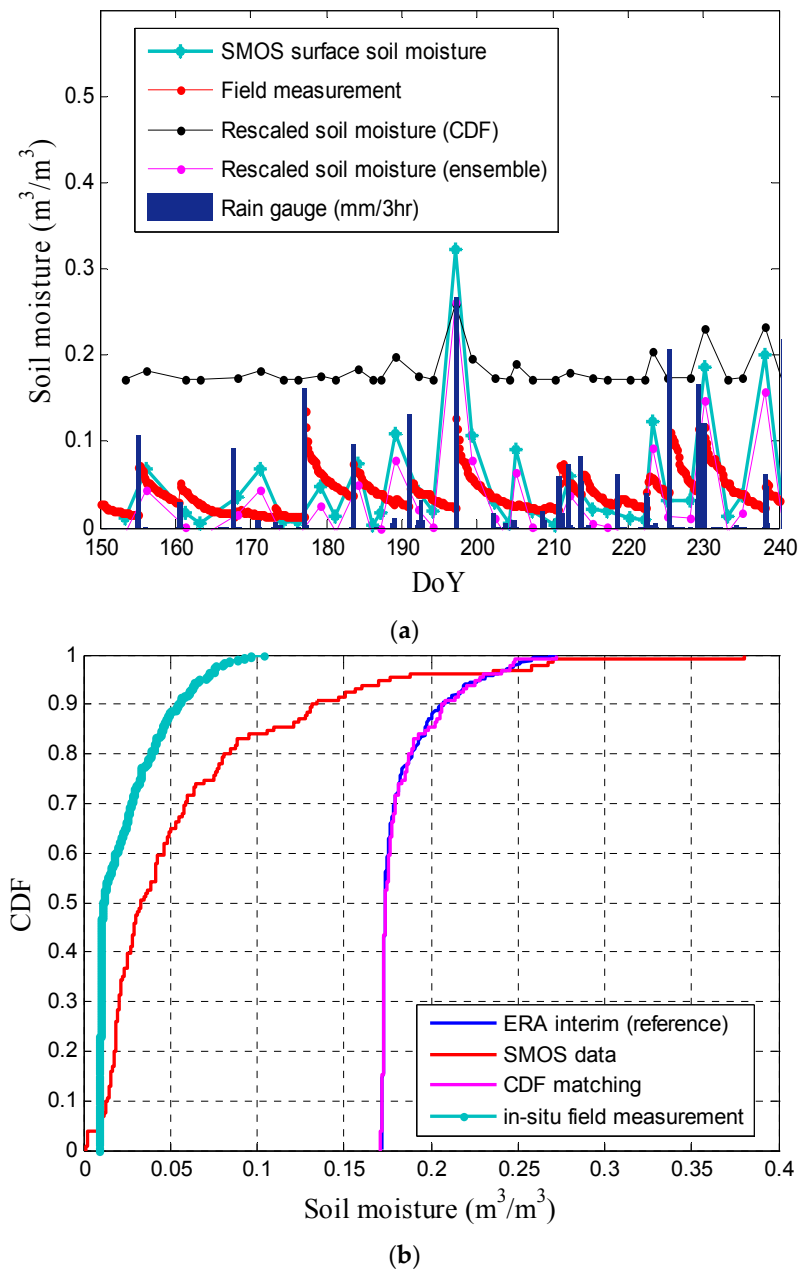
### 3.2. Point-Scale Validation: CDF Matching vs. Ensemble Rescaling

Figure 3 validates the ensemble approach at a point scale, as compared to field measurements and CDF matching results. Figure 3a demonstrates the instantaneous, non-linear, and dynamic error characteristics of SMOS soil moisture in Niger. Such errors are hardly considered by climatology as in CDF matching [12]. For example, in the event of rain on DoY 197 (rain gauge network data: 25 mm/3 h), SMOS soil moisture showed an abrupt and non-linear instantaneous spike. This is because it is difficult for SMOS signals to penetrate the water films formed by rainfall, resulting in non-linear and discontinuous errors on that specific day. Retrieval ensembles (specifically Equation (5)) mitigated such instantaneous SMOS errors arising from rain events in Niger.

In contrast, CDF matching overestimated soil moisture over dry soils in Niger throughout the experiment period. A level of overestimation exceeded SMOS soil moisture errors. This is attributed to errors in the reference data, as shown in Figure 3b by the cumulative distributions of reference data departed from those of *in-situ* field measurements in Niger. More specifically, CDF matching inappropriately shifted SMOS soil moisture towards the ERA interim reference data in an opposite direction to the field measurements, although a cumulative distribution of SMOS soil moisture was better matched with that of *in-situ* field measurements than that of the reference data. This problem is because CDF matching is, in all conditions, biased towards the reference data, but reference data is

