forests

Article

Biomass assessment and carbon sequestration in post-fire shrublands by means of Sentinel-2 and Gaussian processes

David Vinué-Visús[1,*], Ricardo Ruiz-Peinado[2,3], David Fuente[4] , José Vicente Oliver-Villanueva[5], Eloína Coll-Aliaga[6] and Victoria Lerma-Arce[7]

[1] ITACA Research Institute. Universitat Politècnica de València; davivi1@upv.es
[2] Centro de Investigación Forestal (INIA, CSIC); ruizpein@inia.es
[3] iuFOR, Forest Management Research Institute, UVa-INIA
[4] CZECHGLOBE - Global Change Research Institute, CAS; dafueher@upv.es[5]ITACA Research Institute. Universitat Politècnica de València; vlerma@upv.es
[6] ITACA Research Institute. Universitat Politècnica de València; joolvil@upv.es
[7] ITACA Research Institute. Universitat Politècnica de València; ecoll@upv.es
* Correspondence: davivi1@upv.es; Universitat Politècnica de València; Camino de Vera s/n. 46022 Valencia.

Citation: Vinué-Visús, D.; Ruiz-Peinado, R.; Fuente, D.; Oliver-Villanueva, J.V.; Coll-Aliaga, E.; Lerma-Arce, V. Biomass assessment and carbon sequestration in post-fire shrublands by means of Sentinel-2 and Gaussian processes. *Forests* **2022**, *13*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

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

Abstract: Biomass and carbon stock in a post-fire area covered by a young oak coppice of *Quercus pyrenaica* Willd. associated with shrubs, mainly of *Cistus laurifolius* L., has been assessed. This area was burned during the forest fire event of Chequilla (Guadalajara, Spain) in 2012. Sentinel-2 imagery has been used together with own forest inventories in 2020 to assess the total biomass of the target area through machine learning methods. Inventory includes plots of total dry weight biomass ranging between 5 and 14 Mg*ha^{-1}, with individuals up to eight years old. Non-linear, non-parametric Gaussian process regression methods have been used to link reflectance values from Sentinel-2 imagery with total shrub biomass. With a short inventory of only 33 plots for up to 136 ha the total biomass can be assessed with a RMSE of 1.58 Mg*ha^{-1} and a bias of 0.05 achieving an error between 11.2 and 28.8% for the gathered biomass, which means a good assessment associated with a low forest inventory intensity and cost. Results are good enough to settle both forest monitoring and management, being the methodology affordable in terms of time, cost, efforts, and computing capacity.

Keywords: machine learning; remote sensing; Gaussian process regression; forest inventory

1. Introduction

Biomass inventories are the key to design and to assess policies concerning carbon stocks and carbon sequestration [1], including forest management proposals to maximize the protection and the dynamics of both the reservoirs and sinks. Consequently, a robust carbon accounting must be performed [2].

Due the recurrent fire events in the Iberian Peninsula, it is of special interest to assess the carbon recovery after fires due natural vegetation recovery. After a fire, vegetation is sometimes recovered as a shrubland structure and shrub biomass quantification has been revealed to be costly in terms of time and efforts although some standard methodologies have been proposed for its estimation [3].

In addition, the assessment of shrub biomass is necessary in forest fire prevention to avoid wildfires, as they are a significative part of the so-called fuel models [4], that are a mere collection of vegetative features giving wide intervals of biomass values [5]. Consequently, they cannot be used for precision counting of biomass and carbon.

Shrub biomass of woody plants has been also studied by means of biophysical values, as leaf area index (LAI). As an example, for stands of *Pinus halepensis* Mill. in Spain, the relationship between LAI and total biomass displays a decreasing trend, highly remarked with age [6]. However, the extended period needed to find significant variances make this method not suitable for a temporal analysis nor for affordable cost and time spend.

