Article

Detection of Crop Hail Damage with a Machine Learning Algorithm Using Time Series of Remote Sensing Data

Leandro Sosa [1,*] , Ana Justel [2] and Íñigo Molina [1]

1   Department of Surveying and Cartography Engineering, Universidad Politécnica de Madrid, Campus Sur UPM, C/ Mercator 2, 28031 Madrid, Spain; inigo.molina@upm.es
2   Mathematics Department, Universidad Autónoma de Madrid, 28049 Madrid, Spain; ana.justel@uam.es
*   Correspondence: leandroleonelsosa@gmail.com

Abstract: Hailstorms usually result in total crop loss. After a hailstorm, the affected field is inspected by an insurance claims adjuster to assess yield loss. Assessment accuracy depends largely on in situ detection of homogeneous damage sectors within the field, using visual techniques. This paper presents an algorithm for the automatic detection of homogeneous hail damage through the application of unsupervised machine learning techniques to vegetation indices calculated from remote sensing data. Five microwave and five spectral indices were evaluated before and after a hailstorm in zones with different degrees of damage. Dual Polarization SAR Vegetation Index and Normalized Pigment Chlorophyll Ratio Index were the most sensitive to hail-induced changes. The time series and rates of change of these indices were used as input variables in the K-means method for clustering pixels into homogeneous damage zones. Validation of the algorithm with data from 91 soybean, wheat, and corn plots showed that in 87.01% of cases there was significant evidence of differences in average damage between zones determined by the algorithm within the plot. Thus, the algorithm presented in this paper allowed efficient detection of homogeneous hail damage zones, which is expected to improve accuracy and transparency in the characterization of hailstorm events.

Keywords: algorithm; agricultural; pattern classification; radar polarimetry; remote sensing; spectral analysis; time series; unsupervised learning; vegetation mapping

1. Introduction

In the last decades, both the frequency and intensity of extreme weather and climate events have increased worldwide, leading to huge economic losses [1]. In Argentina, the number of intense storms due to climate change has already increased in the last few years [2]. The frequency of hail events is higher in midlatitudes during the warm season [2]. The Pampas region of Argentina, the most important crop production area in the country, is located in midlatitudes, and the time of higher frequency of hail events overlaps both the stage of the highest vulnerability of winter crops (such as wheat, which reaches the fruiting stage in the warm season) and the whole growth period of summer crops such as soybeans and corn.

Farmers aim to mitigate the consequences of natural catastrophes through management strategies within their farming operations, but certain risks, such as hail events, are beyond their financial means. For such cases, farmers usually transfer the risk to specialized risk management companies by taking out agricultural insurance policies, with hail damage policies being very widespread in the country.

Following a hailstorm, an insurance claims adjuster goes to the field and estimates the yield losses due to hail damage. The accuracy of estimates relies heavily on the detection of Homogeneous Damage Zones (HDZs), which allows extrapolating the data from field samples to the whole area of the damaged field. Currently, estimates of the surface area of HDZs are performed visually, and the adjuster has no information about either the affected area within the field or the degree of damage before getting to the site for inspection. Since

crop fields in Argentina are often extensive, with areas of 500 hectares being usual, it is unfeasible to determine HDZs in the field, particularly in the case of tall crops and with very heterogeneous degrees of damage.

One alternative is to apply machine-learning techniques to remote sensing data for the detection of crop changes due to hailstorms. The ability and accuracy of remote sensing to detect and classify changes in the physical, physiological, and chemical properties of a crop vary both with the type of sensor and algorithms used for data processing. Thus, for correct implementation, it is necessary to have precise knowledge of the advantages and limitations of each sensor and the machine-learning techniques available. Two types of sensors were used in this research: active microwave Synthetic Aperture Radars (SAR), and passive multispectral. Data obtained with these sensors are pre-processed and used as input variables for the machine-learning algorithm to detect HDZs.

The total scattered microwave signal that returns to the sensor from a vegetated surface is known to be affected directly by plant cover, by the soil (attenuated by the canopy), and by vegetation-underlying soil interactions [3]. Thus, the microwave signal is sensitive to changes in canopy architecture and geometric configuration, whether such changes have been caused by normal plant growth or by the occurrence of a hailstorm. Since microwave signals are also sensitive to surface roughness and the dielectric coefficient of the materials observed [4], backscatter between two successive captures might be more influenced by wind or rainfall, which affect surface roughness or the dielectric properties of the materials, than by the growth of the crop itself [5] or by hail-caused changes. Several authors [6–10] have proposed using indices calculated with SAR data to estimate biophysical crop parameters, with procedures for smoothing the undesired characteristics in the backscattered signal. The microwave spectrum range, on the other hand, has a clear advantage over multispectral data in not requiring generally atmospheric corrections since it is not sensitive to atmospheric interferences.

Optical sensors capture the sunlight reflected by objects, thus allowing accurate identification of crop characteristics. Filters placed in the sensor record the reflectance received in different layers of an image as wavelength ranges. Optical images require correction of intermediate sun glint caused by atmosphere interferences as well as detection of clouds and cloud shadows. Once this drawback is overcome, optical images are easier to interpret than those obtained from SAR sensors. Since each type of sensor has its advantages, it is reasonable to infer that a combination of both types of signals, microwave and multispectral, would optimize the detection of changes in crop biophysical parameters.

Following this approach, an arithmetic formulation including two or more wavelength ranges in the multispectral signal or types of polarization in the microwave signal is used to generate vegetation indices to describe crop biophysical variables. Although vegetation indices are easy to use, the limited number of variables they include may not reflect the complexity of field conditions [11]. This disadvantage can be minimized by selecting the indices most sensitive to the changes in the variable of interest, which will allow differentiating plant groups with different degrees of change.

Once the biophysical variables to be assessed along the crop growth period have been defined, the challenge is to achieve a correct classification of the changes observed, with machine-learning techniques having been successfully applied for this purpose in agriculture [12]. These techniques generate specific, accurate, and automated outputs from high volumes of remote-sensing complex data [13]. However, implementing machine-learning methods is not easy, and the literature provides contradictory information regarding the key steps in the process. The complex decisions involved include the choice of algorithm, requirements of input data, selection and optimization of user-defined parameters, and computational costs [14].

ML techniques are most generally classified into supervised and unsupervised. They differ from each other in that the former uses labels in a data subset to train the model to be subsequently applied to the whole set, whereas the latter searches for patterns in unlabeled data. Since there was no training sample available for this research, we used an

unsupervised technique. The technique was applied to identify damage patterns in data from different crops affected by hailstorms at different phenological stages.

Although Earth satellite data have been available for more than 30 years, to the best of our knowledge no research has been conducted on the use of SAR data for hail damage to crops. Multispectral data have been used to analyze some local and sporadic cases [15–17] which, however, pose practical limitations for regional and operational application [18]. Thus, the purpose of this work was to develop an algorithm for determining HDZs in crops affected by hailstorms, following a standardized and accurate methodology for processing satellite data. Such a development will confer objectivity and transparency to crop damage assessment.

The following section describes the study area and the importance of agricultural insurance in the region. The characteristics and pre-processing of satellite data are also included in this section. Next, we explain the methodology used to select the indices with the highest sensitivity to hail-induced changes in the crop, as well as the steps followed by the algorithm to classify pixels into HDZs and the algorithm validation methodology. Section 3 details the results of the sensitivity analysis for the different vegetation indices and identifies the most sensitive for each signal type. The results and discussion of the algorithm steps are also included in this section, which finishes with the results and discussion of algorithm validation with in situ data. Conclusions are presented in Section 4.

