*Article*

# Vegetation Characterization through the Use of Precipitation-Affected SAR Signals

**Ramses A. Molijn** [1,*] , **Lorenzo Iannini** [1], **Paco López Dekker** [1], **Paulo S.G. Magalhães** [2] and **Ramon F. Hanssen** [1]

1   Geoscience and Remote Sensing, Delft University of Technology, 2628 CN Delft, The Netherlands; L.Iannini@tudelft.nl (L.I.); F.LopezDekker@tudelft.nl (P.L.D.); R.F.Hanssen@tudelft.nl (R.F.H.)
2   Faculdade de Engenharia Agrícola (FEAGRI), Unicamp, Campinas 13083-875, Brazil; graziano@g.unicamp.br
*   Correspondence: R.A.Molijn@tudelft.nl

check for updates

**Abstract:** Current space-based SAR offers unique opportunities to classify vegetation types and to monitor vegetation growth due to its frequent acquisitions and its sensitivity to vegetation geometry. However, SAR signals also experience frequent temporal fluctuations caused by precipitation events, complicating the mapping and monitoring of vegetation. In this paper, we show that the influence of a priori known precipitation events on the signals can be used advantageously for the classification of vegetation conditions. For this, we exploit the change in Sentinel-1 backscatter response between consecutive acquisitions under varying wetness conditions, which we show is dependent on the state of vegetation. The performance further improves when a priori information on the soil type is taken into account.

## 1. Introduction

Land cover classification offers baseline information for several land monitoring applications such as food security monitoring and early warning, water use efficiency studies or fraud detection, to cite a few. Over the last decade, the spatial and temporal resolution of land cover monitoring products underwent significant enhancements, mainly due to advances in the provision of Analysis-Ready Data (ARD) from existing satellite platforms and due to the increasing computational resources for processing large amounts of remote sensing data. Global land cover maps such as Global Food Security-Support Analysis Data at 30-m (GFSAD30) and GlobeLand30 at 30-m resolution [1–3] and Global Land Cover Project 2000 (GLC 2000), MODIS and GlobCover at 300-m to one-kilometer resolution [4–7] are available and are mainly based on optical imagery produced by the Landsat, SPOT, MODIS and MERIS missions. These land cover products commonly feature broad land cover classes and generally do not distinguish between crop types.

Although the information provided by these products can be valuable for generating regional overviews, important shortcomings were identified, as well. A study comparing the maps of GLC 2000, MODIS and GlobCover produced until 2011 [8] focused specifically on the thematic accuracies between these products. It was reported that the overall accuracies of these maps are 0.68, 0.75 and 0.67, respectively, and even higher disagreements are found for cropland and forest classes. This is critical for studies where precise spatial and temporal discrimination of vegetation classes is required. In the same study, it was stressed that the errors in the maps were higher than the detected changes, making them ineffective for land cover change detection. Although maps produced since 2011 show accuracy improvements, these were found not to be significant given the advances in satellite data

acquisition and algorithms [9]. In addition, all of the mentioned land cover products provide the land cover type, but usually lack phenological information.

The use of Synthetic Aperture Radar (SAR) for land cover classification, and in particular crop classification, is only successful when time series are taken into account [10]. Since the backscatter in a single SAR image can be practically identical for different crops, the focus must lie on the temporal backscatter differences, when vegetation structure changes, especially during the seed and development phase [10]. Recognizing that different wavelengths feature different interactions with vegetation and hence give different classification performances, for C-band SAR specifically, it was found that saturation effects, occurring generally from a vegetation height of 1–1.5 m [10–12], cause classification to be most accurate for lower biomass crops (e.g., wheat, hay, pasture) [10]. Generally, in these studies, it is mentioned that signals in cross-polarization are favored over signals in co-polarization for effective crop classification, since these are more sensitive to volume scattering, as well as to soil moisture [13].

Even though the introduction of Sentinel-1, providing free and temporally-dense C-band SAR imagery, yields a unique opportunity for incorporating these signals into the generation process of land cover monitoring products, surprisingly few works were found that promote such efforts along with descriptions on the SAR-related methodologies and consequential improvements. Limiting to publications that describe the utilization of temporal Sentinel-1 backscatter (intensity) signals for producing land cover products, the studies by [14–16] are illustrative, all applying machine learning algorithms (random forest). None of the works mention the incorporation of contextual information with daily precipitation events and incidence angle. Presumably, the applied models translate the corresponding signal variations into uncertainty in the classification results. An exception is the study by [16], which found that the effect of the incidence angle on the resulting classification performance was minimal. A work that explicitly mentioned the removal of precipitation-affected SAR signals was the study by [17], which evaluated the radiometric stability and quality of Sentinel-1. In addition, the studies by [13,18] emphasized the importance of incorporating daily irrigation and precipitation information when monitoring crop fields using SAR signals, and the latter showed that the signals are most sensitive to soil moisture when taken over bare ground as compared to full vegetation cover. However, none of the above-mentioned works proposed methods on how to treat precipitation-affected signals for improved vegetation classification.

Based on these literature findings, we underline the importance of incorporating precipitation information when performing temporal land cover classification, and in particular vegetation classification, based on SAR signals. In this paper, we focus on the variations in SAR signals that arise from precipitation events and how these could be used for advanced discrimination of vegetation conditions. The objective is to manifest the effect of these variations on the classification performance and the improvements that can be achieved when taking into account precipitation and backscatter information with temporally-consecutive SAR observations. In addition, we quantify the effect on the classification performance when taking into account information with SAR incidence angle and soil type. The aim is to propose an approach on how to handle precipitation-affected SAR signals and the advantages of incorporating such information, along with available knowledge on incidence angle and soil type, regardless of the SAR-based land cover monitoring or classification model. As such, it is not intended that the presented methodologies be used as stand-alone techniques for SAR-based vegetation characterization.

## 2. Study Area and Data Description

### 2.1. Study Area

The study area is São Paulo state, which covers nearly 250,000 $km^2$, 3% of Brazil. It is inhabited by 46 million people, 20% of Brazil's population, most of them living in the eastern part close to São Paulo city. The state territory is located in the sub-tropics and covers various climatic types, mainly

influenced by extratropical and tropical weather systems. The average monthly temperature generally reaches close to 30 °C during summer and close to 20 °C during winter. Most of the precipitation occurs during the rainy season in the summer months; from November until January, almost 40% of the annual precipitation falls. The winter months are usually dry, occasionally with no precipitation during several consecutive weeks. The state's interior receives 1000–1600 mm annual rainfall, whereas the forest strip along the coast receives more than 3000 mm annual rainfall. When precipitation occurs, the average of daily cumulative rainfall is slightly more than 10 mm. The majority of the state is elevated between 300 m and 900 m above sea level.

The forest strip along the coast consists of protected native (mainly Atlantic) forest and divides the coastline areas from the state's interior. The interior is covered by a mix of sugarcane, annual crops including mainly corn, soybean and bean, as well as citrus trees and forest crops including mainly eucalyptus and pine. The state also hosts numerous watersheds and three major river basins. An indication of the coverage of the land cover types is provided in Table 1, and their spatial distribution is illustrated by Figure 1. For more details on the origin of this dataset, see Section 2.2.3. Mid-vegetation includes mainly shrubland and citrus; grassland includes mainly pasture; and forest includes high bushes, native forests and production forests. There are two dominant soil types in São Paulo state, which are different mainly in clay content, as will be further described in Section 2.2.4.