About field sampling, it is a fact that biomass assessment, and therefore carbon stocks, can be achieved at a large scale by traditional forest inventories, such as the National Forest Inventory (NFI) in Spain [7], using biomass models. However, the inventoried areas and the gap between plots in the NFI (typically, 1 Km) are rather large since both are designed for a global assessment. Moreover, the time between consecutive field works, i.e., actual lag accounts for more than 14 years for most of the Spanish territory, is also another limitation. Thus, there is a lack of accurate information of specific areas of interest.

,Large areas of shrublands after fire events have been studied with a general classification regardless of the shrubland composition [8].

Indeed, shrub biomass models are not usually available in literature. In Spain, for example, where shrublands occupy a notable area, there is little information for biomass estimation at formation level [3,9].

Recently, biomass equations based on vegetation soil cover and height measurements are available for most of the shrub structures in Spain [10]. These equations provide an accurate quantification of biomass for each shrub structure and can give an advantage for biomass assessment if land use is segmented. This allows the application of specific equations to related structures since a forest fire promote inhomogeneous vegetation recovery for the whole affected area. Nevertheless, not all the shrub structures are often included, and additional efforts must be made to cover all shrub formations [10].

From a remote sensing point of view, this ground truth is essential: although the relationship between optical reflectance and biomass is well-known, it is not possible to directly assess biomass from remote sensing imagery without linking this information with field data [11,12]. This means that for any study, previous field data for the purposes of calibration and validation must be sampled as reflectance is ever dependent on these field values.

Notwithstanding the origin of the remote sensing data, both from passive or active sensors, several methodologies for data processing have been tested, including multiple regression analysis, k-Nearest Neighbors (k-NN) and neural networks, among others [13]. Those parametric regression methods have achieved accuracy values of R^2 from 0.56 to 0.74. In these approaches a direct relationship (linear, quadratic, multiple) between reflectance and the observed variable is assumed and multispectral imagery over different kind of forests is applied [14]. But all these studies were performed over homogeneous mature forest and never over shrublands.

Other empirical methods are based on non-parametric regression functions without any explicit assumption about variable dependences neither data distribution. These methods avoid carrying out previous spectral band relationships, transformations, or fitting functions [15].

Alternatively, empirical methods are limited inside the range of values of the training dataset and therefore, it is hard to extrapolate the results to other conditions or biomes [16]. However, some of the non-parametric regression methods have demonstrated their capabilities to be adapted to remote sensing studies. Among them, machine learning non-linear regression algorithms such as artificial neural networks (ANN), support vector machine (SVM) or Gaussian process regression (GPR) have been efficiently applied for biophysical variables assessment from Earth observation data [17,18].

GPR is a collection of finite random variables with a multivariate normal distribution [19]. These processes are related to a collection of indexed random variables that can be defined through a shared density of probability, typically a Gaussian distribution. Application of GPR for biomass assessment and shrubland studies have not been deeply tested up until now.

Previously described developments use a wide range of values and field data, and no information about the number of samples needed for a good performance can be taken as a reference.

Beyond the aprioristic analysis of fire consequences and the subsequent forest recovery, it is fundamental to accurately estimate carbon stocks for the implementation of measures that mitigate climate change. Therefore, a proper evaluation of the biomass in a post-fire shrub structure is highly desirable. Moreover, it is also fundamental to properly assess the carbon sequestration associated with the recovery of vegetation in a quantitative form improving the methodologies based on traditional optical indexes.

Linear regression has not been settled as a trustable method for biomass assessment, not even in post-fire shrubs conditions. Thus, a study concerning non-linear methods applied to Copernicus imagery from Sentinel constellation could be of potential interest for assessing biomass.

The overall purpose of this contribution is the biomass assessment of the post-fire recovered area that experienced a wildfire in 2012 in Chequilla (Guadalajara, Spain). In summary, the main specific objectives are the following: (i) to define an algorithm using an empirical non-parametric method based on GPR machine learning techniques to be applied over Sentinel-2 images; (ii) to evaluate the specific algorithm regarding its capability to assess the total biomass of a characteristic post-fire plant association; (iii) to assess the carbon sequestration derived from post-fire vegetation regeneration in the area over a period of eight years and finally (iv) to obtain a map of biomass and carbon stock over the study area with a spatial resolution of 10 m based on Sentinel-2 imagery.