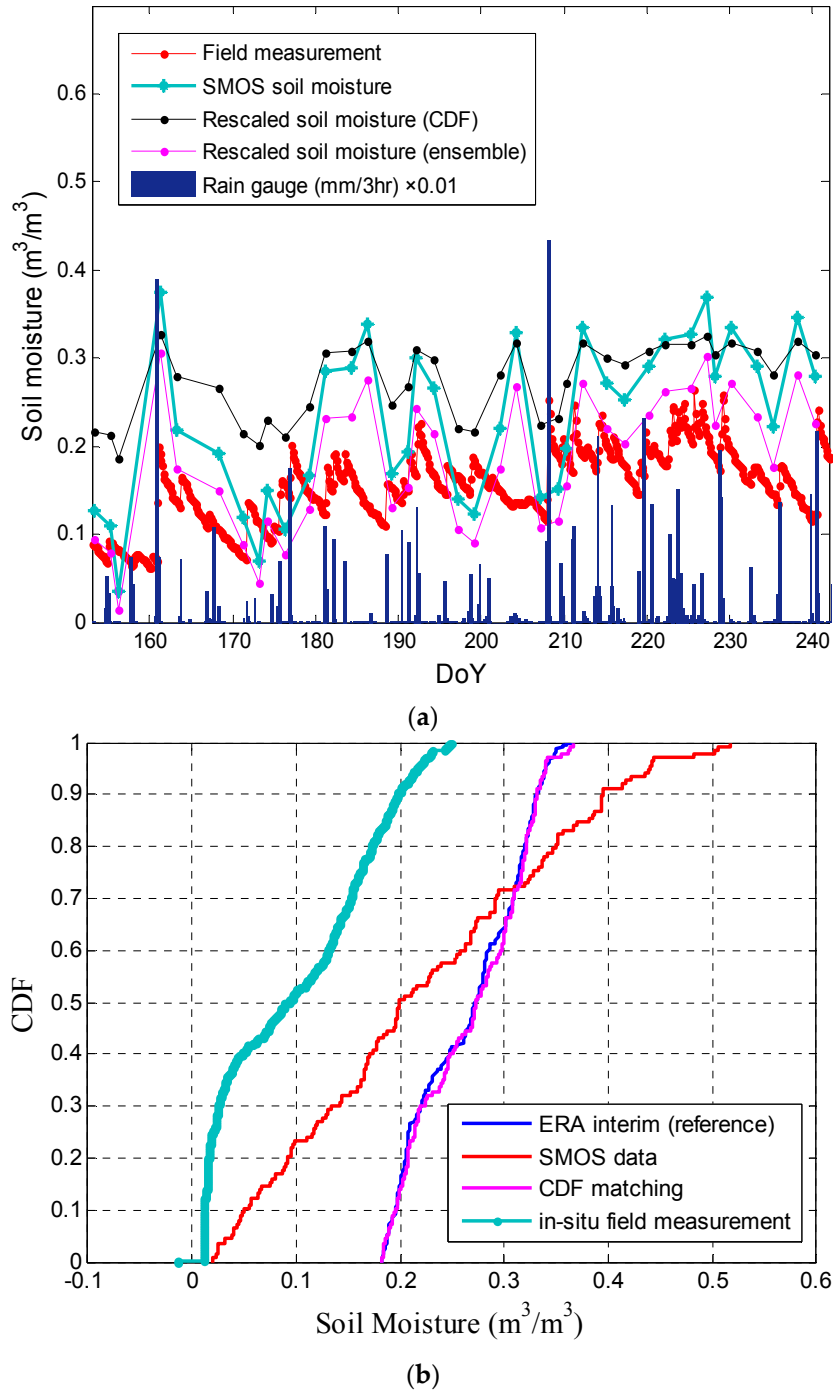


not always perfect. For example, the ERA interim used as reference data in CDF matching contains errors in soil moisture. This is because the lowest limit of soil moisture is set at wilting point when calculating bare soil evaporation [34]. However, soils in Niger (loamy sand) and other arid regions are, in fact, drier than the wilting point, as shown in Figure 3a. Thus, it is not possible to detect soils in extremely dry conditions. For this reason, the ERA interim data as reference data failed to appropriately estimate soil moisture in Niger, resulting in the failure of CDF matching. This limitation is not resolved by replacing the ERA interim data sets with other data sets including NOAA, Community Land Model(CLM), and Interaction Sol-Biosphère-Atmosphère (ISBA) model soil moisture or SMAP brightness temperature data sets, as they all contain their own intrinsic errors [14,23].



**Figure 3.** Soil moisture in Niger: (a) time-series comparisons; (b) cumulative distribution of reference data (ERA interim), SMOS data, CDF matching, and field measurements.