2. Materials and Methods

2.1. Study Area

The study was conducted in the Pampas, a huge sedimentary basin formed on top of the Brasília Massif in South America. The climate in this region is subtropical humid in the north and temperate in the south, with warm summers and cool and variable winters with frequent frosts but no snowfall. The average annual temperature is 17 °C. It presents a monsoon rainfall regime, with annual averages ranging from 1200 mm in the east, an area under the influence of Atlantic winds, to 300 mm in the west, where the dry, hot winds blowing across the Andes Mountains predominate. The Pampas is an extensive plain extending over the Argentine provinces of Buenos Aires, Entre Ríos, Santa Fe, Córdoba, La Pampa, and San Luis (Figure 1). Formerly characterized by rangelands and the absence of trees, today it is the most industrialized region in the country. The industrialization process began with cattle raising and the manufacturing of animal products for export continued with agricultural development and is currently fueled by soybeans, corn, wheat, and their by-products.
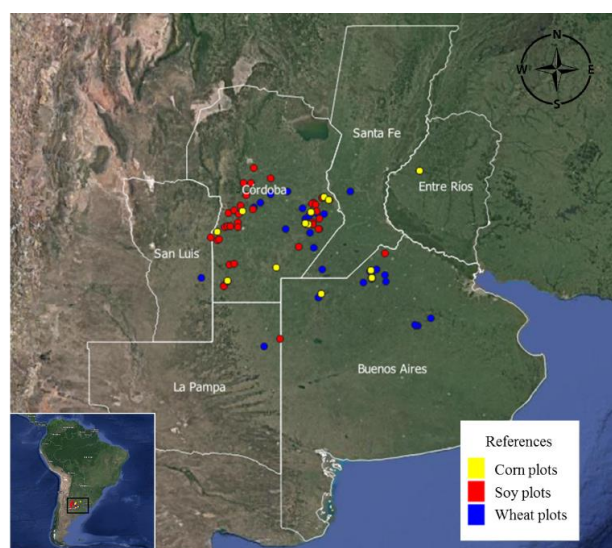
Figure 1. Study area and location of the plots evaluated for methodology validation.

In Argentina, the regions with the highest hailstorm frequencies are located near the Córdoba, Tandil, and La Ventana Mountains, where orographic storms occur, and along the Atlantic coast, where the disturbances caused by cold air currents reaching the continent contribute to the formation of hailstorms [19]. Historically, the damage is greater in winter crops such as wheat since intense storms occur mainly in late spring and early summer, when the filled grains are exposed [20].

The area allocated to crop production shows small variations between years. In the 2019–2020 growing season, the areas sown with soybeans, corn, and wheat were 16.9, 9.5, and 6.9 million hectares respectively (http://datosestimaciones.magyp.gob.ar/ (accessed on 10 March 2020)). In the same period, the policies issued by all the insurance market for Agricultural and Forestry Risks amounted to US$ 250 million (https://www.argentina.gob.ar/superintendencia-de-seguros (accessed on 25 July 2020)), 91.5% of which was taken out for hail damage, as follows: 37.6% for extensive soybean crops, 37.1% for corn, and 16.8% for wheat.

2.2. Data

Data were obtained from Sentinel-1 and -2 Missions of the Copernicus Program (European Space Agency and European Environment Agency, EU), and from in situ damage assessments in 91 plots affected by hailstorms between 2017 and 2020. All plots are located within the study area, 44 planted with soybeans, 32 with wheat, and 15 with corn (Figure 1). The evaluated plots had an average surface of 137.97 hectares, with a minimum value of 13 and a maximum of 590 hectares; the values of the quartiles Q1, Q2, and Q3 were 69.75, 97.5, and 162.75 hectares respectively. Table S1 shows the central point of each plot, the crop name, and the date of the hailstorm which caused the damage.

Sentinel-1 is equipped with C-band, dual-polarization Synthetic Aperture Radar (SAR) sensors operating at a center frequency of 5.405 GHz. Sensors are mounted on two satellite platforms (A and B), each of which registers images of the same area every 12 days. In most of the Argentinian territory, images overlap, and their incidence angles differ considerably between the platforms. To avoid this problem, when data from platforms A and B overlapped, the first to capture images after a given hailstorm was selected.

The images used were Ground Range Detected, processed from Single Look Complex products with Sentinel-1 Toolbox (http://step.esa.int/main/toolboxes/snap (accessed on 12 December 2018)). This process generates a calibrated, ortho-rectified product, with thermal noise removal, and radiometric calibration, and terrain correction using the SRTM 30 digital elevation model, generating a product with a 10 m spatial resolution.

Sentinel-2 is a remote sensing system that uses the sensors MSI (MultiSpectral Instrument) capturing images in the visible light spectrum, near-infrared, and short-wavelength infrared bands, to record solar radiation reflected by objects on Earth in thirteen 16-bit spectral ranges. Similarly to Sentinel 1, two MSI sensors are mounted on two S2 platforms (A and B). The wavelength ranges used were blue (B2), red (B4), and near-infrared (B8), all of them with 10 m spatial resolution and obtained from a product with Level-1C Processing. The data thus obtained represent the reflectance in the upper layer of the atmosphere. The time resolution of Sentinel-2 is 5 days, or 2 and 3 days in sectors with orbit overlap.

For each assessed plot, data were obtained from satellite images from both missions, within the plot boundaries, with a time frame of 60 days prior to and 60 days after the hailstorm. When the time frame included any date prior to sowing or after harvest of each crop, the series was interrupted to avoid entering data from other crops. Satellite data were retrieved from and processed in the Google Earth Engine platform.

In situ damage assessment for each crop was performed following the United States Department of Agriculture and Federal Crop Insurance Corporation guidelines [21–23]. At least ten sampling stations were established in each of the 91 plots analyzed, totaling 1174 observations (646 in soybean plots, 153 in corn, and 375 in wheat). Latitude, longitude, and estimated damage percentage were recorded in each sampling station.

2.3. Satellite Data Pre-Processing

SAR images are formed through coherent processing of the return of successive RADAR pulses. When a RADAR sheds light upon a rough surface, the distances between the elementary scatterers and the receiver vary due to surface roughness. As a result, the backscattered waves are coherent in frequency but not in phase. The signal received is strong if the waves add together in a relatively constructive way, whereas it is weak if phase shifts lead to intensity variations between adjacent pixels. These variations appear as a granular pattern, called speckle [24], which makes visual and digital image interpretation more difficult. The technique most widely used is to remove speckle noise in spatial averaging. This technique, however, lowers image resolution significantly, resulting in blurry images. Morphological filters reduce speckle noise without losing image focus [25] and also have an excellent computational performance. These filters transform images by applying successive convolutions. To do this, a kernel function (e.g., maximum, minimum, mean, median) is assigned to a moving window. The kernel moves over the original image performing the calculations for each pixel. The value obtained as output is assigned as the new value of the central pixel in the image. Thus, the pixel values in the new image are modified according to the values of its neighboring pixels in the original image. Kernels can be circular, square, cross-or diamond-shaped, or octagonal, and their size is defined by the user. To remove speckle noise, the median for each pixel was calculated on a circular kernel with a 15-pixel radius. A median rather than an average was used because speckles appear in random positions and their values vary greatly from other pixels representing the same object, which could lead to distortions in the transformation of speckled areas of an image if an average were used. Kernel size was defined according to plot size and the areas to be differentiated within the plot, whereas kernel shape was defined according to the shape of the areas to be differentiated. The crop changes to be detected are determined by soil variables such as salinity, texture, and topography before the hailstorm, and by hail after the storm. A variation in any of these variables will be strongly correlated with the spatial variable, generating irregular patterns [26]. The circular kernel best fitted this pattern.

Although there are numerous methodologies to correct atmospheric interferences in multispectral data, not all can be applied to Sentinel-2 data because it does not record data in the thermal infrared wavelength range. Since atmospheric interferences appear as unexpected and occasional changes along with successive images in a collection, and crop changes manifest themselves gradually along time, temporal methods such as Tmask [27] and MAJA [28] are considered most suitable to perform atmospheric corrections. However, Tmask and MAJA have a high computational cost. Their application to HDZs would slow the algorithm, making it unsuitable for the quick responses required for crop damage insurance claims. Thus, a quick-running time method was selected to filter clouds and their shadows and correct other atmospheric interferences, namely the mobile linear regressions method, which is applied separately to each pixel time series [29]. A 15-day bandwidth was used to ensure interference removal without adversely affecting the values by data degradation due to excessive smoothing.

2.4. Selection of Vegetation Indices

Vegetation indices allow identifying and describing structural, phenological, and biophysical parameters, which is a temporal line basis help detecting changes in vegetation [30]. To define the variables to be used for pixel classification into HDZs, we considered the vegetation indices which are most informative about crop condition and can be calculated from Sentinel-1 and -2 satellite missions' data. For this purpose, five microwave indices (calculated with the S1 data set) and five multispectral indices (calculated with the S2 data set) were evaluated. Subsequently, one index was selected for each signal type and named I_{S1} and I_{S2}.