2. Materials and Methods

2.1. Study area

In the province of Guadalajara (Spain), in the middle of the Iberian Peninsula, the burnt area affected by a forest fire in the surroundings of the Municipalities of Chequilla, Checa, Alcoroches and Traid has been studied.

The fire started on 1st August 2012 and finished on 4th August 2012 affecting up to 1,151 ha, of which 788 ha of them were included in the forest category mainly dominated by *Pinus sylvestris* L. with some individuals of *Pinus nigra* Arnold. and *Quercus pyrenaica* Willd. in the understory.

This area has a typical temperate oceanic climate, *Cfb* by Köppen, without dry season and a warm summer where the average temperature during the coldest month is barely below 0 °C, the annual average temperature is above 22 °C and up to seven months average above 10 °C. Chequilla records an average annual temperature of 9.1 ºC, a total annual rain of 703 mm and lies at 1,355 m.a.s.l. The soils are mainly schist or slate-based schist with some areas dominated by sandstone and quartzite.

In 2020 the burnt area was naturally regenerated only in 136.52 ha composed by *Quercus pyrenaica* Willd. as coppice and shrubs of *Cistus laurifolius* L. and *Genista florida* L., which is a typical association of degraded forest according to Ruiz de la Torre [20]. This regenerated areaof young oak coppice covering 136.52 ha is our object of study.

The center of the area of study is located at [604700, 4495300] coordinates referenced to the UTM 30TWL zone over datum WGS-84, corresponding to 40° 36′ 7′′N, 1° 45′ 45′′E. Neighbor to this area remain significant forest structures corresponding to *Pinus sylvestris* L. (DGCN, 2006), burned areas and agrarian and urban settlements (Figure 1).

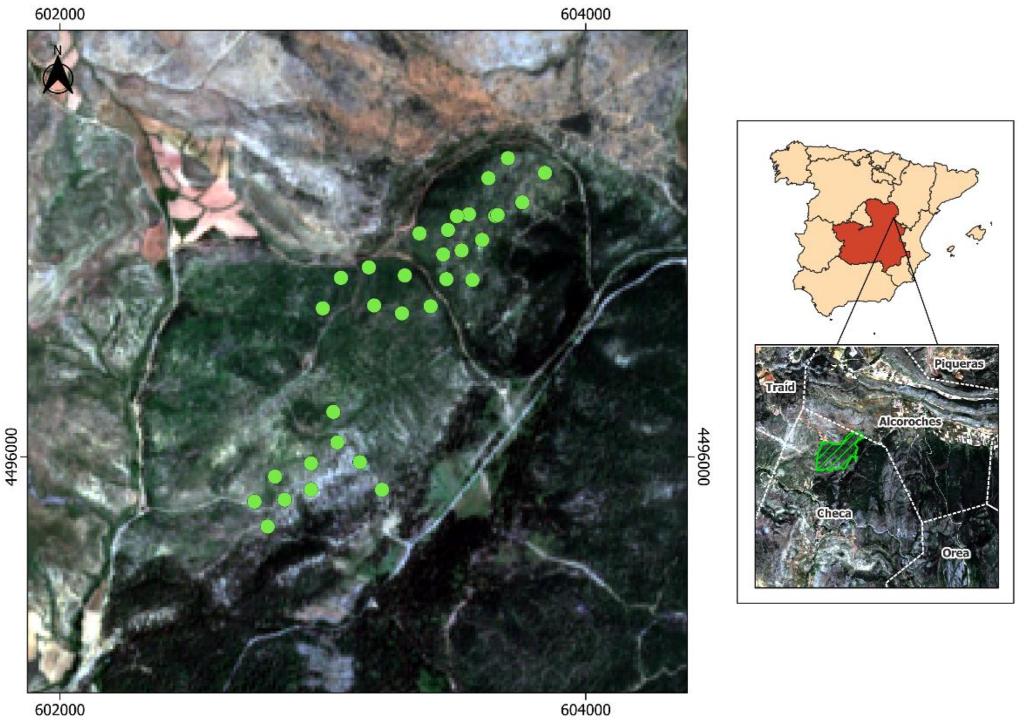


Figure 1. Study area and location of field plots to estimate shrub biomass in the burnt area (Chequilla, Spain). 2020.