In Benin, retrieval ensembles successfully mitigated vegetation attenuation errors. As vegetation optical depth increases, brightness temperature decreases, and the sensitivity of brightness temperature to soil moisture variations also tends to decrease, resulting in SMOS overestimation and increase in soil moisture retrieval uncertainty [37,38]. In vegetated sites at Benin, as shown in Figure 4a, retrieval ensembles (*i.e.*, Equation (5)) significantly decreased the SMOS overestimations. There is no need to make a substitution or re-calibration of rescaling parameters as in CDF matching to take into account vegetation errors different from soil conditions in Niger. As shown in Figures 3a and 4a, the degree of rescaling varies by soil wetness.



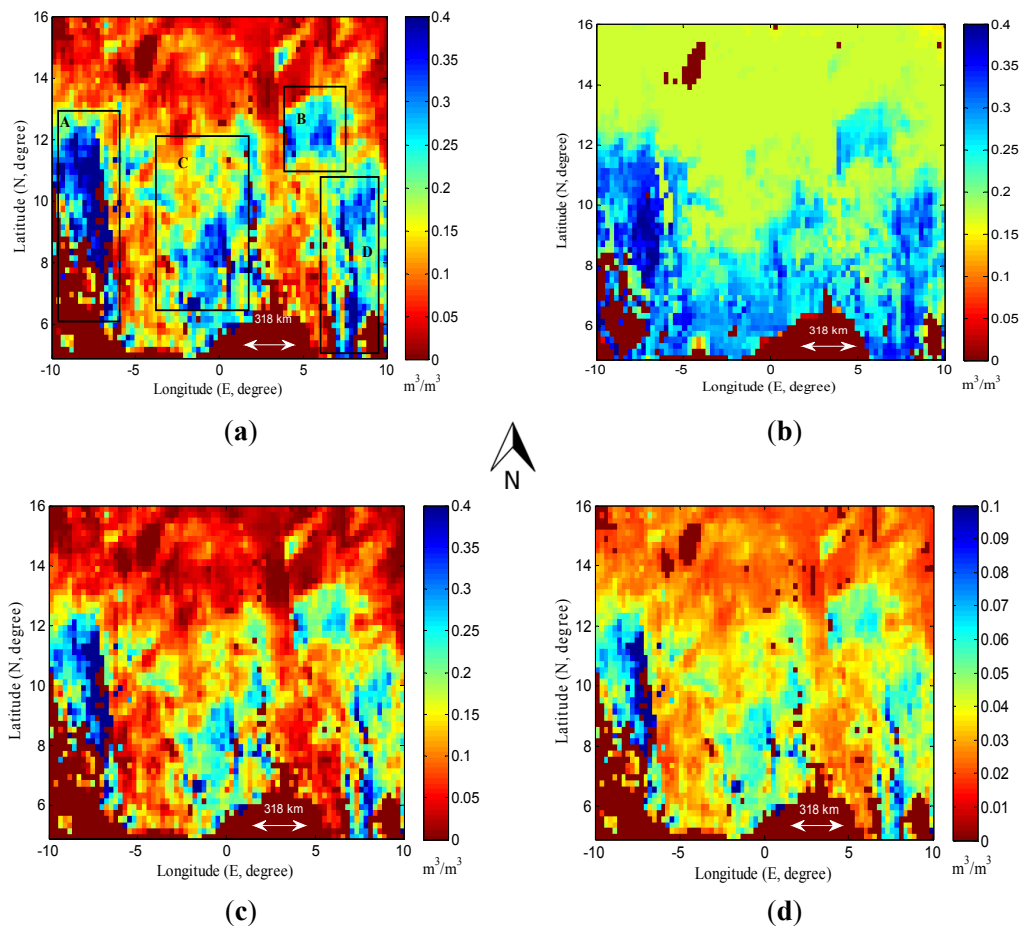
**Figure 4.** Soil moisture in Benin: (a) time-series comparisons; (b) cumulative distribution of reference data (ERA interim), SMOS data, CDF matching, and field measurements.

In contrast, CDF matching converged to the reference data instead of the field measurements. As shown in Figure 4b, a cumulative distribution of the reference data largely deviated from that of the field measurements. As a consequence, CDF matching transferred large errors inherent in the reference data to the final analysis. It was shown that a deviation of CDF matching from the field measurements was farther than that of the original SMOS soil moisture data, suggesting a risk of the approach matching with reference data.

### 3.3. Spatial Comparison: CDF Matching vs. Ensemble Rescaling

As discussed in Section 3.2, the SMOS retrieval errors are related to brightness temperature errors arising from rainfall and vegetation [14,31,33]. In this section, the correction effects for those errors are illustrated by spatial distributions.

The SMOS overestimations shown in red spikes in Figure 5a are related to two rain events (see [14] for Tropical Rainfall Measuring Mission (TRMM) rainfall), indicated by black boxes “A” and “B”, fallen on ECOCLIMAP clay soils, implying a reduced infiltration rate and wetter conditions as a consequence. In reduced infiltration soils, the water film formed by rainfall prevents the penetration of SMOS signals, resulting in brightness temperature errors and spuriously high soil moisture retrievals. In addition to rainfall effects, the SMOS overestimations are relevant to vegetation attenuations indicated by black boxes “C” and “D” (see Supplementary Materials Figure S1 for a latitudinal gradient of LAI distribution) in Figure 5a.