The microwave indices considered were:

- Dual-Pol Diagonal Distance (DPDD) [6]:

$$\text{DPDD} = \frac{\left(\sigma_{vh(i)} - \sigma_{vv(i)}\right)}{\sqrt{2}} \tag{1}$$

where $\sigma_{vv(i)}$ is the value of pixel i in the co-polarized signal, and $\sigma_{vh(i)}$ the value of pixel i in the signal with cross-polarization. A low DPDD value represents bare soil, increasing gradually as biomass increases, and reaching maximum values with thick vegetation.

- Inverse Dual-Pol Diagonal Distance (IDPDD) [6]:

$$\text{IDPDD} = \frac{\left(\sigma_{vv(max)} - \sigma_{vv(i)}\right) + \sigma_{vh(i)}}{\sqrt{2}} \tag{2}$$

where $\sigma_{vv(i)}$ and $\sigma_{vh(i)}$ are defined as in (1) and $\sigma_{vv(max)}$ is the highest value of the co-polarized signal observed in each image within the plot. IDPDD is better than DPDD to discriminate between vegetated pixels and pixels with bare soil but presents some degree of inaccuracy to differentiate vegetation from water bodies.

- Vertical Dual Depolarization Index (VDDPI) [6]:

$$\text{VDDPI} = \frac{\left(\sigma_{vh(i)} - \sigma_{vv(i)}\right)}{\sigma_{vv(i)}} \tag{3}$$

where $\sigma_{vv(i)}$ and $\sigma_{vh(i)}$ are defined as in (1). This index quantifies biomass by measuring the ratio between total polarized power and co-polarized power.

- Microwave Polarization Difference Index (MPDI) [31]:

$$\text{MPDI} = \frac{\left(\sigma_{vv(i)} - \sigma_{vh(i)}\right)}{\left(\sigma_{vv(i)} + \sigma_{vh(i)}\right)} \tag{4}$$

where $\sigma_{vv(i)}$ and $\sigma_{vh(i)}$ are defined as in (1). This index is particularly sensitive to surface roughness and soil moisture, both being considerably altered after a hailstorm.

- Dual Polarization SAR Vegetation Index (DPSVI) [6]:

$$\text{DPSVI} = \frac{\sigma_{vh(i)}\left[\left(\sigma_{vv(max)}\sigma_{vh(i)} - \sigma_{vv(i)}\sigma_{vh(i)} + \sigma_{vh(i)}^2\right) + \left(\sigma_{vv(max)}\sigma_{vv(i)} - \sigma_{vv(i)}^2 + \sigma_{vh(i)}\sigma_{vv(i)}\right)\right]}{\sqrt{2} * \sigma_{vv(i)}} \tag{5}$$

where $\sigma_{vv(i)}$ and $\sigma_{vh(i)}$ are defined as in (1), and $\sigma_{vv(max)}$ as in (2). DPSVI is a combination of VDDPI and IDPDD which keeps the advantages of both indices while allowing differentiating vegetation from water bodies.

The spectral indices considered were:

- Normalized Difference Vegetation Index (NDVI) [32]:

$$\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} \tag{6}$$

where NIR is the pixel value in the near-infrared range (B8); Red is the pixel value in the red range (B4). NDVI is associated with the amount of chlorophyll in the vegetation. High NDVI values correspond to areas with higher reflectance in the near-infrared spectrum, which correspond to thicker, healthier vegetation. This index reduces noises such as differing amounts of light, cloud shadows, atmospheric attenuation, or certain topographic variations [27]. However, it saturates with high biomass and is very sensitive to variations in soil underlying the canopy [33].

- Enhanced Vegetation Index (EVI) [34]:

$$EVI = 2.5 * \frac{(NIR - Red)}{(NIR + 6 * Red - 7.5 * Blue + 1)} \tag{7}$$

where NIR and Red are defined as in (6), and Blue is the pixel value in the blue range (B2). EVI performs better than NDVI in high biomass situations. It also has a better response to canopy structure variations such as leaf area index, canopy type, plant shape, and canopy architecture [35]. EVI requires a thorough previous removal of both external noise and sensor noise to estimate surface reflectances [34].

- Soil Adjusted Vegetation Index (SAVI) [33]:

$$SAVI = \frac{NIR - Red}{NIR + Red + 0.428} * (1 + 0.428) \tag{8}$$

where NIR and Red are defined as in (6). SAVI is derived from the NDVI formula, to which a coefficient is added for the correction of soil brightness in areas with low vegetative cover. This correction is important in open canopies where the underlying soil affects the reflectance recorded.

- Advanced Vegetation Index (AVI) [36]:

$$AVI = [NIR * (1 - Red) * (NIR - Red)]^{1/3} \tag{9}$$

where NIR and Red are defined as in (6). Like NDVI, AVI uses the red and near-infrared reflectances. However, since AVI uses a power degree of the infrared response, it is more sensitive to slight differences in canopy density [36].

- Normalized Pigment Chlorophyll Ratio Index (NPCRI) [37]:

$$NPCRI = \frac{Red - Blue}{Red + Blue} \tag{10}$$

where Red and Blue are defined as in (6) and (7). This index is sensitive to carotenoid: chlorophyll content ratios. Being especially useful for characterizing leaf senescence and fruit ripening, it also differentiates senescent crops from bare soil.

Indices I_{S1} and I_{S2} were selected by analyzing data from the hailstorm on February 10th, 2019, near Lozada town, province of Cordoba, Argentina. The analysis was performed in a 90-hectare soybean field (plot 79 in Table S1, hereafter referred to as model plot) in the R3 phenological stage according to the Fehr scale [38]. At the time of the hailstorm, the plot presented total and homogeneous crop cover. After the storm, damage ranged from null to total crop loss, with intermediate degrees of damage in some sectors.

All Sentinel-1 and -2 images containing the model plot were selected from 12 December 2018 (60 days before the storm) to 11 April 2019 (60 days after the storm). Each image was cut out to fit the plot shape, and two datasets were built: S1 with images from mission Sentinel-1, and S2 with images from Sentinel-2. Datasets were pre-processed as described in Section 2.3. In each dataset, the five vegetation indices for each signal were calculated for each image pixel. A graph was built for each index showing the time series of the model plot pixels. Standard deviations of index values were calculated for each image. Next, the maximum values of the standard deviations pre-storm $(Max(S_j)_{ant})$ and post-storm $(Max(S_j)_{post})$ were calculated for each of the ten vegetation indices I_j ($j = 1, \dots, 5$ with S1 dataset, and $j = 6, \dots, 10$ with S2 dataset). Then we selected the lowest quotient among the vegetation indices for each dataset S1 and S2, defined as I_{S1} and I_{S2} respectively and calculated as:

$$I_{S1} = arg\ min_{j=1,\dots,5}\left(\frac{Max(S_j)_{ant}}{Max(S_j)_{post}}\right) \ and \ I_{S2} = arg\ min_{j=6,\dots,10}\left(\frac{Max(S_j)_{ant}}{Max(S_j)_{post}}\right) \tag{11}$$

Thus, the microwave index and the spectral index with the lowest pre-storm backscatter compared to the post-storm backscatter were selected.

2.5. Algorithm for Identification of HDZs

An unsupervised learning algorithm is proposed to identify HDZs within a plot using Sentinel-1 and-2 satellite images. The steps of the algorithm are: (1) selection and pre-processing of satellite data to remove speckle noise, presence of clouds, and other atmospheric phenomena; (2) calculation of vegetation indices with the highest sensitivity to hail-caused damage to constructing the data matrix for pixel classification; (3) preparation of the variables for classification and data matrix construction; (4) unsupervised pixel classification into HDZs.

(1)   Selection and pre-processing of satellite data

For each plot, satellite images within the 60-day pre- and post-storm time window, as described in Section 2.2, were selected. Each image was cut out to fit the plot shapes and datasets S1 and S2 were assembled. Both datasets were pre-processed by correcting speckle on microwave signals and atmospheric interferences on the multispectral signal, as described in Section 2.3.