2.2. Satellite imagery

To avoid the influence of illumination effects related to geometry and acquisition angles on reflectance values, only the closest image to the summer solstice of 2020 has been used. Sentinel-2A image has been downloaded in a Level2A reflectance (at a bottom of atmosphere) [21] and any further post-processing of the image have been performed. For the study area, a split area of 10 × 10 km has been used.

The image acquisition is related to the field sampling data (Table 1). Considering that annual biomass growth is gathered in spring and summer months and assuming only a unique total annual rate (as a discrete value, not as a continue increase), just a single satellite image can be used to obtain the final biomass map.

Table 1. Sentinel-2 imagery data

Mission/ Sensor	Level	Acquisition date	Reference	UTM	Azimuth angle	Elevation angle
Sentinel 2 /MSI	L2A	2020/06/22	R094	N T30TWL	139.09	21.8433

2.3. Sampling method and evaluation

Concerning sampling methods in shrublands, many methodologies have been suggested. For example, the line intercept method developed by Canfield [22] is a technique that was employed to estimate cover in the grasslands of southwestern USA. It is the most used and adapted for rangeland inventory and monitoring applications, including the assessment of biomass in burned areas in a shrub-grassland mosaic [23]. However, due the aim of the actual research of reducing the sampling cost and maximizing the representativeness of the data base, a random inventory has been developed.

Plots have been designed as a circle of 15m of radius and have been selected in a random way along the transects ensuring a minimum distance between them of at least 100m to avoid overlapping of the Sentinel-2 pixels related to the plot.

For each one of the 33 plots sampled the species composition was checked to ensure that it belongs to the shrub formation composed by *Quercus pyrenaica* and *Cistus laurifolius* and soil cover and mean height of the structure was measured. The field campaign was performed between 22nd and 24th June 2020.

To better ensure the reliability of the sample selection procedure, Monte-Carlo method was applied [24]. This method proceeds with a data analysis based on random sampling to generate the values for the probabilistic components and compares the sampling with simulated databases to evaluate the representativeness of the sampling, regardless the small number of plots sampled. Then, this method combines statistical concepts as random sampling with the generation of pseudo-random numbers, and it is based on systematic sampling of random variables.

An evaluation of the radiometric value distribution was performed following the Monte-Carlo procedure, comparing the NDVI distribution values of inventoried plots against the NDVI distribution over the whole site (10 × 10 km^2). As the number of plot pixels and the area of study are completely different, it is not statistically consistent to compare two NDVI histograms. Then, a comparison between the NDVI cumulative frequencies of both distributions was carried out to benchmark the actual frequency to a randomly shifted sampling pattern.

Thanks to this method (i) a cumulative frequency of the N pixel of NDVI values corresponding to the sampling is computed. Afterwards, (ii) the cumulative frequency of NDVI on a randomly shifted sampling design is computed and (iii) step (ii) is repeated 199 times with 199 different random translation vectors.

This method provides a total population of N = 199 + 1 (actual) cumulative frequency on which a statistical test with a significance level of 0.05 is applied: for a given NDVI level, if the actual plot density function is located between two limits defined by the five highest and five lowest values of the 200 cumulative frequencies, the null hypothesis is accepted, otherwise rejected. In the first case, both NDVI distributions are supposed to be equivalent, assuming those NDVI distributions between the area of study and the sampling plots.

*2.4. Allometric equations*Allometric equation for shrub formation composed by *Quercus pyrenaica* and *Cistus laurifolius* has been applied [25] (Equation 1).

$$\text{Biomass (Mg ha}^{-1}\text{)} = 0.2693*(\text{soil cover}*\text{mean height})^{0.5232} \quad (1)$$

Where soil cover of the vegetation in percentage (%) and mean height of the formation in cm.