**Table 1.** Coverage of land cover types in São Paulo state based on the 2015 classification map produced by Delft University of Technology (TU Delft) [19,20].

| Land Cover | Coverage |
|---|---|
| Sugarcane | 19% |
| Annual crop | 4% |
| Grassland | 33% |
| Mid-vegetation | 15% |
| Native forest | 23% |
| Production forest | 1% |
| Water | 3% |
| Built-up | 2% |
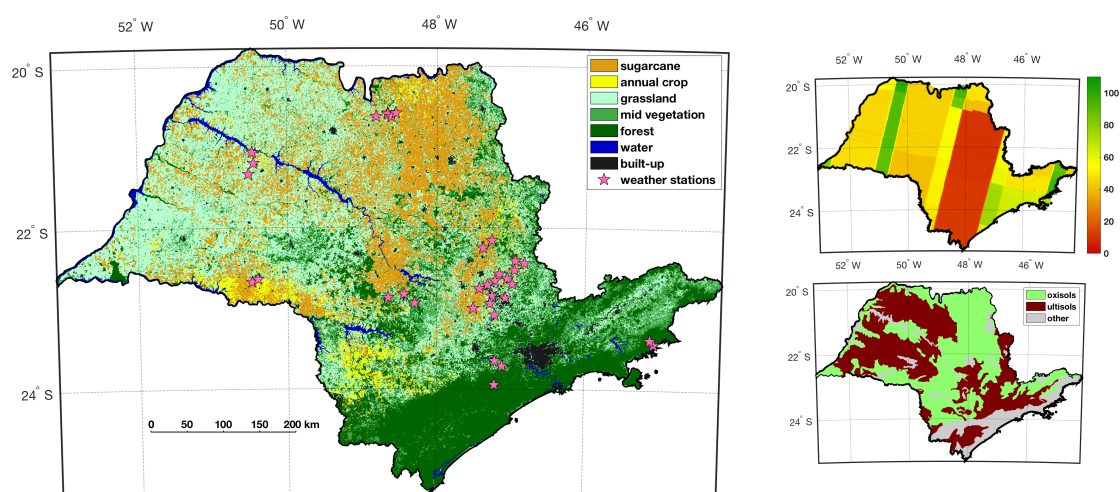


**Figure 1.** São Paulo state land cover map for 2015 based on Landsat data, produced by TU Delft [19,20] (left), the number of Sentinel-1 Interferometric-Wide acquisitions (top right) and the location of the two soil types (bottom right).

The São Paulo state sugarcane sector is important to Brazil, producing more than half of the national volume of bio-ethanol, which makes Brazil the largest sugarcane and second largest bio-ethanol producer in the world. Sugarcane acreage more than doubled over the last fifteen years [21,22]. This expansion, dominantly over grasslands, causes pasture fields to displace to areas

outside of the state and may indirectly affect the preservation of natural vegetation [23,24]. Sugarcane is a semi-perennial crop with a crop cycle of 12–18 months. After 5–7 cycles of growth from the same root system, new shoots are planted due to decreased yield. Usually, the plants reach up to 4 m high [25]. The annual crops in São Paulo state are usually rain-fed, but occasionally irrigated during growth and usually have a growth cycle of between three and five months, depending on the season and crop type. Different harvest practices are applied, especially in the case of corn, whereby the plant is sometimes left for drying of the corn kernels, after which the stalks are cut at 0.5 m above ground, and the field is tilled at a later stage, sometimes months later [26]. The grasslands are usually used as pastures for cattle grazing and can differ considerably in the state of vegetation. Overgrazing results in degraded pasture fields that are dry and with little or no grass cover, occasionally causing severe soil erosion [27]. Healthy pastures usually contain more grass cover, with grass height usually between 0.1 and 0.5 m [26]. When it grows beyond this point, the vegetation approaches shrubland, which belongs to the mid-vegetation class together with citrus trees up to 4 m in height. The forest crop class includes eucalyptus and pine trees, and both have a growth cycle of about seven and eleven years, respectively, and are planted mainly for paper, pulp and timber production. The native forest class mainly comprises Atlantic forest, a moist tropical forest with multiple tree canopies and plants ranging up to 30 m. Native forest is mainly present at the coastal zone, but it can also be found inland, usually covering water drainage channels between crop fields, which are protected by governmental regulations to preserve water presence in the soil, as well as in national parks and isolated patches scattered throughout the state. Built-up is defined as all man-made structures with an impervious surface and covers a broad range of objects, from roads and railways to buildings and industrial factories. Water is defined as any surface covered by a water layer, both natural and artificial water bodies.

*2.2. Study Data*

2.2.1. Sentinel-1 SAR

São Paulo state is covered by Sentinel-1 SAR products in three different modes: Interferometric-Wide (IW), Extra-Wide (EW) and StripMap (SM). The acquisitions in EW and SM modes occur west of São Paulo city and over the city, respectively. Since these modes feature different resolutions and since IW is the common mode, only the Sentinel-1 acquisitions in IW mode were used in this study. Due to the presence of the three different modes, the availability of IW acquisitions varies over São Paulo state. The upper right figure in Figure 1 illustrates the heterogeneous temporal coverage of images over São Paulo state. The wide swath of the EW acquisitions, resulting in very few IW acquisition, is clearly visible in the middle of the state.

The analysis is based on Ground Range-Detected (GRD) products, which were pre-processed applying thermal noise removal, radiometric calibration and terrain correction using the SRTM 30-m resolution Digital Elevation Model (DEM). The nominal IW resolution (range by azimuth) of 20 m by 22 m was spatially averaged to 30 by 30 m grid resolution. All acquisitions were taken during the descending orbit and occurred between 08:20 and 09:02 UTC ($-2$–$-3$ h local time). The time frame ranges from October 2014 until December 2016, resulting in 739 images and 13 billion pixels in time and space. Together with the cross-polarization backscatter values, the corresponding incidence angles (ranging from 29.1°–46.0°) were extracted, as well.

In this study, we have restricted the analyses to backscatter data from the cross-polarized (VH) channel only, for two reasons. First of all, VH is known to be more sensitive to variations of the vegetation cover than the co-polarized channels and is therefore typically used for vegetation monitoring (see Section 1). Secondly, processing and storage capacities limited the data from one channel only, considering the area of interest and time frame.

2.2.2. Precipitation Condition

The precipitation data were extracted from products provided by the Global Precipitation Measurement (GPM) mission operated by NASA and JAXA. GPM offers three-hourly images gridded over 0.1° by 0.1°, which corresponds to approximately 11 km by 11 km in São Paulo state. The specific variable used, precipitationCal, is the multi-satellite precipitation estimate with gauge calibration. It is produced through the assimilation of observations from the Core Observatory satellite, launched in February 2014 and hosting a microwave imager and dual-frequency radar, with sensor data from various other precipitation-relevant passive microwave satellites, and at a later, stage post-processed with data from ground weather stations [28]. For this study, the daily cumulations were taken over the same time frame as the Sentinel-1 data.

In order to assess the reliability of the GPM data, we compared the precipitation measurements against in situ weather stations spread over São Paulo state (indicated by the pink stars in Figure 1). The distance between weather stations ranges from several kilometers to more than 600 km. For each site, at least two weather stations within 10 km were selected and cross-checked to ensure that each station actually records precipitation and is in acceptable agreement with other weather stations on short distances. It was concluded that the level of agreement of detecting precipitation between weather stations themselves is more than 0.85 on short distances (smaller than 5 km), after which, it rapidly drops below 0.75 (around 100 km), and over long distances, the rate reduces to approximately 0.6. Given the scattered positions of available weather stations in São Paulo state, this illustrates the need for precipitation data that are regularly sampled over large areas, as offered by GPM.