**Figure 5.** Spatial distribution of surface soil moisture,  $m^3/m^3$ : (a) SMOS (control run); (b) rescaled by CDF matching; (c) ensemble rescaling; (d) difference between (a) and (c).

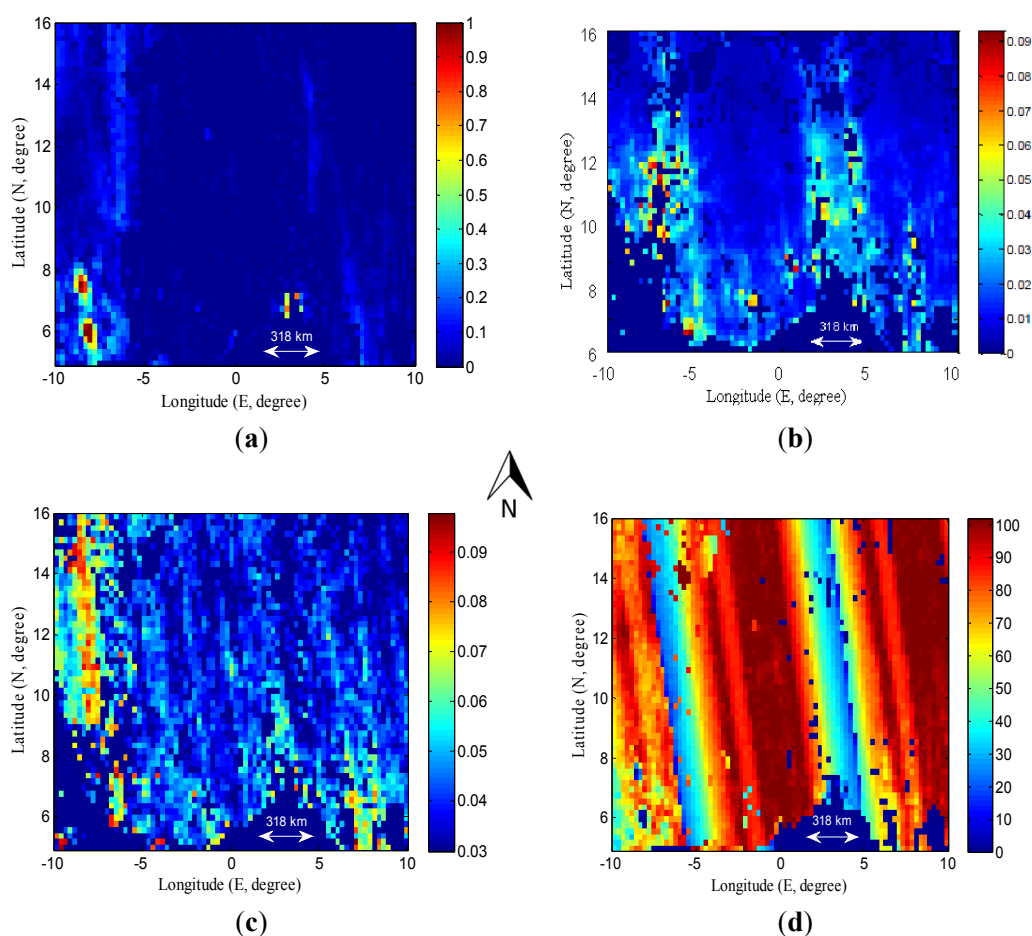
As compared to the original SMOS data in Figure 5a, CDF matching in Figure 5b furthermore overestimated soil moisture over the entire study domain, due to the intrinsic errors in the ERA interim reference data, as discussed above. In contrast, Figure 5c shows that an ensemble approach (*i.e.*, Equation (5)) effectively mitigated SMOS overestimations, appropriately representing dry soils. Those overestimation parts shown in black boxes “A” and “B” were alleviated in Figure 5c. Vegetated areas shown in black boxes “C” and “D” were also mitigated in Figure 5c. All the differences before and after ensemble rescalings are indicated by Figure 5d.

From a spatial distribution, it was noted that mitigation and modulation rates of the ensemble approach are spatially heterogeneous and instantaneous. This effect cannot be achieved by the replacement of retrieval models or a threshold approach for the following reasons. If brightness temperature measurements are contaminated by errors, a SMOS retrieval algorithm using a Bayesian approach retrieves faulty parameters to be matched with erroneous brightness temperature, although an ensemble approach assumes that a model is perfect. In addition, a threshold approach has a limitation because soils drier than a threshold level are hindered. Finally, for better understanding of a large-scale circulation, it is very important to appropriately present dry soils by satellite observations, as most of the climate models usually neglect a negative feedback between dry soils and convective rainfall [34,39,40].