(2)   Calculation of vegetation indices

Indices I_{S1} and I_{S2} were calculated for each pixel of datasets S1 and S2 respectively, as described in Section 2.4.

(3)   Preparation of classification variables and construction of data matrix

To determine HDZs in a field with hail-damaged crops, pixels were classified using the data in Sentinel-1 and -2 images. The number of pixels depended on the size of the affected field, with analyzed fields having an average area of 60 hectares. Thus, for the average field size, 6000 pixels of 10 m resolution (100 m^2) each were analyzed.

The algorithm was expected to classify pixels into HDZs. Since the pixels in a given data matrix always corresponded to a single plot, they invariably represented a single crop and variety, sown on the same date. Classification into HDZs considered the pre- and post-hailstorm crop evolution. The resulting HDZs were expected to cluster plants with a similar response to the biophysical variables estimated by the indices, both prior to and after the hailstorm.

In order to capture the diversity and complexity of possible hail-induced crop changes, both microwave and multispectral signals were used. As mentioned in Section 1, microwave signals are more sensitive to changes in structure and geometry, while the multispectral signal is more sensitive to physiological changes manifesting themselves in canopy coloration.

Since the revisit time of the Sentinel 2 mission is shorter than that of Sentinel 1, Sentinel 2 will capture a larger number of images than Sentinel 1 (RADAR) for a given time window. For this reason, a linear interpolation of spectral index values was performed to estimate a value for each of the dates of SAR image capture. This allowed having the same amount of data for both microwave and spectral signals in the matrix, thus balancing the influence of each dataset. The procedure also provided multispectral images with temporal equidistance from each other, which allowed calculating the rates of index value changes.

The rates of change (derivative functions) were calculated by differentiating each of the time series obtained when calculating the different indices of vegetation in the plots. Differentiating an equispaced time series consists of subtracting the immediately preceding value from each value and, therefore, it served us to calculate the derivative of the function. Exploration of both series, the original and the derivative, is interesting since in certain cases the different groups may differ due to the indices observed in a given period, but in other cases due to their change ratios [39]. It was not possible to calculate the derivative of the first image due to a lack of previous data, so it was removed from the matrix after calculating the derivatives. Thus, the data matrix for each plot had as many rows as pixels, and the variables I_{S1}, I_{S2} and their corresponding derivatives (ΔI_{S1}, ΔI_{S2}) calculated for

each observation within the time window. Thus, if the sowing (harvesting) date was prior to (following) the time window (60 days pre- and post-storm date), 9 observations were obtained (4 pre-storm and 5 post-storm) for each variable (I_{S1}, I_{S2}, ΔI_{S1}, ΔI_{S2}), resulting in a maximum of 36 values in the vector for each pixel.

(4)　　Unsupervised Pixel Classification

HDZs will be defined using unsupervised machine learning. Unsupervised learning aims to find the underlying structure of a dataset and cluster its data by similarity. The methods most widely used include K-means [40], hierarchical clustering [41], and mixture modeling [42]. Hierarchical clustering requires calculating the distance matrix between all data, which increases processing costs considerably in the case of big datasets. Mixture modeling requires assuming restrictive hypotheses on the probability distribution of the clusters defined. We used K-means [40] because it is simple, easy to implement, and quick and cost-effective for computing big datasets [43]. K-means is an iterative clustering algorithm used for partitioning n observations (x_1, x_2, ... , x_n) into k groups ($S = \{S_1, S_2, ... , S_k\}$). Each observation is allocated to the cluster with the nearest mean value in order to minimize within-cluster variances, based on the following equation:

$$\underset{s}{arg\ min}\ \sum_{i=1}^{k} \sum_{\chi_j \in S_i} \parallel \chi_j - \mu_i \parallel^2 \tag{12}$$

where μ_i is the mean of observations in S_i.

This method, which has been proven to have an optimal performance in scientific research for analyzing data clusters, requires specifying the number of groups into which data will be classified. For HDZ detection, plot pixels were grouped into three categories ($k = 3$). In this sense, three classes were defined per plot because it is the number of homogeneous damage zones most frequently detected in the field using visual techniques; thus, it is validated by the expertise of claims adjusters seeking to simplify and standardize field surveys. The Euclidean distance was used to measure proximity to centroids. For different applications of the algorithm, this parameter can be modified to include other classification objectives.

The algorithm was programmed in Python 3 (https://www.python.org/ (accessed on 31 January 2021)), using the pandas, GeoPandas, ee, eeconvert and DateTime libraries.

2.6. Algorithm Validation with In Situ Data

The quality of clusters provided by the algorithm proposed in this work was validated by external data, different from the data used to determine HDZs in each plot. Even though no information was available about the actual HDZ to which each pixel belonged for any of the plots analyzed, we had in situ damage measurements for some spots.

Pixels for which in situ data were identified in each plot. These pixels constituted a sub-sample of the three HDZs identified with the algorithm, for which actual damage data was available from in situ measurements. To validate the quality of pixel clusters into HDZs, we compared the means of the damage percentages calculated with in situ data in the three HDZs identified by the algorithm. If the algorithm identified HDZs correctly, the mean actual damage percentage had to differ between the zones compared. Homogeneity between HDZs was tested using one-way ANOVA tests. Differences were considered statistically significant if p-value < 0.05.

3. Results and Discussion

### 3.1. Selection of Vegetation Indices

Hail damage detection using satellite data is based on the premise that strong winds, rainfall, and hail impacts on crop architecture may alter the physical characteristics of the canopy and underlying soil, thus changing the energy backscattered from the crop to active and passive sensors [44]. The most sensitive indices were selected using the model plot described in Section 2.4. Figure 2 shows a timeline with image capture dates from both

missions and some representative images of the plot. Storm damage was heterogeneous, with in situ assessed yield losses ranging from 0 to 100% along a latitudinal gradient with increasing intensity from north to south. Total (100%) damage is defined as areas with the complete destruction of plant tissue and in situ observation of senescent plant remains and bare soil.
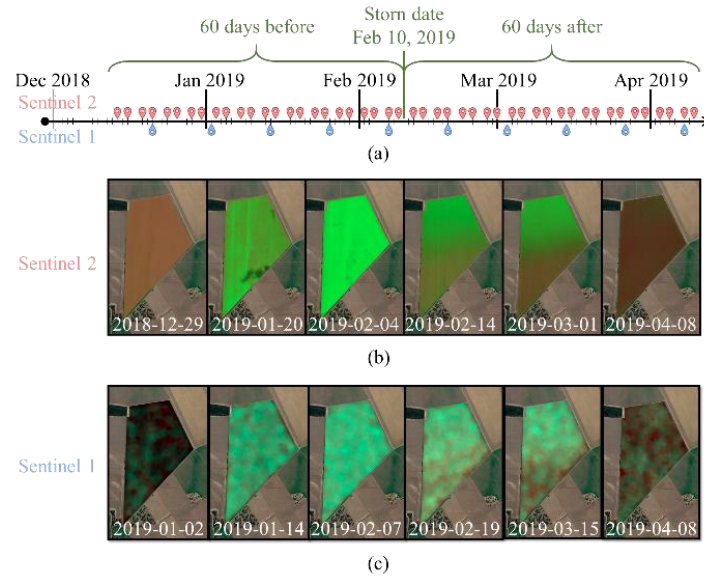
Figure 2. (**a**) Sentinel-1 and -2 images timeline for the model plot; (**b**) optical images without atmospheric interference, represented in RGB composition: 'B11', 'B8', and 'B4'; (**c**) despeckled SAR images, represented in RGB composition: 'VH', 'VV', and 'VV'.