When comparing the detected precipitation by GPM with the weather stations, the rate of agreement is 0.70. This discrepancy can be explained by several reasons. First of all, there could be precipitation detected by GPM that is not detected by weather stations, due to the coverage difference of GPM pixels with respect to point measurements of weather stations and due to precipitation evaporation before reaching the ground level. Vice versa, the weather stations may be dependent on a minimum detectable amount of precipitation due to the instrument characteristics. Since accurate precipitation flagging of Sentinel-1 is essential for this study, we analyzed the effects of varying the threshold of measured precipitation on the agreement between the two data sources. From this, it followed that increasing the threshold of the weather stations rapidly decreases the rate of agreement to chance agreement (0.5). Conversely, when increasing the GPM threshold, the agreement increases to a maximum of 0.78 at 10 mm, after which the agreement declines. Hence, in order to increase the confidence in the GPM-derived rain-flagging, we adopted the following criteria for the labeling of Sentinel-1 pixels:

- All Sentinel-1 pixels within a GPM resolution cell were labeled as 'affected by precipitation' if GPM measured more than 10 mm of precipitation for that GPM cell and the eight adjacent GPM cells. This should furthermore hold for the day preceding the Sentinel-1 acquisition, as well as for the day of acquisition.
- All Sentinel-1 pixels within a GPM resolution cell were labeled as 'not affected by precipitation' if GPM measured zero precipitation for that GPM cell and the eight adjacent GPM cells. This should furthermore hold for all three days preceding the Sentinel-1 acquisition, as well as for the day of acquisition.

It should be noted that this labeling approach is conservative and that, consequently, numerous data-points were not rain-labeled and, therefore, not used in the analysis.

As an illustration of the effect of vegetation wetness on Sentinel-1 VH backscatter, Figure 2 shows high backscatter gains from the dry condition to the wet condition for marginal vegetation (low backscatter, colored as blue), approximating the backscatter from high vegetation (high backscatter, colored as red) for both conditions.
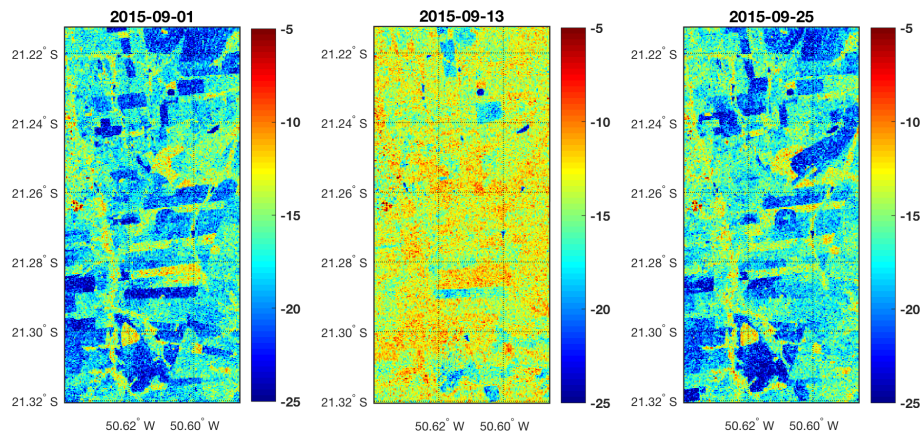
**Figure 2.** Three consecutively-acquired Sentinel-1 Interferometric-Wide (IW) images over the same area illustrating the effects of precipitation on VH backscatter (in dB). The area spans approximately 5 by 10 km. The right and left images show the backscatter, average −15.8 dB and −15.5 dB, respectively, when at least two days of dryness occurred. The middle image, average −13.0 dB, shows the same location with 40 mm of precipitation on the day before the acquisition and 15 mm on the day of acquisition. The blue patches in the dry images are most of all marginally vegetated (harvested fields or early stage crops), and the yellow and red patches are vegetated (mature crops or forests).

### 2.2.3. State of Vegetation

Since the publicly available land cover products do not offer information on the state of vegetation (such as phenology) or vegetation height on a similar time interval as the Sentinel-1 acquisitions (see Section 1), we use land cover maps that were generated by the TU Delft [19,20] (see Figure 1 for the map of 2015). In short, these maps are based on NDVI and NDWI derived from optical Landsat-5 and Landsat-8 data. Based on an adopted version of the Hidden Markov Model (HMM), the state of the land cover is estimated with a regular interval of 16 days (i.e., the temporal revisit of the Landsat satellites). For temporally-dynamic land cover types, specifically annual crops and sugarcane, the growth cycles are divided into respectively seven and 14 states. These states can be roughly linked to phenological stages, ranging from the first two states that can be regarded as bare ground or little vegetation to the last states that can be regarded as senescent plants just before harvest. For grassland, mid-vegetation and forest, no state distinctions are made throughout the season; hence, they are treated as state-stationary in time. The overall accuracies of these maps for the considered years, 2015 and 2016, are between 85% and 90%. The training and validation datasets for accuracy assessments were based on ground reference datasets acquired in São Paulo state, Brazil, including [25,26]. The most notable confusions were found between grasslands and mid-vegetation (high grasslands classified as mid-vegetation and young citrus orchards belonging to mid-vegetation classified as grasslands) and between grasslands, sugarcane and annual crops when gaps in optical observations due to clouds caused misregistration of harvest and early growth events.

The land cover maps, including the temporal states estimated on a 16-day interval, allow us to group states together into discrete vegetation categories. The following acronyms are used throughout this article for concise referencing to the states of vegetation:

- NV: No to marginal Vegetation (indicatively 0 m–0.1 m high): states of annual crop associated with bare ground and the germination stage.
- LV: Low Vegetation (indicatively 0.1 m–1 m high): states of sugarcane associated with early growth stages and grasslands.
- MV: Medium Vegetation (indicatively 2 m–4 m high): states of sugarcane associated with the maturation and senescence stages.
- HV: High Vegetation (indicatively 5–30 m high): forest.

The reason for taking only the first states of annual crops (and not sugarcane) for NV is because these fields are usually cleared or tilled before the growth cycle commences, consequently generally exposing bare ground. The first states of sugarcane on the other hand usually represent the land cover condition where straw material produced during the past sugarcane harvest event covers the remaining sugarcane sprouts for protection against plant dehydration. As such, in order to minimize the plant material present on the ground in this category, these sugarcane states are excluded. Conversely, LV does not include states of annual crop, predominantly because the annual crops may be irrigated during their vegetative stage, which is not captured by the GPM observations. In addition, the annual crops class can include corn and soybean, which differ in plant height and geometry. For MV, the mid-vegetation class was not included in order to avoid the erroneous reported inclusion of low to medium vegetation, such as high grasslands, and medium to high vegetation, such as citrus orchards.

### 2.2.4. Soil Type

For this study, two dominant soil types present in São Paulo state were taken into account, latossolos and argissolos, according to the Brazilian taxonomy convention, and oxisols and ultisols, respectively, according to the Soil Taxonomy of the USDA [29–31]. The location of these two soil types is depicted in the lower right plot of Figure 1. Each of these soil types covers approximately 40% of the state's area and in total supports more than 95% of the state's cultivated land.

Both soil types consist of mineral material, highly weathered by high temperatures and precipitation, and are typical for tropical regions [30]. The main difference is that the oxisols generally have higher clay content and are uniform in clay content, approximately 380 g/kg in the top one-meter layer, whereas for ultisols, a gradual increase in clay content in deeper soil layers is observed (158 g/kg from 0–0.2 m, 201 g/kg from 0.4–0.6 m and 334 g/kg from 0.8–1 m) [32].