#### 3.4. Spatial Comparison: QC Science Flag vs. Ensemble Rescaling

This section shows that the rescaling discussed in Sections 3.2 and 3.3 cannot be predicted with the QC science flag provided by the SMOS Data Analysis Product (DAP) and User Data Product (UDP) data set [5]. The QC science flag includes information about brightness temperature measurement data availability, RFI, radiance measurements, retrieved parameters, default *a priori* parameters, and auxiliary data products.

Some QC examples are provided in Figure 6 in order to compare them with the effects of ensemble rescaling in Figure 5d. Figure 6a indicates the RFI fraction by  $(RFI_x + RFI_y)/M\_AVA0$  where  $M\_AVA0$  is the initial number of measurements in L1c. The mitigation effects in Figure 5d are unlikely to overlap with Figure 6a, as the pixels highly contaminated by RFI did not undergo any retrieval [27]. Figure 6b shows the soil moisture Data Quality index (DQX) indicating the quality of retrievals. The DQX is a measure of *a posteriori* standard deviation of SMOS soil moisture retrievals. As the default value is set to zero, any positive number implies uncertainty in retrievals. As there is no overlap with the improvements shown in Figure 5d, it is unlikely that retrieval errors can be predicted from the DQX information. Figure 6c shows the Chi-square index, which is another retrieval fit quality index provided by SMOS DAP data sets. It informs us about a cost function obtained at the end of the retrieval process. The smaller the value, the better the fit between the SMOS brightness temperature measurements and the simulation. Some part of Figure 5d (shown by black box “A” in Figure 5a) was consistent with the Chi-square index. In Figure 6d, the Measurements AVAilable ( $M\_AVA$ ) indicate the number of Level 1c brightness temperature measurements available for the retrievals, after filtering out faulty or dubious pixels for several reasons. It does not appear related to ensemble rescaling in Figure 5c. Taken together, it is difficult to mitigate or predict SMOS retrieval errors solely with the QC science flag. Although there is some degree of consistency between the Chi-square index and some parts of the ensemble rescaling in Figure 5d, the Chi-square index did not completely detect all the effects of rainfall and vegetation, nor did it provide the absolute values of retrieval errors in the soil moisture unit of  $m^3/m^3$ .



**Figure 6.** Quality control flags: (a) RFI fraction  $((RFI_x + RFI_y)/M\_AVA0)$ ; (b) soil moisture DQX; (c) Chi-square; (d) M\_AVA.

#### 4. Discussion

As compared to one-dimensional field measurement-based error analysis, triple collocation, and CDF matching, the proposed approach suggests several operational merits. First, this ensemble approach mitigates non-linear IFOV errors not predicted by the QC science flag or climatology information. Secondly, this ensemble approach is self-contained, without relying on any reference data or relative comparison with other data sets. In other words, as the ensemble approach uses the error propagation of its own retrieval algorithm, the estimation of uncertainty does not change by other data sets. In contrast, the consistency check approach (*i.e.*, satellite observation–minus–model estimates) as in CDF matching or triple collocation is based upon comparative judgement so that the estimation of satellite errors changes if changing data sets to be compared with satellite data. If both satellite and model estimates concurrently make overestimations, for example in the event of rain, that approach would provide the misleading information that they are accurate. Thirdly, this ensemble approach provides error information without scaling issues. Since one-dimensional field measurement-based error estimation is not useful in practice due to the scarcity of field measurement data across the globe, a lack of spatial heterogeneity information, and upscaling errors [12,38,41,42], the pixel-scale IFOV error estimation from the retrieval ensemble spread adds useful information to the existing QC. Finally, the ensemble approach does not hinder extreme conditions, unlike a threshold approach.

Future studies will further improve limitations of the proposed approach such as how to reduce the SMOS biases that occur under the condition of high vertical gradients of soil moisture. As shown

in Figure 3a, a reduction rate of SMOS overestimation to be made in the event of rain (DoY 197) fallen on dry and sandy soils was not very remarkable, as much as it is needed. To this end, there are several things to consider. In the present study, only brightness temperature errors were considered to generate the SMOS ensemble mean. This is based upon the findings of Lee *et al.* [14] on the statistics of the retrieval ensemble spread. However, if other SMOS error sources such as land cover fraction errors affecting a selection of retrieval models are included for perturbations, this may make a further improvement. In fact, this is a matter of trade-off. If we add more and more error sources and complicated perturbation schemes, the computational cost will increase accordingly. It may make it difficult to apply this ensemble approach to near-real-time operations due to computational cost. Secondly, the ensemble mean may not be able to be considered for retrieval bias correction if the error distribution is not Gaussian or random. The error distribution may be difficult to define in other regions. Thirdly, as the present study produced a rescaling function from ensembles over West Africa, Equation (5) may not be applicable to other regions. In addition, a retrieval algorithm always evolves. Thus, for a future study, it may be useful to explore a short time-integration of retrieval ensembles over a single pixel, rather than directly adopting Equation (5). Finally, this ensemble approach which utilizes a retrieval evolution cannot mitigate errors of instrument measurements such as brightness temperature or backscattering, although this can mitigate soil moisture retrieval errors at end-product levels.