The five microwave indices were sensitive to hail-caused damage in field pixels (Figure 3). The graphs highlight the behavior of pixels with no damage, the values of which were obtained by averaging 100 pixels from the north part of the field, where the crop was unaffected by the storm. Curves for the indices evaluated showed a homogeneous pre-storm behavior and split after the storm.
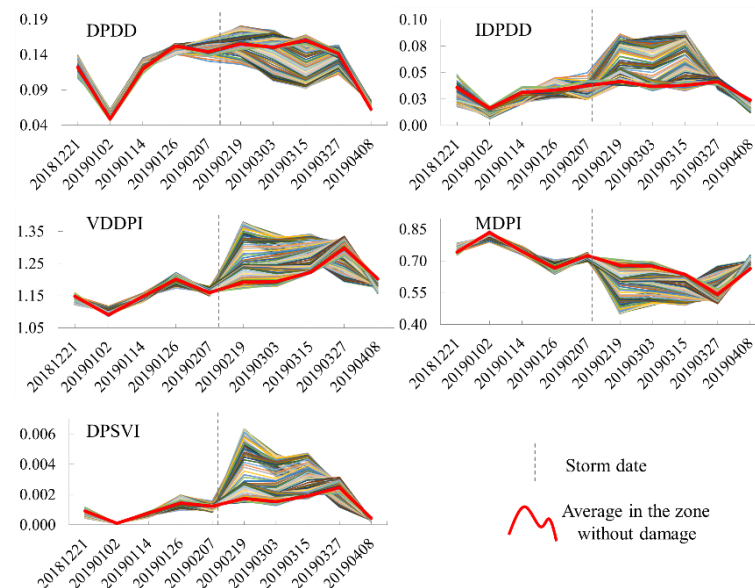
Figure 3. Microwave indices were calculated for the model plot pixels, before and after the hailstorm. Each line represents the index values for each pixel along time. The average values in the zone without damage were obtained by averaging 100 pixels from the northern part of the field, where the crop was unaffected by the storm.

Maximum pre- and post-storm standard deviations for each microwave index and their quotients are shown in Table 1. The microwave index selected for classification (I_{S1}) was DPSVI, the one with the lowest quotient value (Max_{ant}/Max_{post}), 0.1535 (Table 1). According to these results and considerations, it can be stated that DPSVI is of crucial importance to identify the scattering effects caused by hail on the observed crops.

Table 1. Pre- and post-storm maximum standard deviations of microwave indices.

|	DPDD	IDPDD	VDDPI	MPDI	DPSVI
Max_{ant}	0.0080	0.0071	0.0113	0.0158	0.0002
Max_{post}	0.0286	0.0205	0.0650	0.0794	0.0016
Max_{ant}/Max_{post}	0.2788	0.3488	0.1742	0.1986	0.1535

The DPSVI model is based on the scatter diagram pattern built by the backscatter coefficient of VV and VH polarizations and highlights the contribution of cross-polarization. The three parameters ($\sigma_{vv(i)}$, $\sigma_{vh(i)}$, and $\sigma_{vv(max)}$) in the equation for the DPSVI model (5) have a great influence on biomass characterization [6]. When the biomass production of a single crop increases, the cross-polarized backscatter values are higher [45]. The transmitted vertical polarization signal changes its polarization state partially or totally due to a change in the orientation angle caused by several factors such as the nature, topography, orientation, and distribution of the target. Thus, both rough surfaces and the presence of vegetation cause a significant depolarization of the illuminated signal, whereas smooth surfaces depolarize the signal to a lesser extent. Since C-band microwave signals penetrate through leaves and undergo multiple dispersion within small stalks of the crop, the signal is useful for assessing crop biomass [46]. C-band has a surface penetration ability of about 1 m, reaching saturation at around 60–70 tons/ha [47]. This might account for its good performance to detect hail-induced changes in the model plot.

Analogously, all spectral indices were sensitive to hail damage (Figure 4). The wavelength ranges used in the indices are sensitive to both healthy vegetation and bare soil. The values obtained with indices NDVI, EVI, SAVI, and AVI were positively correlated with the amount of healthy vegetation in the pixel -the higher the soil cover by vegetation growth, the higher the index value. NPCRI, on the other hand, had a negative correlation with healthy vegetation, showing lower and lower negative values as crop growth increased soil cover. It is worth pointing out that the highlighted curve for the sector with no damage had intermediate values in all indices before the hailstorm, becoming the most extreme curve after the storm with positive values in indices NDVI, EVI, SAVI, and AVI, and negative in NPCRI (Figure 4). This shows the sensitivity of spectral indices to changes in biomass amount in the pixel.

A comparison of the quotient between the pre- and post-storm maximum standard deviation for each spectral index (Table 2) showed that NPCRI had a remarkably better performance than all other indices (Max_{ant}/Max_{post} = 0.1483).

Table 2. Pre- and post-hailstorm maximum standard deviations for spectral indices.

|	NDVI	EVI	SAVI	AVI	NPCRI
Max_{ant}	0.0356	0.0618	0.0340	0.0232	0.0145
Max_{post}	0.1383	0.1813	0.1129	0.0743	0.0977
Max_{ant}/Max_{post}	0.2578	0.3406	0.3016	0.3121	0.1483
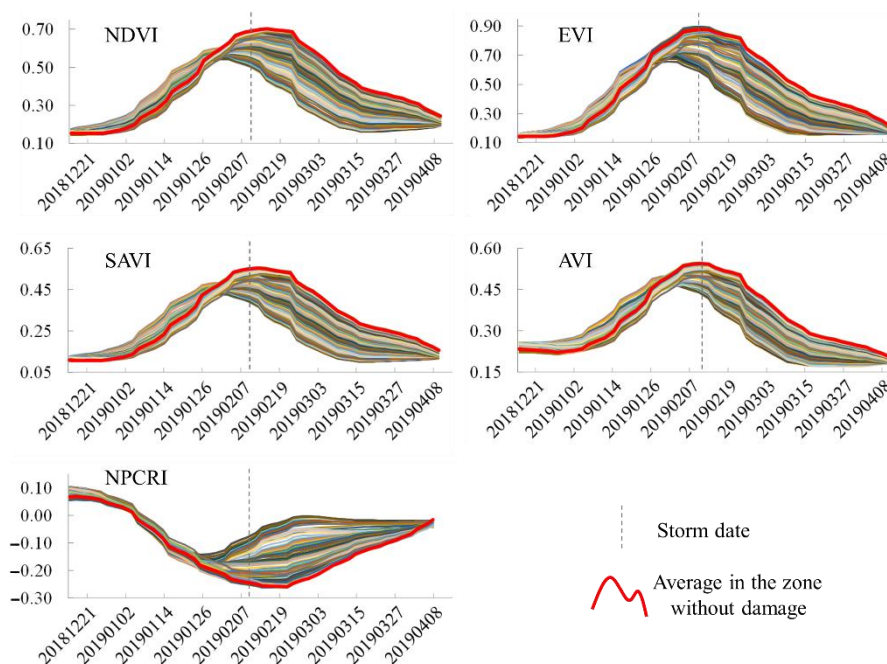
Figure 4. Spectral indices were calculated to plot the model pixels, before and after the hailstorm. Each line represents the index values for each pixel along time. The average values in the zone without damage were obtained by averaging 100 pixels from the northern part of the field, where the crop was unaffected by the storm.

This better performance might be due to the ability of NPCRI to differentiate between bare soil and healthy or senescent vegetation. NDVI is a normalized quotient of the near-infrared and red reflectance values, while EVI, SAVI, and AVI are equations also based on the vegetation response to these wavelengths. NPCRI is a normalized quotient of the red and blue reflectance values and takes positive values when soil predominates in the pixel (beginning of the curve) and close-to-zero negative values when senescent vegetation is predominant (end of curves). The other four indices show similar values at the beginning and end of curves, without discriminating bare soil predominance from senescent vegetation predominance. When hitting upon the canopy, hail destroys leaves and stems, leading to their senescence. Plant tissues undergo color changes due to modifications in content and relative proportions of pigments [48]. Changes in tissue color in senescent plants are mainly related to increasing chlorophyll degradation and a relative increase of carotenoids. Externally, color shifts from green to yellow, followed by brown when the tissue becomes necrotic. Carotenoids and chlorophylls exhibit strong absorption in the blue spectral range, the reflectance values used in NPCRI [37]. After a storm, when the crop begins to manifest hail-caused senescence, each pixel shows different fractions of bare soil and healthy or senescent vegetation. The ability of NPCRI to differentiate these components might explain its higher performance.