This difference also has a direct effect on the SAR microwave interaction through the change in the medium's dielectric constant. As is explained in more detail in [33,34], the dielectric constant is a major contributor to microwave backscatter and is directly related to the soil chemical and physical properties. One of the main variables is soil moisture; the effect of water content on the medium's dielectric constant is dependent on the ability of the water molecule's dipole moments to align along an applied field. In soil, the first added water molecules are tightly bound to the soil particle's surface due to matric and osmotic forces, preventing the free movement of the dipole moments and hence causing a small increase in the dielectric constant. When the water content surpasses the transition (moisture) point, the water molecules become free particles, and the dielectric constant becomes much larger than that of soil. The well-known semi-empirical mixing dielectric model (SMDM) proposed by [35] allows for the computation of the permittivity of microwaves in soil as a function of its frequency, soil moisture and soil texture properties. Specifically, for the two São Paulo soils of interest, from the application of the model, it can be concluded that soil moisture has a major effect on the backscatter magnitude (roughly 5 dB increase). The difference in backscatter magnitude between the two soils themselves is negligible when both are dry and slightly higher backscatter for oxisols (higher clay content) when both are wet (smaller than 1 dB).

Specifically regarding the dry state of these specific soils types, the oxisols are characterized by the presence of deep, well-drained soils [32]. This leads to the additional assumption that the moisture content in soils that are assumed dry under the criteria reported in Section 2.2.2 is lesser for oxisols (i.e., resulting in lower backscatter) than for ultisols. Hence, from the SMDM and the characteristics of the soils, we expect that, on average, the backscatter from oxisols with respect to ultisols is slightly lower when both are dry and is slightly higher when both are wet. Finally, as a general note, C-band microwaves penetrate in the order of a few centimeters for wet soil and a few tens of centimeters for dry soil.

## 3. Methodology

### 3.1. Feature Extraction

In order to label each Sentinel-1 gridded observation with the incidence angle, precipitation condition, state of vegetation and soil type, these data sources were all projected onto the geographical Sentinel-1 grid with a pixel spacing of approximately 30 by 30 m, for each acquisition. Subsequently, only those samples were selected for which the precipitation condition of the consecutive acquisitions were known. From these data, the combined training and validation subset was randomly sampled per land cover type. For each of these samples, the backscatter and backscatter difference between consecutive epochs were taken, referred to as $\sigma_0$ and $\Delta\sigma_0$, respectively. Here, $\sigma_0$ is taken from the first of the two acquisitions, and $\Delta\sigma_0$ is the increase in $\sigma_0$ of the first acquisition to the second acquisition, i.e., $\Delta\sigma_0 = \sigma_0^{t+1} - \sigma_0^{t}$.

### 3.2. Configurations and Precipitation Information Scenarios

We define seven scenarios with different precipitation conditions (Table 2). We introduce acronyms for these scenarios that are referred to throughout this article. The table also shows the percentual occurrences of the precipitation conditions with respect to the total amount of extracted samples. This illustrates the effect of the strict precipitation detection criteria, as was delineated in Section 2.2.2. It shows that less than 25% of the samples were labeled as precipitated or non-precipitated and that less than 10% of the samples were labeled as condition pairs for consecutive acquisitions.

**Table 2.** Precipitation information scenarios, with acronyms used as references. The occurrences show the percentages of detected precipitation condition with respect to the total number of extracted samples. The occurrences between parentheses show the same, but normalized relative to their grouped sum. The variate analysis indicates which variables are used for classification.

| Scenario | Precipitation Information | Occurrence | Variate Analysis |
|---|---|---|---|
| None | None | 100% (100%) | Uni ($\sigma_0$) |
| NP | Non-Precipitated | 18% (73%) | Uni ($\sigma_0$) |
| P | Precipitated | 6% (27%) | Uni ($\sigma_0$) |
| P2NP | Precipitated to Non-Precipitated | 0.9% (12%) | Bi ($\Delta\sigma_0, \sigma_0$) |
| NP2P | Non-Precipitated to Precipitated | 0.8% (10%) | Bi ($\Delta\sigma_0, \sigma_0$) |
| P2P | Precipitated to Precipitated | 0.7% (9%) | Bi ($\Delta\sigma_0, \sigma_0$) |
| NP2NP | Non-Precipitated to Non-Precipitated | 5% (69%) | Bi ($\Delta\sigma_0, \sigma_0$) |

The resulting amount of training samples for each scenario (Table 3) shows that for NV and MV, significantly lower amounts of samples are available than for LV and HV. The relative differences between the precipitation scenarios do not necessarily correspond to those of Table 2 because of the non-uniform occurrence of precipitation and land cover type in São Paulo state.

Finally, in order to evaluate the effect of incidence angle and soil type on the classification performance, we consider five configurations. These differ in ranges of incidence angles and soil types. The ranges of incidence angles were selected such that they were sufficiently disjoint, while still containing enough samples. The configurations overview is given by Table 4, which is placed just before the results and discussion (Section 4) for quick reference.

**Table 3.** Number of training samples for each vegetation state and precipitation scenario. NV, No to marginal Vegetation; LV, Low V; MV, Medium V; HV, High V.

| Scenario | NV | LV | MV | HV |
|----------|------|------|------|------|
| None | 376,809 | 2,885,966 | 210,071 | 4,078,977 |
| NP | 352,150 | 2,177,407 | 180,758 | 3,260,595 |
| P | 24,659 | 708,559 | 29,313 | 818,382 |
| P2NP | 12,093 | 416,469 | 8469 | 489,782 |
| NP2P | 70,982 | 283,664 | 4452 | 284,219 |
| P2P | 12,566 | 292,090 | 20,844 | 328,600 |
| NP2NP | 281,168 | 1,893,743 | 176,306 | 2,976,376 |

### 3.3. Parametrization of Distributions and Classification

Based on the univariate and bivariate histograms for each configuration and each scenario, the Probability Density Functions (PDFs) were fitted using maximum likelihood estimation. As an illustrative example, Figure 3 shows the bivariate histogram of NV per precipitation condition. The center location of the distribution $(\Delta\sigma_0, \sigma_0)$ changes from $(-15$ dB, $-13$ dB) for P2NP to $(+7$ dB, $-21$ dB) for NP2P. For the condition pairs with no precipitation condition change, P2P and NP2NP, the center $\sigma_0$ location is similar to the P2NP and NP2P, respectively, and the center $\Delta\sigma_0$ location is approximately zero.
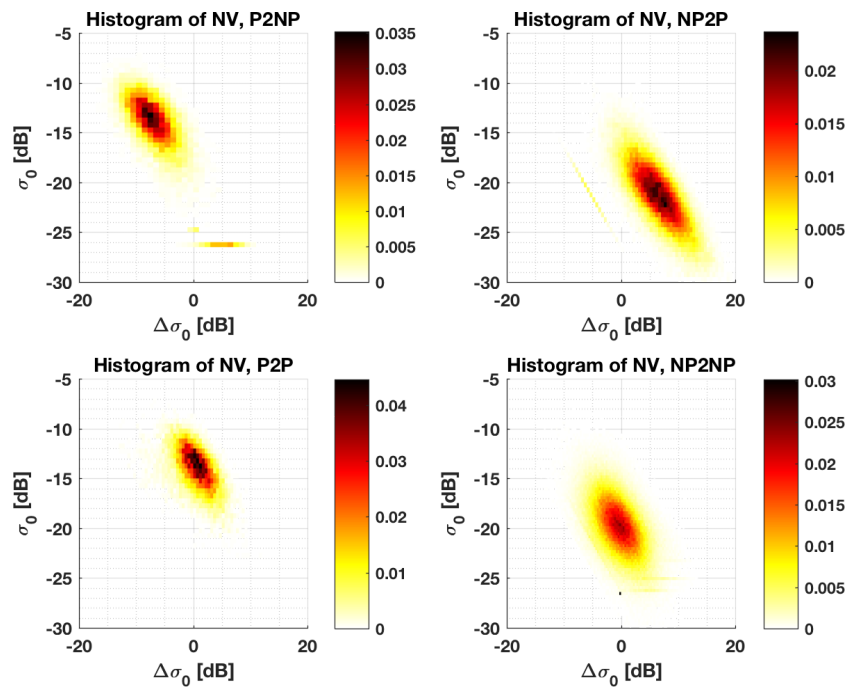


**Figure 3.** Bivariate histograms for NV per precipitation condition pair. The histograms are normalized by the probability density function estimate, i.e., the sum of the bin volumes approximates one.