2.5. Carbon sequestration assessment

The total biomass stock recovered after a forest fire will be considered as a carbon sequestration at CO_2 equivalent form ($CO_{2\,eq.}$). First, a carbon fraction of dry biomass of 0.4799 has been calculated according to previous studies [26,27]. Then, the molecular weight of the CO_2 and the stoichiometric relationship between carbon and CO_2 have been considered to achieve the total amount of $CO_{2\,eq.}$ This approach shows a transforming coefficient from biomass to $CO_{2\,eq.}$ of 1.759.

2.6. Data analysis with Gaussian processes regression (GPR)

Gaussian process regression (GPR) models are non-parametric kernel-based probabilistic models. Meanwhile linear regressions estimate the error from the database itself, a GPR generates a response from an interval of variables of the training data and a new vector as an input by introducing variables from a Gaussian process.

The main assumption is that a set of random variables at any finite combination of them are distributed along the Gaussian curve and then any number of observed variables will be distributed as Gaussian as well. Consequently, a GPR model is a probabilistic method and thus, it is possible to predict the outcome intervals from training models.

Moreover, the results are not fitting a line of responses but lie over a probability interval [19].

The covariance function between input and output variables shows the similarity between them. This kernel function and its mean (adjusted to zero for simplify) define the [28]. In this study, a kernel function with separated length scale for each predictor is applied over the data, where an automatic relevance determination (ARD) method is used to order the inputs according to their importance [29].

GPR is a very suitable method for remote sensing analysis as it is not limited by the large number of parameters needed for the implementation of methods like neural networks [30] and its computation requirements are less demanding than those one based on pixel-by-pixel inversion methods [15]. However, GPR has not been tested yet on areas requiring a huge effort of inventory to simplify the field data.

GPR computation [31] was trained with Sentinel-2 spectral bands 2 to 8A (490 to 865 nm), both included, and SWIR bands 11 (1610 nm) and 12 (2190 nm) as inputs, whereas total biomass represents the output. Any additional synthetic band is created because non-parametric methods can extract all the relevant information of the bands without user intervention.

Field data was randomly split in six subsets with well-distributed values of biomass.

Moreover, five subsets were used for algorithm training in each iteration and one subset was utilized for the result validation. This procedure is done six times in total so that each subset is used once as a validation group during the statistical analysis. The obtained estimations and corresponding errors result from the average of the six iterations. The aim of the iterations is the generation of every possible combination between training and validating subsets to make the validation more robust by using all the subsets for this purpose. The performance has been evaluated with the absolute mean square root error (RMSE) and the coefficient of determination (R^2) as overall indicators of accuracy.

. To understand the fit between observed and predicted values, a major axis regression (MAR) line corresponding to the well-adjusted slope has been included [32]. Although it has been demonstrated that for adult forest structures it is necessary to split the reflectance values according to their features [33] the post-fire regenerated shrubs structures have been tested as a single class.

3. Results

3.1. Inventory and database

After the random application of the sampling method, a great variability of biomass including stories and weight has been observed (Figure 2 and Table 2) to characterize a post-fire shrub structure with just a limited number of plots (33 sampled plots, plus 5 non-forested soil plots). $CO_{2\,eq.}$ has been calculated for the database (Table 2).

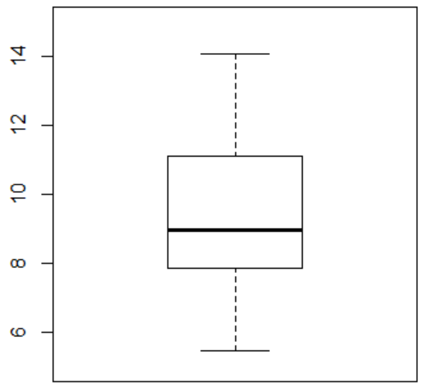


Figure 2. Biomass sampling of *Quercus pyrenaica* Willd. in Chequilla (Mg*ha^{-1}), 2020.