3.2. Algorithm Application for Identification of HDZs in the Model Plot

This section details the results of the steps followed by the algorithm applied to the model plot satellite data for post-storm HDZs identification. To corroborate how data pre-processing improves storm effect visualization, the vegetation indices were calculated with and without speckle filtering and atmospheric correction for each type of signal (Figure 5). Storm effects were difficult to observe in indices without the pre-processing step (Figure 5a) but easily observed in graphs of indices with pre-processed data (Figure 5b). Pre-processing completely removed speckle in the microwave signal and atmospheric interferences in the multispectral signal, without altering index sensitivity to hail-caused changes.
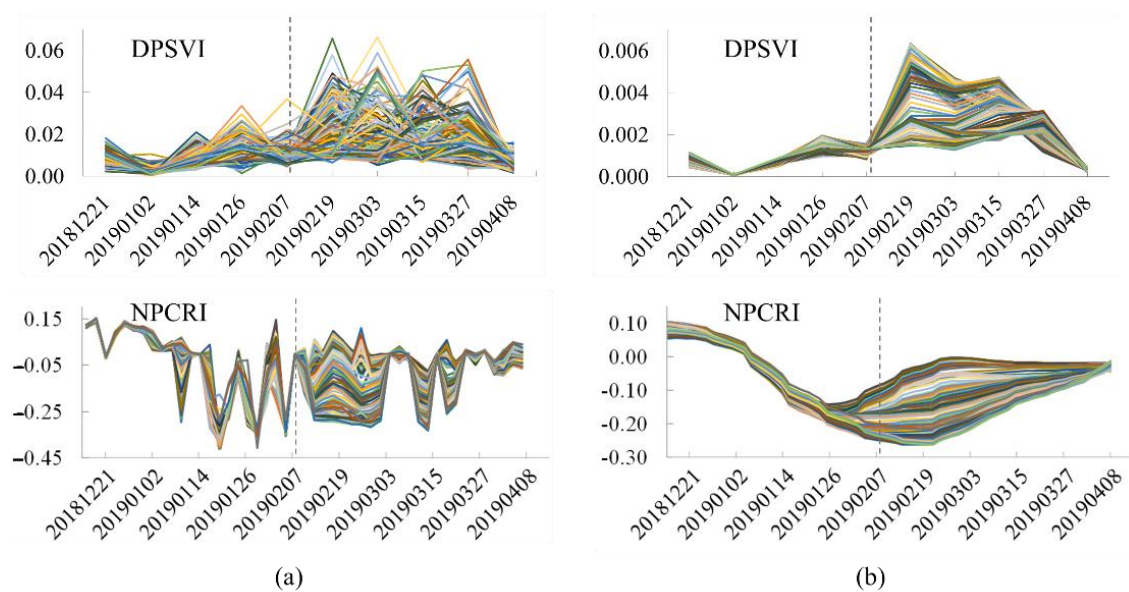
Figure 5. Pre-processing of satellite data for speckle removal in the microwave signal, and atmospheric interference removal in the multispectral signal. The time series for indices applied to model plot pixels without data pre-processing are plotted in (**a**), those with data pre-processing in (**b**).

The data matrix for the model plot was constructed with the four variables (indices DPSVI, NPCRI, and their corresponding derivatives), and the images recorded at 9 dates during each crop growing season. Figure 6a maps in grayscale the values of indices and their derivatives, and the time series of indices for each pixel are shown in Figure 6b. DSPVI had an immediate response to hail damage since the microwave signal is sensitive to changes in the physical properties, roughness, and geometry of a crop caused by hailstorms. NPCRI, on the other hand, had a gradual response along with the series. Changes in crop color, detected by the multispectral signal, occur gradually, as relative pigment amounts are modified by the destruction of plant structures. The DSPVI change rates confirmed these changes in the first post-storm image, whereas NPCRI change rates showed alterations in the pre-storm image due to the linear regression applied to the S2 collection as a smoothing procedure. This resulted in the visualization of storm-induced changes before the storm date. Since the K-means method does not take into account image date for image processing, this effect did not alter the result of classifications.

Next, the pixels in the model plot data matrix were subjected to unsupervised classification into HDZs using the K-Means method. The in situ damage assessments confirmed the good performance of the algorithm for detecting HDZs (Figure 6c).

To the best of our knowledge, the k-means method has not been applied to define HDZs in crops, but this technique has been widely used to identify homogeneous zone for site-specific management in precision agriculture [49]. A management zone is a sub-region of a field with relatively homogeneous soil and landscape attributes and responds in a similar way to certain management practices. The problems posed for management zone delimitation are comparable to those addressed in our study to detect HDZs. In the field of site-specific management, several authors have obtained satisfactory results using K-means to delimit homogeneous management zones [50–52].

3.3. Algorithm Validation with In Situ Data

The results of algorithm application in each of the 91 plots for which in situ data were available are shown in Table 3. The *p*-values obtained with the one-way ANOVA test in each plot are shown in Table S1. In situ data of hail damage in corn showed that in 66.67% of plots (15 out of 5) there was significant evidence of differences in mean damage percentage between the three HDZs determined by the algorithm. For wheat, mean damage differed between HDZs in 78.13% of the plots (25 out of 32 analyzed), while in soybeans, damage
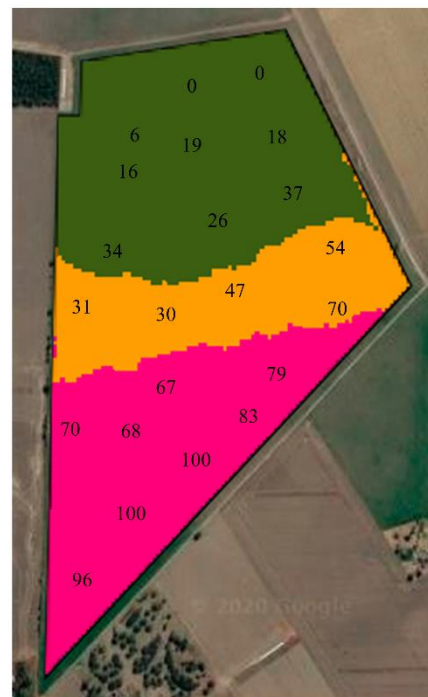
differences between algorithm-determined HDZs were confirmed in 72.70% of plots (32 out of 44).
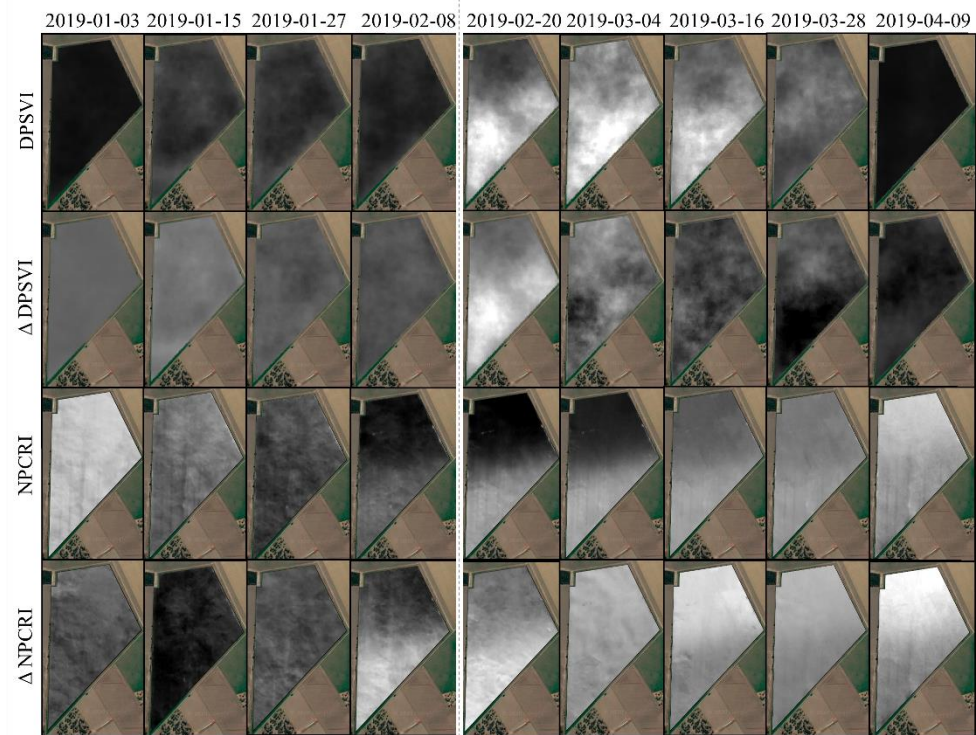
(a)
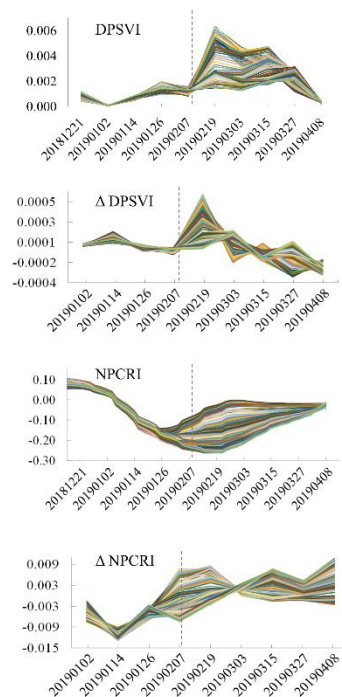
(b)

(c)

Figure 6. Model plot data matrix, with data prepared for classification shown in a map (**a**) and as a graph (**b**). In Figure (**c**), different colors in the plot model indicate HDZs classified by the algorithm, and the numbers show damage percentages recorded in situ and used for algorithm validation.