The estimated PDFs from the NV bivariate histograms are also visualized as the contour lines, $f_{75\%}$, of the 75% Highest Density Region (HDR), defined as $R(f_{75\%}) = \{x : f(x) \geq f_{75\%}\}$, such that $P(X \in R(f_{75\%})) \geq 0.75$ in Figure 4. The differences in location of the distributions for the four precipitation condition pairs is now more clearly illustrated. In addition, the percentile plots of HV illustrate that the $\Delta\sigma_0$ location of the distribution is centered around zero for all precipitation condition pairs. In other words, the effect of precipitation on cross-polarized $\sigma_0$ is not observed for high vegetation, which can be explained by the dominant contribution of microwave volume scattering. Finally, the $\sigma_0$ location of HV is similar to the $\sigma_0$ location of NV for the P2NP and P2P conditions.
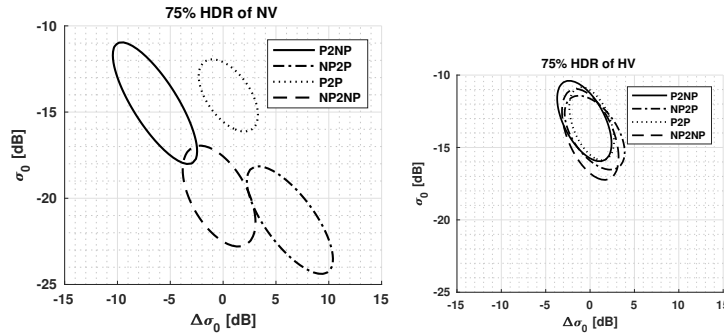
**Figure 4.** The 75% Highest Density Region (HDR) plots for NV and HV for all precipitation condition pairs.

For classification, the log-likelihoods are computed based on the univariate and bivariate distributions given the configuration and precipitation scenario. The vegetation state yielding the maximum log-likelihood is consequently compared to the true state for validation. Resulting confusion matrices are accompanied by the following:

- Overall accuracy: percentage of total number of correctly classified samples with respect to all classified samples.
- Producer's accuracy: fraction of correctly classified samples with respect to all samples of the truth, which is directly related to the omission error and can hence be interpreted as the accuracy from the classification map maker's perspective.
- User's accuracy: fraction of correctly classified samples with respect to all samples classified as this class, which is directly related to the commission error and can hence be interpreted as the accuracy from the classification map user's perspective.

All accuracies are corrected for the differences in the numbers of classified samples per class. Because of this, the overall accuracy is equal to the average of the producer's accuracies. In addition, the classification procedure was carried out ten times with randomly sampled sets of training data (70%) and validation data (30%), allowing for the computation of the averages and standard deviations of the accuracies. The differences between Configurations 2, 3, 4 and 5 with respect to Configuration 1 are presented as the corresponding accuracy improvements with a significance level of 0.05.

*3.4. Hellinger Distances*

To quantitatively describe the similarity of the PDFs for each precipitation scenario, we employ the Hellinger distance statistic, which is used for various applications related to classification techniques [36–40]. This metric $H$, which is closely related to the Bhattacharyya distance [41,42], is dependent on two bivariate normal distributions, $P \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $Q \sim \mathcal{N}(\mu_2, \Sigma_2)$, where $\mu$ and $\Sigma$ are the mean and covariance matrix of the distributions, respectively. Its generic expression [36,41,43] is:

$$\begin{aligned} H(P,Q) &= \sqrt{\frac{1}{2} \int \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx} \\ &= \sqrt{1 - \int \sqrt{f(x)\,g(x)\,dx}} \end{aligned} \tag{1}$$

Here, $H(P,Q) \in [0,1]$ and a value of $H(P,Q) \to 0$ indicates high similarity, and $H(P,Q) \to 1$ indicates low similarity between $P$ and $Q$ [36].

**Table 4.** Configurations with varying incidence angle and soil type used for performance comparisons.

| Configuration | Incidence Angle | Soil Types |
|---|---|---|
| 1 | All | All |
| 2 | 29.1°–35.9° | All |
| 3 | 39.2°–46.0° | All |
| 4 | All | Oxisols |
| 5 | All | Ultisols |

## 4. Results and Discussion

### 4.1. Hellinger Distances

Under Configuration 1, based on Hellinger distances as a metric for the differences between the PDFs of the vegetation states per precipitation scenario (Figure 5), the following observations may be listed:

1.  Generally, the PDFs of P and P2P are less discriminative than the PDFs of the other scenarios. When precipitation occurs, the backscatter of all vegetation states approximate the backscatter of HV, as was also illustrated by Figure 4. For both P and P2P, the distance between NV and HV is smaller than the distances between the other vegetation states and HV. This illustrates the severe effect of precipitation on backscatter when the ground is practically bare, seriously complicating accurate discrimination under such conditions. In addition, the distances between the vegetation states is generally smaller for P than for P2P, implying that the incorporation of two consecutive precipitation-affected acquisitions improves the discriminatory power as compared to a single precipitation-affected acquisition.

2.  The distances for the scenarios with no information (None) and NP are similar to each other due to the relative underrepresentation of samples associated with scenario P (also see Table 3). As for the precipitation condition pairs, NP2P and NP2NP are similar to each other, but differ from P2NP. The latter mainly has larger differences between the PDF of NV and the PDFs of other vegetated states (LV, MV and HV). These latter states also show smaller differences between each other. This may be explained by P2NP causing higher remaining moisture content in vegetation (i.e., for LV MV and HV) after the first precipitation event as compared to the remaining moisture in soil when NV due to faster evaporation for bare ground.