All 33 plots range a fractional cover between 50% and 95% with 25 plots with at least 70% of fractional cover. Height of *Quercus pyrenaica* Willd. ranges between 0.6 and 1.9 m and the plot area with a radius of 15 m is shared with *Cistus laurifolius* L. and *Genista florida* L., but normally with a lower height than that of *Quercus pyrenaica* Willd.

Table 2. Statistics of the shrub biomass values (Mg*ha^{-1}) and its correspondent CO_2 equivalent (Mg*ha^{-1}) for Chequilla study area. 2020.

Variable	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Biomass (Mg*ha^{-1})	5.480	7.850	8.980	9.367	11.110	14.040
$CO_{2\ eq.}$ (Mg*ha^{-1})	9.639	13.808	15.795	16.476	19.542	24.696

For the shrub sampling, the mean (9.367 Mg*ha^{-1}) is slightly higher than the median value (8.980 Mg*ha^{-1}) and the difference between median and the third quartile (11.110 Mg*ha^{-1}) is higher than the difference between the median and the first quartile (7.850 Mg*ha-1). As a result, the database displays a slightly positive skewed distribution, but close to a normal one. The range lies between 5.480 and 14.040 Mg*ha^{-1}, that is, it is not especially wide for a shrub structure of the same age. In addition, remarkable differences cannot be found by a visual inspection. However, even with a minor asymmetry in the distribution of the biomass, those values and ranges have been also observed by other authors [12], and consequently sampling can be accepted as adequate for this study.

For calibration purposes and to include the maximum variability of reflectance values in the database, six plots of bare soil have been randomly selected. Those plots range a maximum NDVI of 0.11, which ensures the absence of significant vegetation.

3.2. Field samplig evaluation

Once biomass values are well distributed, an evaluation of the radiometric value distribution has been done comparing the NDVI distribution values of inventoried plots against the NDVI distribution over the whole scene by means of the Monte-Carlo procedure (Figure 3). For a methodological demonstration, a global area of 10 × 10 km^2 has been selected, including the burned area. Nevertheless, this burned area represents only around 5% of the total territory included in the study area and a wide surrounding area that was not burned.

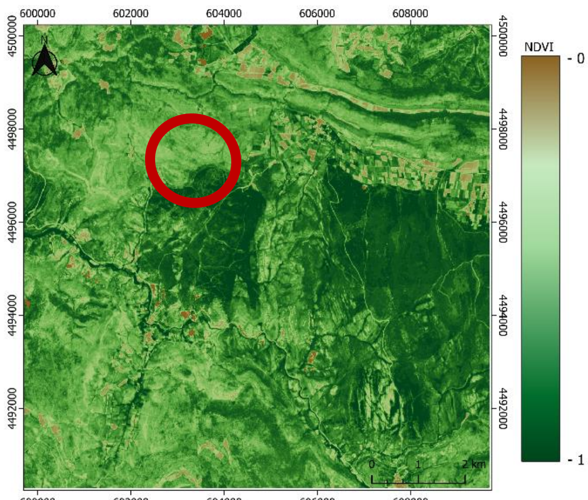


Figure 3. NDVI distribution in Chequilla, Spain. 23/062020. In the red cercle, the study area.

Monte-Carlo procedure has been performed to avoid any extreme value neither outlier in the sampled plots and to confirm that reflectance values included will not promote any bias and to confirm that no imagery processing issues are include in the database

The figure 4 displays that plotted values (green dots) are rather representative for the other pixels in the whole area and any bias is forced. After iterating groups of random pixels against sampled pixels it is noticed that the NDVI values distribution for the field sampling is in accordance with those values of the study area. Moreover, real and simulated distributions are equivalent as shown in figure 4, ranging the NDVI values of the plots between 0.44 and 0.80. It is assumed that lower values are not related to forest due the high soil cover fraction values of the plots (0.5 to 0.95).