Table 3. Algorithm validation comparing the means of damage percentages obtained from in situ assessments for each HDZ. Plots with correct HDZ determination (those with significant differences between HDZs) and plots without significant differences between HDZs are shown in columns 2 and 3. Differences between HDZs were calculated using one-way ANOVA with a *p*-value < 0.05.

Crop	Plots Analyzed	Plots with Correct HDZ Determination	Plots with Non-Significant Differences between HDZs
Corn	15	10	5
Wheat	32	25	7
Soybeans	44	32	12
Total	91	67	24

The ANOVA test is frequently used to evaluate the performance of clustering methods [53] and previous researchers [49,54] have reported similar results in validating homogeneous management zones determination. Cluster validation by analysis of variance is demanding since the presence of outliers might hide significant differences between groups; however, the method is highly reliable when it identifies significant differences.

We have identified some particular situations in plots with no significant evidence of differences in degrees of crop damage between the three HDZs defined by the algorithm. In six plots, damage values were highly similar over the whole surface area. In such situations, the k-means algorithm is forced to perform artificial discrimination between pixels, since it works on the assumption of the existence of three different pixel clusters. Thus, no differences in mean damage values were found between HDZs, not due to a wrong classification but to the fact that all the plot pixels belonged to a single HDZ category. Also, no significant evidence of differences in crop damage between HDZs was found in 8 plots which presented highly heterogeneous vegetation conditions before the hailstorm. Plots with excess surface water during the growing season were a clear example of this anomaly. In this case, hail damage differences between HDZs were less conspicuous than the pre-storm crop differences.

If these 14 plots presenting particular situations are excluded, the overall result of algorithm validation amounts to 87.01%, making it very reliable for the detection of areas with similar degrees of damage. The tool is expected to increase the accuracy and fairness of damage adjustments for all the parties involved in the insurance contract: insurance and reinsurance companies, and policyholders.

4. Conclusions

In this research, we have proposed an algorithm for the automatic determination of crop damage by hail, using satellite data from Sentinel-1 and -2 missions. The algorithm selected the satellite data to be used based on the information recorded for each plot (sowing and harvest dates, storm date, and plot geometry), then performed a pre-processing step by applying morphological filters to microwave data, and linear regressions for atmospheric correction of multispectral data time series. The efficiency of data pre-processing was confirmed in the model plot. The next step consisted in calculating indices with sensitivity to hail damage: DPSVI for the microwave signal, and NPCRI for the multispectral signal. After setting the classification variables, a data matrix was constructed by calculating the change rates of index values to enhance the classifications. HDZ determination with the K-means method had excellent results in the model plot, in which in situ assessments were intensified. Overall validation showed significant evidence of differences in mean damage values between algorithm-determined HDZs in 87.01% of the plots analyzed. The proposed methodology provides transparency to HDZ determination and improves the scope of damage assessment since remote sensors allow sampling of entire plots. Thus, the method avoids the need to extrapolate the scarce in situ observations performed by insurance claims adjusters. Future extensions of the algorithm should adapt index selection

to each particular crop and its phenological stages, and also include the different soil types of the region where the crop is grown.

References

1. Halsnæs, K.; Larsen, M.A.D.; Kaspersen, P.S. Climate Change Risks for Severe Storms in Developing Countries in the Context of Poverty and Inequality in Cambodia. *Nat. Hazards* **2018**, *94*, 261–278. [CrossRef]
2. Barros, V.R.; Boninsegna, J.A.; Camilloni, I.A.; Chidiak, M.; Magrín, G.O.; Rusticucci, M. Climate Change in Argentina: Trends, Projections, Impacts and Adaptation. *Wiley Interdiscip. Rev. Clim. Chang.* **2015**, *6*, 151–169. [CrossRef]
3. Attema, E.P.W.; Ulaby, F.T. Vegetation Modelled as a Water Cloud. *Radio Sci.* **1978**, *13*, 357–364. [CrossRef]
4. Arciniegas, G.A.; Bijker, W.; Kerle, N.; Tolpekin, V.A. Coherence- and Amplitude-Based Analysis of Seismogenic Damage in Bam, Iran, Using ENVISAT ASAR Data. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1571–1581. [CrossRef]
5. Karam, M.A.; Member, S.; Fung, A.K.; Lang, R.H.; Chauhan, N.S. Model for Layered Vegetation. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 767–783. [CrossRef]
6. Periasamy, S. Significance of Dual Polarimetric Synthetic Aperture Radar in Biomass Retrieval: An Attempt on Sentinel-1. *Remote Sens. Environ.* **2018**, *217*, 537–549. [CrossRef]
7. Kim, Y.; van Zyl, J. A Time-Series Approach to Estimate Soil Moisture Using Polarimetric Radar Data. *IEEE Trans. Geosci. Remote. Sens.* **2009**, *47*, 2519–2527.
8. Huang, Y.; Walker, J.P.; Gao, Y.; Wu, X.; Monerris, A. Estimation of Vegetation Water Content from the Radar Vegetation Index at L-Band. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 981–989. [CrossRef]
9. Sahadevan, D.K.; Sitiraju, S.; Sharma, J. Radar Vegetation Index as an Alternative to NDVI for Monitoring of Soyabean and Cotton. In Proceedings of the XXXIII INCA International Congress (Indian Cartographer), Jodhpur, India, 19–21 September 2013; pp. 91–96.
10. Kim, Y.; Jackson, T.; Bindlish, R.; Lee, H.; Hong, S. Radar Vegetation Index for Estimating the Vegetation Water Content of Rice and Soybean. *IEEE Geosci. Remote Sens. Lett.* **2012**, *9*, 564–568. [CrossRef]
11. Verrelst, J.; Camps-Valls, G.; Muñoz-Marí, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.P.W.; Moreno, J. Optical Remote Sensing and the Retrieval of Terrestrial Vegetation Bio-Geophysical Properties—A Review. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 273–290. [CrossRef]
12. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef]
13. Camps-Valls, G. Machine Learning in Remote Sensing Data Processing. In Proceedings of the 2009 IEEE International Workshop on Machine Learning for Signal Processing, Grenoble, France, 1–4 September 2009; pp. 1–6.
14. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
15. Chandler, O.; Apan, A.; Pullinger, R.; Bullen, K. Quantifying Hail Damage for Crop Loss Assessment: Techniques Using Remote Sensing and Geographic Information Systems. In Proceedings of the 11th Australasian Remote Sensing and Photogrammetry Conference (ARSPC 2002): Images to Information, Brisbane, Australia, 2–6 September 2002; pp. 412–421.