3.  Overall, there is relatively high confusion between LV and MV and to a lesser extent between LV and NV. This can be ascribed to the broad range of grasslands in the LV class and the confusion of higher grasslands with bushland; also see Section 2.2.3.
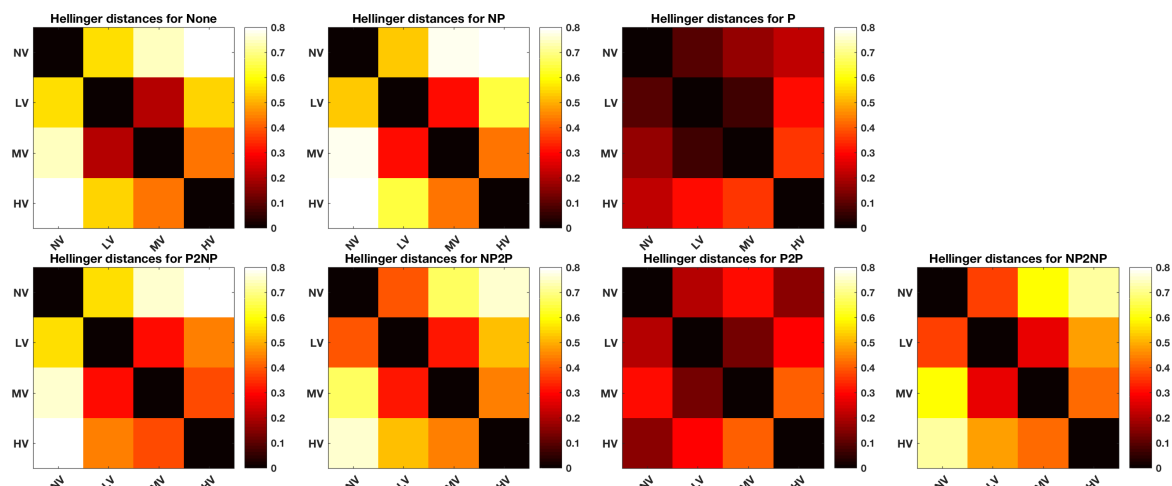


**Figure 5.** Hellinger distances between the PDFs of all vegetation states and for all precipitation scenarios.

## 4.2. Classification Results with Precipitation Conditions

The overall accuracies from the classification under Configuration 1 (Figure 6) first of all show that omitting P-related observations results in a marginal improvement of the classification performance from None to NP. Conversely, only taking P-related data significantly deteriorates the performance. This asymmetric behavior, which also applies to the producer's and user's accuracies, can be ascribed, similarly to before, to the difference in the number of data samples, i.e., their corresponding occurrences. The producer's accuracies for NV for both P and P2P are relatively low with respect to the user's accuracy due to the classification confusion with the other vegetation states (as was described in Discussion Point 1 in Section 4.1), particularly with HV for P2P.
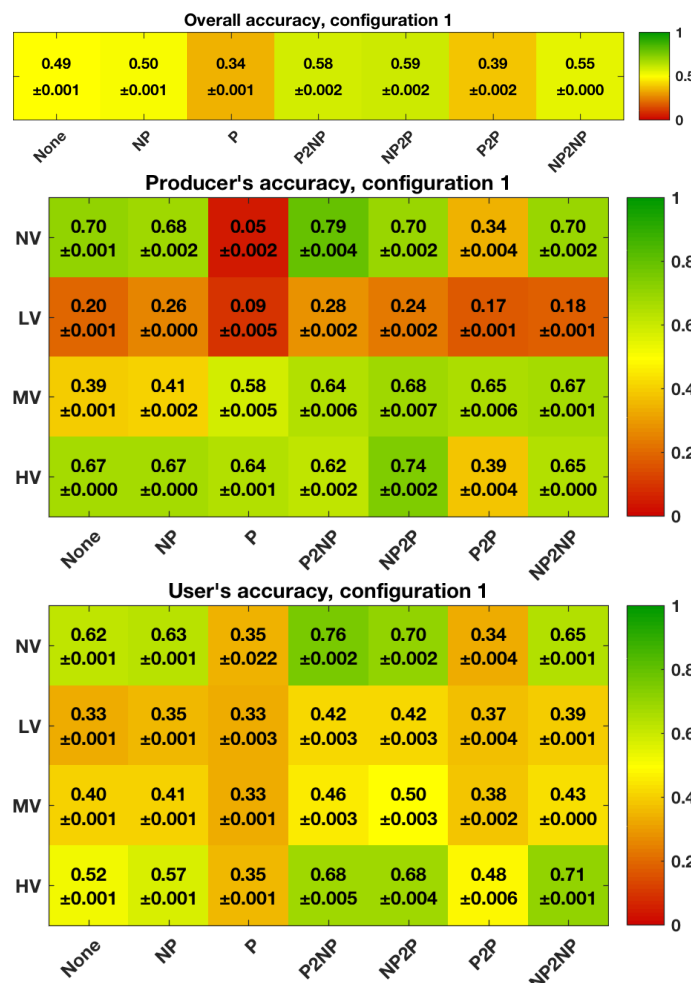


**Figure 6.** The averages and standard deviations of the overall (top), producer's (middle) and user's (bottom) accuracies for Configuration 1.

A greater classification improvement can be achieved when the bivariate information is taken into account, firstly demonstrated by the NP2NP-related accuracies with respect to the None-related accuracies. Hence, even though the center $\triangle\sigma_0$ location is close to zero (as indicated in Section 3.3), the second observation still adds valuable information to the classifier. In addition, the absolute difference in overall accuracies between NP2NP and P2P is similar to the difference between NP and P.

The best classification performances can be reached when changing precipitation conditions occur for consecutive acquisitions, as demonstrated by the increase in overall accuracies for P2NP and NP2P. With respect to NP2NP, the increases in the overall accuracies of P2NP and NP2P are approximately equal. Nevertheless, the producer's and user's accuracies show differences, both commonly for NV.

This is reflected by the larger Hellinger distances between NV and the other states for P2NP with respect to NP2P, as was described and explained by Discussion Point 2 in Section 4.1.

Finally, the general confusion of LV with the other states, as was laid out by Discussion Point 3 in Section 4.1, consequently results in relatively low producer's accuracies for all precipitation scenarios.

### 4.3. Effect of Incidence Angle on Classification Results

Restricting the training and evaluation to a narrower range of incident angles (Configurations 2 and 3) has a marginal effect on the classification accuracy (Figure 7). The slight deterioration could be ascribed to the non-uniform geographical distribution of the land cover classes and incidence angles, causing lower number of samples per subset and potentially an over-representation of clusters within one land cover type (especially for LV).
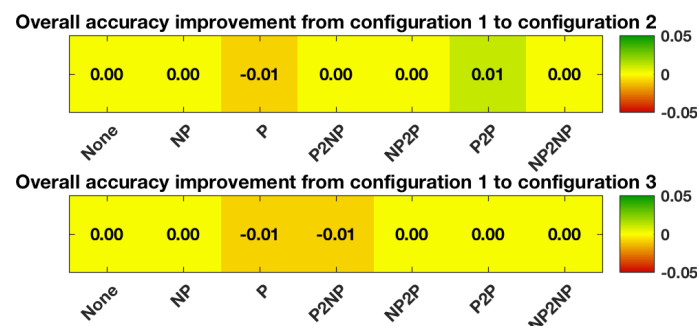


**Figure 7.** Overall accuracy improvements from Configuration 1 to Configuration 2 (top) and from Configuration 1 to Configuration 3 (bottom). Only the improvements with a significance level of 0.05 are non-zero.

### 4.4. Effect of Soil Type Information on Classification Results

The soil-related differences, on the other hand, show significant improvements, particularly for Configuration 5 (ultisols, lower clay content) for scenarios P, NP2P and P2P. These can be mainly attributed to less confusion of NV with the other vegetation states. A comparison of the PDF of NV with the PDF of MV for NP2P and the PDF of NV with the PDF of HV for P2P for Scenarios 4 and 5 with Scenario 1 (Figure 8) first of all indicates that the PDFs of Scenario 4 (oxisols, higher clay content) closely resemble the PDFs of Scenario 1 (all soil types combined), which is evidently based on data where the oxisols soil is over-represented.