## 5. Conclusions

We illustrated that it is possible to mitigate SMOS overestimation with retrieval ensembles in a computationally inexpensive way (*i.e.*, ensemble size of 12 and no time-integration). We overcame the existing perception that an ensemble approach is not operationally feasible due to high computational cost [38]. This is ground-breaking because it was shown that an ensemble approach can be used for near-real-time operations if an ensemble generation process is appropriately optimized. Ensemble-based satellite bias correction is more appropriate for assessing the RMSE goal of SMOS and SMAP missions, as they are not climatology as in CDF matching or triple collocation. Such an ensemble approach also provides important observational error information for data assimilation analysis.

Validations at Niger and Benin sites showed that the time-average RMSEs significantly decreased with an ensemble approach, while they increased with the CDF matching being biased towards the reference data in all conditions. This is because CDF matching is biased to the imperfect reference data contaminated by several errors. The spatial comparison results showed that the ensemble rescaling mitigated SMOS overestimation arising from vegetation attenuation and rain events. Those improvements are not predicted from the QC science flag.

**Acknowledgments:** This research was supported by National Space Lab Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT, & Future Planning (Grant: NRF-2013M1A3A3A02042391 and NRF-2015R1C1A1A02037224). We also thank the “Centre Aval de Traitement des Données SMOS” (CATDS) operated for the “Centre National d’Etudes Spatiales” (CNES, France) by IFREMER (Brest, France) for the SMOS data. The AMMA data set was obtained within the framework of the AMMA project. Based on a French initiative, AMMA has been constructed by an international group and is currently funded by a large number of agencies, especially from France, the UK, the US and Africa. It has been the beneficiary of a major financial contribution from the European Community’s Sixth Framework Research Programme. Detailed information on the scientific coordination and funding is available on the AMMA international web site (<https://www.amma-eu.org>).

**Author Contributions:** Ju Hyoung Lee designed this study, and led the data analysis and manuscript writing. Jung-ho Im contributed to the research design, the discussion of the results, the manuscript writing, and served as the corresponding author.