16. Bentley, M.L.; Mote, T.L.; Thebpanya, P. Using Landsat to Identify Thunderstorm Damage in Agricultural Regions. *Bull. Am. Meteorol. Soc.* **2002**, 363–376. [CrossRef]

17. Capellades, M.A.; Reigber, S.; Kunze, M. Storm Damage Assessment Support Service in the U.S. Corn Belt Using RapidEye Satellite Imagery. In *Remote Sensing for Agriculture, Ecosystems, and Hydrology XI*; Neale, C.M.U., Maltese, A., Eds.; SPIE: Bellingham, WA, USA, 2009; Volume 7472, pp. 56–69. [CrossRef]

18. de Leeuw, J.; Vrieling, A.; Shee, A.; Atzberger, C.; Hadgu, K.M.; Biradar, C.M.; Keah, H.; Turvey, C. The Potential and Uptake of Remote Sensing in Insurance: A Review. *Remote Sens.* **2014**, *6*, 10888–10912. [CrossRef]

19. Sierra, E.M.; Beltran, A.B.; Maio, S. Peligrosidad del granizo para los cereales en la región Pampeana. *Rev. Fac. Agron.* **1993**, *14*, 35–43.

20. Saluzzi, M.; Nuñez, J.M. Comportamiento de granizadas sobre diversas áreas cultivadas del país. *Geoacta* **1975**, *7*, 77–90.

21. USDA; FCIC. *Corn Loss Adjustment Standards Handbook*; USDA: Washington, DC, USA; FCIC: Washington, DC, USA, 2019.

22. USDA; FCIC. *Soybean Loss Adjustment Standards*; USDA: Washington, DC, USA; FCIC: Washington, DC, USA, 2019.

23. USDA; FCIC. *Small Grains Adjustment Standards*; USDA: Washington, DC, USA; FCIC: Washington, DC, USA, 2016.

24. Lee, J.S.; Jurkevich, I.; Dewaele, P.; Wambacq, P.; Oosterlinck, A. Speckle Filtering of Synthetic Aperture Radar Images: A Review. *Remote Sens. Rev.* **1994**, *8*, 313–340. [CrossRef]

25. Kher, A.R.; Mitra, S. Optimum Morphological Filtering to Remove Speckle Noise from SAR Images. In *Image Algebra and Morphological Image Processing IV*; Dougherty, E.R., Gader, P.D., Serra, J.C., Eds.; SPIE: Bellingham, WAS, USA, 1993; Volume 2030, pp. 97–108. [CrossRef]

26. Mlynarczuk, M.; Porzycka-Strzelczyk, S. Speckle filtering in SAR images using morphological filters. *Int. Multidiscip. Sci. GeoConference-SGEM* **2018**, *18*, 175–182. [CrossRef]

27. Zhu, Z.; Woodcock, C.E. Automated Cloud, Cloud Shadow, and Snow Detection in Multitemporal Landsat Data: An Algorithm Designed Specifically for Monitoring Land Cover Change. *Remote Sens. Environ.* **2014**, *152*, 217–234. [CrossRef]

28. Hagolle, O.; Huc, M.; Desjardins, C.; Auer, S.; Richter, R. MAJA ATBD Algorithm Theoretical Basis Document. *Development* **2017**, *1*, 1–39.

29. El Hajj, M.; Bégué, A.; Lafrance, B.; Hagolle, O.; Dedieu, G.; Rumeau, M. Relative Radiometric Normalization and Atmospheric Correction of a SPOT 5 Time Series. *Sensors* **2008**, *8*, 2774–2791. [CrossRef]

30. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the Radiometric and Biophysical Performance of the MODIS Vegetation Indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]

31. Becker, F.; Choudhury, B.J. Relative sensitivity of Normalized Difference Vegetation Index (NDVI) and Microwave Polarization Difference Index (MPDI) for vegetation and desertification monitoring. *Remote Sens. Environ.* **1988**, *24*, 297–311. [CrossRef]

32. Gitelson, A.A.; Merzlyak, M.N. Remote estimation of chlorophyll content in higher plant leaves. *Int. J. Remote Sens.* **1997**, *18*, 37–41. [CrossRef]

33. Huete, A.R. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

34. Liu, H.Q.; Huete, A. A feedback based modification of the NDVI to minimize canopy background and atmospheric noise. *IEEE Trans. Geosci. Remote Sens.* **1995**, *33*, 457–465. [CrossRef]

35. Gao, X.; Huete, A.R.; Ni, W.; Miura, T. Optical-biophysical relationships of vegetation spectra without background contamination. *Remote Sens. Environ.* **2000**, *74*, 609–620. [CrossRef]

36. Roy, P.S.; Sharma, K.P.; Jain, A. Stratification of density in dry deciduous forest using satellite remote sensing digital data—An approach based on spectral indices. *J. Biosci.* **1996**, *21*, 723–734. [CrossRef]

37. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiol. Plant.* **1999**, *106*, 135–141. [CrossRef]

38. Fehr, W.R.; Caviness, C.E. *Stages of Soybean Development*; Iowa State University. Agricultural and Home Economics Experiment Station: Ames, IA, USA, 1977; pp. 1–12.

39. Justel, A.; Svarc, M. A divisive clustering method for functional data with special consideration of outliers. *Adv. Data Anal. Classif.* **2017**, *12*, 637–656. [CrossRef]

40. Wong, M.A.; Hartigan, J.A. Algorithm as 136: A k-Means Clustering Algorithm. *J. R. Stat. Soc. Ser. C Appl. Stat.* **1979**, *28*, 100–108.

41. Basak, J.; Krishnapuram, R. Interpretable hierarchical clustering by constructing an unsupervised decision tree. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 121–132. [CrossRef]

42. Davis, R.L.; Greene, J.K.; Dou, F.; Jo, Y.K.; Chappell, T.M. A practical application of unsupervised machine learning for analyzing plant image data collected using unmanned aircraft systems. *Agronomy* **2020**, *10*, 633. [CrossRef]

43. Arthur, D.; Vassilvitskii, S. K-means++: The advantages of careful seeding. *Chem. Eng.* **2006**, *8*, 26–28.

44. Jedlovec, G.J.; Nair, U.; Haines, S.L. Detection of storm damage tracks with EOS data. *Weather. Forecast.* **2006**, *21*, 249–267. [CrossRef]

45. Chauhan, S.; Srivastava, H.S. Comparative evaluation of the sensitivity of multi-polarised SAR and optical data for various land cover. *Int. J. Adv. Remote Sens. GIS Geogr.* **2016**, *4*, 1–14.

46. Ghasemi, N.; Sahebi, M.R.; Mohammadzadeh, A. A review on biomass estimation methods using synthetic aperture radar data. *Int. J. Ofgeomatics Geosci.* **2011**, *1*, 776–788.

47. Nizalapur, V.; Jha, C.; Madugundu, R. Estimation of above ground biomass in Indian tropical forested area using multi-frequency DLR-ESAR data. *Int. J. Geomat. Geosci.* **2010**, *1*, 167–178.

48. Buchanan-Wollaston, V. The molecular biology of leaf senescence. *J. Exp. Bot.* **1997**, *48*, 181–199. [CrossRef]
49. Córdoba, M.; Bruno, C.; Costa, J.; Balzarini, M. Subfield management class delineation using cluster analysis from spatial principal components of soil variables. *Comput. Electron. Agric.* **2013**, *97*, 6–14. [CrossRef]
50. Arno, J.; Martinez-Casasnovas, J.A.; Ribes-Dasi, M.; Rosell, J.R. Clustering of grape yield maps to delineate site-specific management zones. *Span. J. Agric. Res.* **2011**, *9*, 721. [CrossRef]
51. Haghverdi, A.; Leib, B.G.; Washington-Allen, R.A.; Ayers, P.D.; Buschermohle, M.J. Perspectives on delineating management zones for variable rate irrigation. *Comput. Electron. Agric.* **2015**, *117*, 154–167. [CrossRef]
52. Ikenaga, S.; Inamura, T. Evaluation of Site-specific management zones on a farm with 124 contiguous small paddy fields in a multiple-cropping system. *Precis. Agric.* **2008**, *9*, 147–159. [CrossRef]
53. Souza, E.G.; Schenatto, K.; Bazzi, C.L. Creating thematic maps and management zones for agriculture fields. In Proceedings of the 14th International Conference on Precision Agriculture, Montreal, QC, Canada, 24–27 June 2018; pp. 1–17.
54. Bazzi, C.L.; Souza, E.G.; Uribe-Opazo, M.A.; Nóbrega, L.H.P.; Rocha, D.M. Management zones definition using soil chemical and physical attributes in a soybean area. *Eng. Agric.* **2013**, *33*, 952–964. [CrossRef]