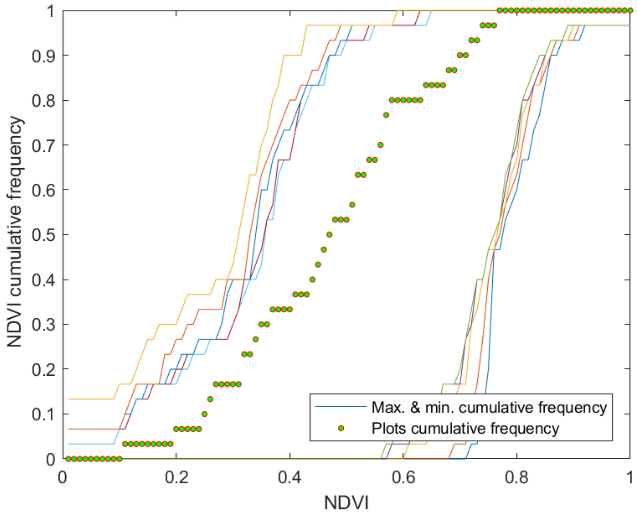


Figure 4. Comparison of NDVI distribution between study area (green dots) and over the whole image (5 highest and lowest values as external lines). Area of the Chequilla fire, June 2020.

This wide area of 10 × 10 km² clearly shows that NDVI values are very similar for both burnt and recovered areas and for other non-disturbed forest parts.

3.3. Biomass and carbon assessment

Figure 5 shows the scatterplot between the best fitted model obtained by GPR and biomass ground data for the oak coppice forest of *Q. pyrenaica* Willd. with other shrub species. Using all these points for training a GPR with Sentinel-2 data, the R^2 between those estimated values against field data can account up to 88% of the variability of the forest.

The error, evaluated as root mean square error (RMSE) reaches 1.58 Mg*ha^{-1} for the gathered biomass range of 5.480 and 14.040 Mg*ha^{-1}, representing a maximum of 28.8 % and a minimum of 11.2 % of error in the biomass assessment produced for both over- and under estimation.