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



## References

1. De Rosnay, P.; Muñoz Sabater, J.; Drusch, M.; Albergel, C.; Balsamo, G.; Boussetta, S.; Isaksen, L.; Thépaut, J.-N. Bias correction for SMOS data assimilation in the ECMWF Numerical Weather Prediction System. In Proceedings of the ESA Living Planet Symposium, Edinburgh, UK, 9–13 September 2013.
2. Margulis, S.A.; McLaughlin, D.; Entekhabi, D.; Dunne, S. Land data assimilation and estimation of soil moisture using measurements from the Southern Great Plains 1997 Field Experiment. *Water Resour. Res.* **2002**, *38*. [[CrossRef](#)]
3. Anderson, J.L. Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter. *Phys. D: Nonlinear Phenom.* **2007**, *230*, 99–111. [[CrossRef](#)]
4. Gruhier, C.; de Rosnay, P.; Kerr, Y.; Mougin, E.; Ceschia, E.; Calvet, J.-C.; Richaume, P. Evaluation of AMSR-E soil moisture product based on ground measurements over temperate and semi-arid regions. *Geophys. Res. Lett.* **2008**, *35*, L10405. [[CrossRef](#)]
5. Reichle, R.H. Data assimilation methods in the Earth sciences. *Adv. Water Resour.* **2008**, *31*, 1411–1418. [[CrossRef](#)]
6. Fertig, E.J.; Baek, S.-J.; Hunt, B.R.; Ott, E.; Szunyogh, I.; Aravéquia, J.A.; Kalnay, E.; Li, H.; Liu, J. Observation bias correction with an ensemble Kalman filter. *Tellus A* **2009**, *61*, 210–226. [[CrossRef](#)]
7. Reichle, R.H.; Koster, R.D. Bias reduction in short records of satellite soil moisture. *Geophys. Res. Lett.* **2004**, *31*, L19501. [[CrossRef](#)]
8. Entekhabi, D.; Yueh, S.; O'Neill, P.; Kellogg, K.; Allen, A.; Bindlish, R.; Brown, M.; Chan, S.; Colliander, A.; Crow, W.T.; *et al.* *SMAP Handbook*; JPL Publication JPL 400–1567; Jet Propulsion Laboratory: Pasadena, CA, USA, 2014; p. 182.
9. Dee, D.P. Bias and data assimilation. *Q. J. R. Meteorol. Soc.* **2005**, *131*, 3323–3343. [[CrossRef](#)]
10. Dee, D.P.; Uppala, S. Variational bias correction of satellite radiance data in the ERA-Interim reanalysis. *Q. J. R. Meteorol. Soc.* **2009**, *135*, 1830–1841. [[CrossRef](#)]
11. Draper, C.S.; Mahfouf, J.-F.; Walker, J.P. An EKF assimilation of AMSR-E soil moisture into the ISBA land surface scheme. *J. Geophys. Res.* **2009**, *114*, D20104. [[CrossRef](#)]
12. Crow, W.T.; Berg, A.A.; Cosh, M.H.; Loew, A.; Mohanty, B.P.; Panciera, R.; deRosnay, P.; Ryu, D.; Walker, J.P. Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products. *Rev. Geophys.* **2012**, *50*, RG2002.
13. Reichle, R.H.; Koster, R.D.; Liu, P.; Mahanama, S.P.; Njoku, E.G.; Owe, M. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). *J. Geophys. Res.* **2007**, *112*, D09108. [[CrossRef](#)]
14. Lee, J.H.; Pellarin, T.; Kerr, Y.H. EnOI optimization of SMOS soil moisture over West Africa. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2015**, *8*, 1821–1829. [[CrossRef](#)]
15. Drusch, M.; Wood, E.F.; Gao, A.H. Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. *Geophys. Res. Lett.* **2005**, *32*, L15403. [[CrossRef](#)]
16. Evensen, G. The Ensemble Kalman Filter: Theoretical formulation, and practical implementation. *Ocean Dyn.* **2003**, *53*, 343–367. [[CrossRef](#)]
17. Stroeve, J.; Hamilton, L.C.; Bitz, C.M.; Blanchard-Wrigglesworth, E. Predicting September sea ice: Ensemble skill of the SEARCH Sea Ice Outlook 2008–2013. *Geophys. Res. Lett.* **2014**, *41*, 2411–2418. [[CrossRef](#)]
18. Koster, R.D.; Cox, P.M.; Dirmeyer, P.; Guo, Z.; Bonan, G.; Chan, E. Regions of strong coupling between soil moisture and precipitation. *Science* **2004**, *305*, 1138–1140. [[CrossRef](#)] [[PubMed](#)]
19. Zhang, J.; Li, Z.L.; Li, J.; Li, J.L. Ensemble retrieval of atmospheric temperature profiles from AIRS. *Adv. Atmos. Sci.* **2014**, *31*, 559–569. [[CrossRef](#)]
20. Hossain, F.; Anagnostou, E.N. A two-dimensional satellite rainfall error model. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1511–1522. [[CrossRef](#)]
21. Olson, W.S.; Kummerow, C.D.; Yang, S.; Petty, G.W.; Tao, W.; Bell, T.L.; Braun, S.A.; Wang, Y.; Lang, S.E.; Johnson, D.E.; *et al.* Precipitation and latent heating distributions from satellite passive microwave radiometry. Part I: Improved method and uncertainties. *J. Appl. Meteor. Climatol.* **2006**, *45*, 702–720. [[CrossRef](#)]