Furthermore, as was earlier conveyed in Section 2.2.4, a slightly lower $\sigma_0$ when dry and slightly higher $\sigma_0$ when wet for Configuration 4 (oxisols, higher clay content) with respect to Scenario 5 (ultisols, lower clay content) was expected and is confirmed by the NP2P and P2P plots, respectively. For NP2P, this consequently leads to a higher $\Delta\sigma_0$ for Scenario 4. The combination of the differences in $\sigma_0$ and $\Delta\sigma_0$ between the two scenarios results in a smaller confusion with particularly MV for NP2P. For P2P the smaller confusion can mostly be attributed to the difference in $\sigma_0$ for NV with respect to HV. Both result in the main accuracy improvements, as is shown in Figure 9.
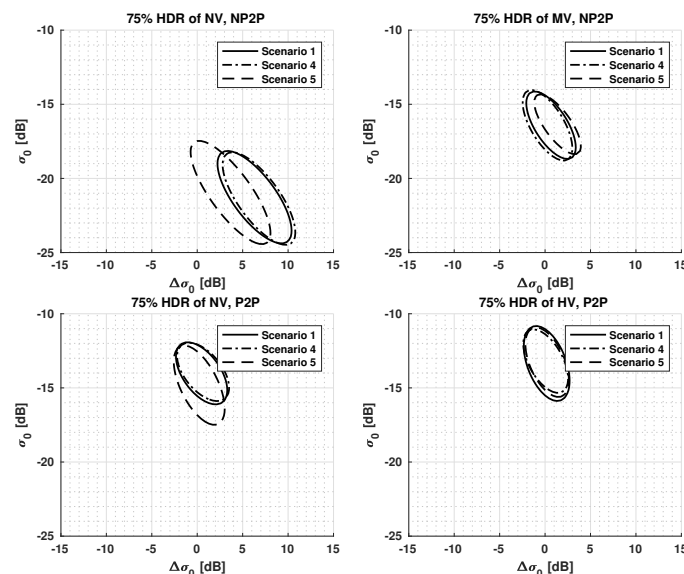
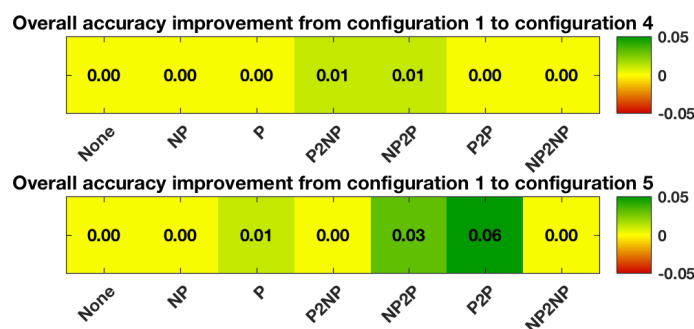**Figure 8.** The 75% HDR plots for NV for NP2P and P2P for Scenarios 1, 4 and 5.



**Figure 9.** Overall accuracy improvements from Configuration 1 to Configuration 4 (top) and from Configuration 1 to Configuration 5 (bottom). Only the improvements with a significance level of 0.05 are non-zero.

## 5. Conclusions

Incorporating information on surface wetness combined with the signal from the consecutive SAR observation can significantly improve the characterization of vegetation. Moreover, a change in wetness condition between acquisitions further increases the characterization accuracy. This is caused by the non-uniform increase in SAR backscatter for the different vegetation states from dry to wet condition and vice versa. In order to maximize the accuracy even further, the classification model may be trained for different soil types as well, which, depending on their conductivity properties, increases the difference between probability density functions and hence enhances the separability between NV and other vegetation states.

Since the effect of precipitation on backscatter differences is dependent on the amount of vegetation, other gradations in vegetation coverage than used in this article may be investigated. As an example, improved discrimination may be possible between degraded grasslands and healthy grasslands, due to their difference in soil contribution on the SAR signals. Furthermore, the potential of co-polarized backscatter (VV in the case of Sentinel-1 IW in this area) for vegetation discrimination based on precipitation information may be studied, as well.

The methodologies as laid out in this article are not recommended for utilization as stand-alone classification techniques. Rather, we stress the concept that variations in surface wetness condition can actually be used advantageously in SAR-based land cover monitoring or classification models.

## References

1. Xiong, J.; Thenkabail, P.S.; Tilton, J.C.; Gumma, M.K.; Teluguntla, P.; Oliphant, A.; Congalton, R.G.; Yadav, K.; Gorelick, N. Nominal 30-m cropland extent map of continental Africa by integrating pixel-based and object-based algorithms using Sentinel-2 and Landsat-8 data on google earth engine. *Remote Sens.* **2017**, *9*, 1065. [CrossRef]

2. Chen, J.; Chen, J.; Liao, A.; Cao, X.; Chen, L.; Chen, X.; He, C.; Han, G.; Peng, S.; Lu, M.; et al. Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS J. Photogramm. Remote Sens.* **2014**, *103*, 7–27. [CrossRef]

3. Jokar Arsanjani, J.; Tayyebi, A.; Vaz, E. GlobeLand30 as an alternative fine-scale global land cover map: Challenges, possibilities, and implications for developing countries. *Habitat Int.* **2016**, *55*, 25–31. [CrossRef]

4. Bartholomé, E.; Belward, A.S. GLC2000: A new approach to global land cover mapping from Earth observation data. *Int. J. Remote Sens.* **2005**, *26*, 1959–1977. [CrossRef]

5. Friedl, M.A.; Sulla-Menashe, D.; Tan, B.; Schneider, A.; Ramankutty, N.; Sibley, A.; Huang, X. MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sens. Environ.* **2010**, *114*, 168–182. [CrossRef]

6. Defourny, P.; Schouten, L.; Bartalev, S.; Bontemps, S.; Caccetta, P.; Bella, C.D.; Gond, V.; Hazeu, G.W.; Heinimann, A.; Herold, M.; et al. Accuracy Assessment of a 300 m Global Land Cover Map: The GlobCover Experience. In Proceedings of the 33rd International Symposium on Remote Sensing of Environment, Sustaining the Millennium Development Goals, Stresa, Italy, 4–8 May 2009; pp. 1–5.

7. Bontemps, S.; Defourny, P.; Bogaert, E.V.; Kalogirou, V.; Perez, J.R. GLOBCOVER 2009 Products Description and Validation Report. *ESA Bull.* **2011**, *136*, 53.

8. Fritz, S.; See, L.; McCallum, I.; Schill, C.; Obersteiner, M.; Van Der Velde, M.; Boettcher, H.; Havlík, P.; Achard, F. Highlighting continued uncertainty in global land cover maps for the user community. *Environ. Res. Lett.* **2011**, *6*, 1–6. [CrossRef]

9. Herold, M.; See, L.; Tsendbazar, N.E.; Fritz, S. Towards an integrated global land cover monitoring and mapping system. *Remote Sens.* **2016**, *8*, 1036. [CrossRef]

10. Steele-Dunne, S.C.; McNairn, H.; Monsivais-Huertero, A.; Judge, J.; Liu, P.W.; Papathanassiou, K. Radar Remote Sensing of Agricultural Canopies: A Review. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2249–2273. [CrossRef]

11. Baghdadi, N.; Boyer, N.; Todoroff, P.; El Hajj, M.; Bégué, A. Potential of SAR sensors TerraSAR-X, ASAR/ENVISAT and PALSAR/ALOS for monitoring sugarcane crops on Reunion Island. *Remote Sens. Environ.* **2009**, *113*, 1724–1738. [CrossRef]

12. McNairn, H.; Brisco, B. The application of C-band polarimetric SAR for agriculture: A review. *Can. J. Remote Sens.* **2004**, *30*, 525–542. [CrossRef]

13. Moran, M.S.; Alonso, L.; Moreno, J.F.; Cendrero Mateo, M.P.; de la Cruz, D.F.; Montoro, A. A RADARSAT-2 Quad-Polarized Time Series for Monitoring Crop and Soil Conditions in Barrax, Spain. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1057–1070. [CrossRef]

14. Kontgis, C.; Warren, M.S.; Skillman, S.W.; Chartrand, R.; Moody, D.I. Leveraging Sentinel-1 time-series data for mapping agricultural land cover and land use in the tropics. In Proceedings of the 2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp), Brugge, Belgium, 27–29 June 2017; pp. 1–4.