22. Balsamo, G.; Albergel, C.; Beljaars, A.; Boussetta, S.; Brun, E.; Cloke, H.; Dee, D.; Dutra, E.; Muñoz-Sabater, J.; Pappenberger, F.; *et al.* ERA-Interim/Land: A global land surface reanalysis data set. *Hydrol. Earth Syst. Sci.* **2015**, *19*, 389–407. [[CrossRef](#)]
23. Lee, J.H.; Pellarin, T.; Kerr, Y.H. Inversion of soil hydraulic properties from the DEnKF analysis of SMOS soil moisture over West Africa. *Agric. For. Meteorol.* **2014**, *188*, 76–88. [[CrossRef](#)]
24. Delon, C.; Galy-Lacaux, C.; Adon, M.; Lioussé, C.; Serça, D.; Diop, B.; Akpo, A. Nitrogen compounds emission and deposition in West African ecosystems: Comparison between wet and dry savanna. *Biogeosciences* **2012**, *9*, 385–402. [[CrossRef](#)]
25. Cappelaere, B.; Descroix, L.; Lebel, T.; Boun, N.; Ramier, D.; Laurent, J.-P.; Favreau, G.; Boubkraoui, S.; Boucher, M.; Moussa, I.B.; *et al.* The AMMA-CATCH experiment in the cultivated Sahelian area of south-west Niger—Investigating water cycle response to a fluctuating climate and changing environment. *J. Hydrol.* **2009**, *375*, 34–51. [[CrossRef](#)]
26. Séguis, L.; Boulain, N.; Cappelaere, B.; Cohard, J.M.; Favreau, G.; Galle, S.; Guyot, A.; Hiernaux, P.; Mougin, É.; Peugeot, C.; *et al.* Contrasted land-surface processes along the West African rainfall gradient. *Atmos. Sci. Lett.* **2011**, *12*, 31–37. [[CrossRef](#)]
27. Kerr, Y.H.; Waldteufel, P.; Richaume, P.; Ferrazzoli, P.; Wigneron, J.-P. *SMOS LEVEL 2 Processor Soil Moisture Algorithm Theoretical Basis Document (ATBD) v1.3h*; SM-ESL (CBSA): Toulouse, France, 2013; p. 141.
28. Kerr, Y.H.; Waldteufel, P.; Richaume, P.; Wigneron, J.-P.; Ferrazzoli, P.; Mahmoodi, A.; Al Bitar, A.; Cabot, F.; Gruhier, C.; Juglea, S.E.; *et al.* The SMOS Soil Moisture Retrieval Algorithm. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1384–1403. [[CrossRef](#)]
29. Burgers, G.; Jan van Leeuwen, P.; Evensen, G. Analysis scheme in the Ensemble Kalman Filter. *Mon. Weather Rev.* **1998**, *126*, 1719–1724. [[CrossRef](#)]
30. Rodriguez-Fernandez, N.; Richaume, P.; Aires, F.; Prigent, C.; Kerr, Y.; Kolassa, J.; Jimenez, C.; Cabot, F.; Mahmoodi, A. Soil moisture retrieval from SMOS observations using neural networks. In Proceedings of the 2014 IEEE International Geoscience Remote Sensing Symposium, Quebec City, QC, Canada, 13–18 July 2014; pp. 2431–2434.
31. Holmes, T.R.H.; Drusch, M.; Wigneron, J.P.; de Jeu, R.A.M. A global simulation of microwave emission: Error structures based on output from ECMWF's operational integrated forecast system. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 846–856. [[CrossRef](#)]
32. Muñoz-Sabater, J. Incorporation of passive microwave Brightness Temperatures in the ECMWF soil moisture analysis. *Remote Sens.* **2015**, *7*, 5758–5784. [[CrossRef](#)]
33. Schlenz, F.; dall'Amico, J.T.; Mauser, W.; Loew, A. Analysis of SMOS brightness temperature and vegetation optical depth data with coupled land surface and radiative transfer models in Southern Germany. *Hydrol. Earth Syst. Sci.* **2012**, *16*, 3517–3533. [[CrossRef](#)]
34. Balsamo, G.; Viterbo, P.; Beljaars, A.; van den Hurk, B.; Hirschi, M.; Betts, A.K.; Scipal, K. A revised hydrology for the ECMWF model: Verification from field site to terrestrial water storage and impact in the Integrated Forecast System. *J. Hydrometeor.* **2009**, *10*, 623–643. [[CrossRef](#)]
35. Scipal, K.; Drusch, M.; Wagner, W. Assimilation of a ERS scatterometer derived soil moisture index in the ECMWF numerical weather prediction system. *Adv. Water Resour.* **2008**, *31*, 1101–1112. [[CrossRef](#)]
36. Chung, D.; Dorigo, W.; Hahn, S.; Melzer, T.; Paulik, C.; Reimer, C.; Vreugdenhil, M.; Wagner, W.; Kidd, R. *ECV Production, Fusion of Soil Moisture Products: Algorithm Theoretical Baseline Document, Ver. 2.0*. ESA soil moisture CCI; ESA: Vienna, Austria, 2014.
37. De Jeu, R.A.M.; Wagner, W.W.; Holmes, T.R.H.; Dolman, A.J.; van de Giesen, N.C.; Friesen, J. Global soil moisture patterns observed by space borne microwave radiometers and scatterometers. *Surv. Geophys.* **2008**, *28*, 399–420. [[CrossRef](#)]
38. Parinussa, R.M.; Meesters, A.G.C.A.; Liu, Y.Y.; Dorigo, W.; Wagner, W.; de Jeu, R.A.M. Error estimates for near-real-time satellite soil moisture as derived from the Land Parameter Retrieval Model. *IEEE Geosci. Remote Sens. Lett.* **2011**, *8*, 779–783. [[CrossRef](#)]
39. Guillod, B.P.; Orłowsky, B.; Miralles, D.G.; Teuling, A.J.; Seneviratne, S.I. Reconciling spatial and temporal soil moisture effects on afternoon rainfall. *Nat. Commun.* **2015**. [[CrossRef](#)] [[PubMed](#)]
40. Taylor, C.M.; de Jeu, R.A.M.; Guichard, F.; Harris, P.P.; Dorigo, W.A. Afternoon rain more likely over drier soils. *Nature* **2012**, *489*, 423–426. [[CrossRef](#)] [[PubMed](#)]

41. Famiglietti, J.S.; Ryu, D.; Berg, A.A.; Rodell, M.; Jackson, T.J. Field observations of soil moisture variability across scales. *Water Resour. Res.* **2008**, *44*, W01423. [[CrossRef](#)]
42. Lu, H.; Yang, K.; Koike, T.; Zhao, L.; Qin, J. An improvement of the radiative transfer model component of a land data assimilation system and its validation on different land characteristics. *Remote Sens.* **2015**, *7*, 6358–6379. [[CrossRef](#)]



© 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 by Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).