15. Balzter, H.; Cole, B.; Thiel, C.; Schmullius, C. Mapping CORINE land cover from Sentinel-1A SAR and SRTM digital elevation model data using random forests. *Remote Sens.* **2015**, *7*, 14876–14898. [CrossRef]

16. Torbick, N.; Chowdhury, D.; Salas, W.; Qi, J. Monitoring rice agriculture across myanmar using time series Sentinel-1 assisted by Landsat-8 and PALSAR-2. *Remote Sens.* **2017**, *9*, 119. [CrossRef]

17. El Hajj, M.; Baghdadi, N.; Zribi, M.; Angelliaume, S. Analysis of Sentinel-1 radiometric stability and quality for land surface applications. *Remote Sens.* **2016**, *8*, 406. [CrossRef]

18. Satalino, G.; Balenzano, A.; Mattia, F.; Davidson, M.W. C-band SAR data for mapping crops dominated by surface or volume scattering. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 384–388. [CrossRef]

19. Iannini, L.; Molijn, R.; Mousivand, A.; Hanssen, R. Monitoring LULC dynamics in the Sao Paulo region through landsat and C-band SAR time series. In Proceedings of the 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Milan, Italy, 26–31 July 2015; pp. 90–93.

20. Iannini, L.; Molijn, R.; Mousivand, A.; Hanssen, R.; Lamparelli, R.A.C. A HMM-based approach for historic and up-to-date land cover mapping through Landsat time-series in the state of Sao Paulo, Brazil. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 5457–5460.

21. Machado, P.G.; Rampazo, N.A.M.; Picoli, M.C.A.; Miranda, C.G.; Duft, D.G.; de Jesus, K.R.E. Analysis of socioeconomic and environmental sensitivity of sugarcane cultivation using a Geographic Information System. *Land Use Policy* **2017**, *69*, 64–74. [CrossRef]

22. INPE. Canasat. Available online: http://www.dsr.inpe.br/canasat/ (accessed on 16 October 2018).

23. Alkimim, A.; Clarke, K.C. Land use change and the carbon debt for sugarcane ethanol production in Brazil. *Land Use Policy* **2018**, *72*, 65–73. [CrossRef]

24. Rudorff, B.F.T.; de Aguiar, D.A.; da Silva, W.F.; Sugawara, L.M.; Adami, M.; Moreira, M.A. Studies on the Rapid Expansion of Sugarcane for Ethanol Production in São Paulo State (Brazil) Using Landsat Data. *Remote Sens.* **2010**, *2*, 1057–1076. [CrossRef]

25. Molijn, R.A.; Iannini, L.; Rocha, J.V.; Hanssen, R.F. Detailed ground reference data for sugarcane biomass estimation in São Paulo state, Brazil. *Sci. Data* **2018**, *5*, 1–18. [CrossRef] [PubMed]

26. Molijn, R.A. Land cover ground reference data in São Paulo state, Brazil, taken in 2015. TU Delft, 2015. Dataset.

27. Merten, G.H.; Minella, J.P. The expansion of Brazilian agriculture: Soil erosion scenarios. *Int. Soil Water Conserv. Res.* **2013**, *1*, 37–48. [CrossRef]

28. Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J. *Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation*; Technical Report; NASA: Washington, DC, USA, 2017.

29. USDA. *Keys to Soil Taxonomy*; Technical Report; USDA: Washington, DC, USA, 2014.

30. Eswaran, H.; Reich, P. World Soil Map. In *Encyclopedia of Soils in the Environment*; Elsevier: Amsterdam, The Netherlands, 2005; pp. 352–365.

31. Embrapa. *Sistema Brasileiro de Classificação de Solos*, 2nd ed.; Embrapa Informação Technológica: Brasília, Brazil, 2006; p. 306, ISBN 85-85864-19-2.

32. Alves, M.R.; Demattê, J.A.M.; Barros, P.P.S. Multiple Geotechnological Tools Applied to Digital Mapping of Tropical Soils. *Rev. Bras. Ciênc. Solo* **2015**, *39*, 1261–1274. [CrossRef]

33. Behari, J. *Microwave Dielectric Behavior of Wet Soils*; Springer: Dordrecht, The Netherlands, 2005; ISBN 978-1-4020-3271-4.

34. Gadani, D.H.; Vyas, A.D. Dielectric mixing model for the estimation of complex permittivity of wet soils at C and X band microwave frequencies. *Indian J. Pure Appl. Phys.* **2015**, *53*, 190–198.

35.   Dobson, M.C.; Ulaby, F.T.; Hallikainen, M.T.; El-Rayes, M.A. Microwave Dielectric Behavior of Wet Soil-Part II: Dielectric Mixing Models. *IEEE Trans. Geosci. Remote Sens.* **1985**, *GE-23*, 35–46. [CrossRef]

36.   Salmon, B.P.; Kleynhans, W.; Van Den Bergh, F.; Olivier, J.C.; Marais, W.J.; Grobler, T.L.; Wessels, K.J. A search algorithm to meta-optimize the parameters for an Extended Kalman filter to improve classification on hyper-temporal images. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 4974–4977.

37.   Silva, W.B.; Freitas, C.C.; Sant'Anna, S.J.; Frery, A.C. Classification of segments in PolSAR imagery by minimum stochastic distances between wishart distributions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2013**, *6*, 1263–1273. [CrossRef]

38.   Salmon, B.P.; Kleynhans, W.; Bergh, F.V.D.; Olivier, J.C.; Marais, W.J.; Wessels, K.J. Meta-optimization of the extended Kalman filter's parameters for improved feature extraction on hyper-temporal images. In Proceedings of the 2011 IEEE International Geoscience and Remote Sensing Symposium, Vancouver, BC, Canada, 24–29 July 2011; pp. 2543–2546.

39.   Erol, H.; Akdeniz, F. A new supervised classification method for quantitative analysis of remotely-sensed multi-spectral data. *Int. J. Remote Sens.* **1998**, *19*, 775–782. [CrossRef]

40.   Genovez, P.C.; Freitas, C.C.; Santanna, S.J.; Bentz, C.M.; Lorenzzetti, J.A. Oil Slicks Detection from Polarimetric Data Using Stochastic Distances between Complex Wishart Distributions. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 463–477. [CrossRef]

41.   Subasi, Y.; Demirekler, M. Quantitative measure of observability for stochastic systems. *IFAC Proc. Vol. (IFAC-PapersOnline)* **2011**, *18*, 4244–4249. [CrossRef]

42.   Girish, V. Hellinger distance. *Commun. Complex.* **2011**, *2011*, 12.

43.   Frasca, M.; Liberati, R. Riemann manifolds from Hellinger distance. In Proceedings of the 2012 Tyrrhenian Workshop on Advances in Radar and Remote Sensing: From Earth Observation to Homeland Security, TyWRRS 2012, Naples, Italy, 12–14 September 2012; pp. 59–61.