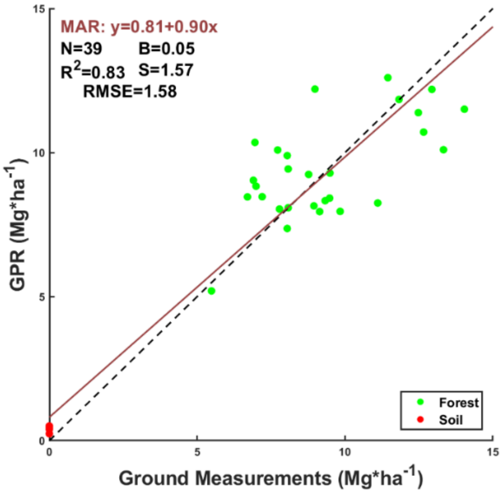


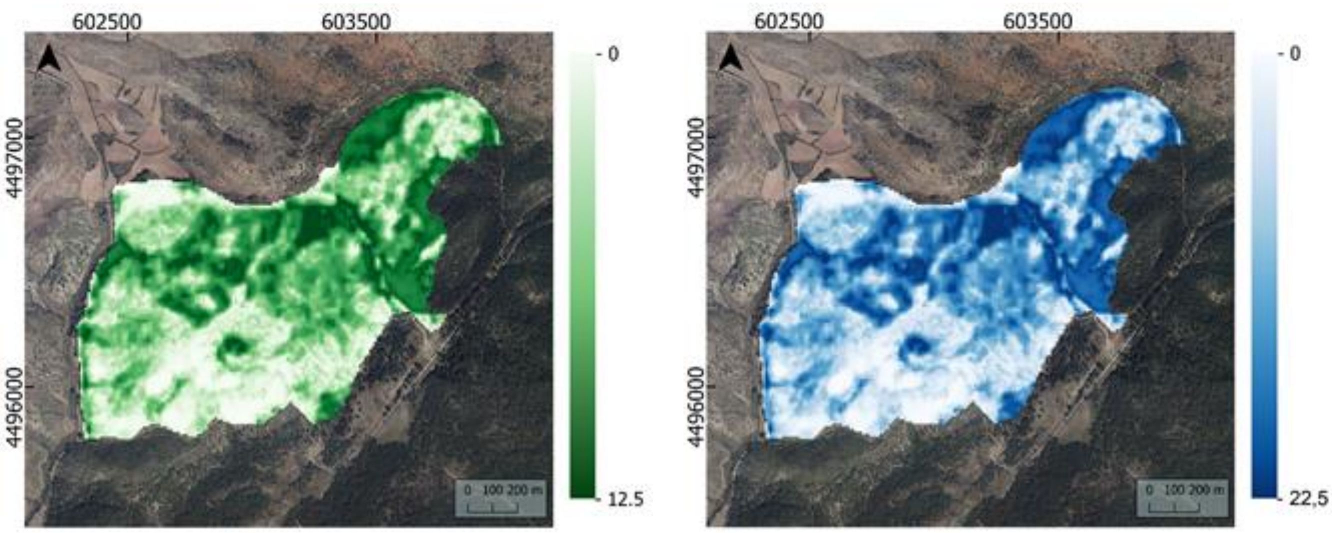
Figure 5. Measured Vs. estimated (GPR) biomass values for Sentinel-2. MAR.: Major Axis Regression fit; N: number of samples; B: mean bias; R^2 coefficient of determination; S: standard deviation of the bias; RMSE: root mean square error. Dash line is the 1:1 line; Continuous line is the MAR adjust. Green points are shrub plot; red points are bare soil plots.

The linear MAR fit between predictions and ground values reveals fitted slopes close to 1 (0.90) and small offsets (0.81 Mg*ha^{-1}) for a general algorithm (Table 3). All these results indicate that the combined use of random field sampling, low intensity of sampling and a minimum segmentation of the database just ensured by equivalent real and simulated distributions can provide excellent results for forest management purposes and carbon accountability by using probabilistic models applied to remote sensing imagery of Sentinel-2.

Table 3. Statistic of the performance of a Gaussian process depending on the satellite imagery and story segmentation. MAR: Major Axis Regression fit; R^2 coefficient of determination; RMSE: root mean square error.

Imagery	Structure	MAR	R^2	RMSE
Sentinel-2A	Shrubs of *Quercus pyrenaica* Willd.	y = 0.81 + 0.90x	0.83	1.58

Finally, the application of the general algorithm obtained from GPR has been applied over the corresponding area to achieve the total biomass in the post-fire regenerated area (Figure 6a). Additionally, the relation between biomass and $CO_{2\,eq.}$ has been considered to estimate the total sequestrated carbon in the next eight years after the wildfire (Figure 6b).



(**a**) (**b**)

Figure 6. Actual shrub biomass and carbon stock eight years after a fire burning a forest of *Pinus sylvestris* L. in Chequilla, Spain. (**a**) biomass of a regenerated shrub structure of *Quercus pyrenaica* Willd. (Mg*ha^{-1}), 2020; (**b**) carbon stock of a regenerated shrub structure of *Quercus pyrenaica* Willd. (Mg C*ha^{-1}), 2020.

For the total area of 136.52 ha a total sequestered carbon of 2,287.21 Mg of $CO_{2\,eq.}$ has been achieved by the young oak coppice due its natural recovery, meaning an average carbon sequestration of 2.095 Mg $CO_{2\,eq}$ ha^{-1} year^{-1}.

3.4. Application of GPR over non-burned areas

Although GPR has been trained for a specific forest structure its application over other type of forest has been tested. After the general algorithm obtained from GPR has been applied over the young oak coppice, an experiment has been performed by applying the same algorithm over 14.8 ha of non-burned areas (Figure 7). The study area for this experiment is the actual boundary between burned and non-burned areas and the forest to the east of the affected area remains as a mature forest dominated by *Pinus sylvestris* L.

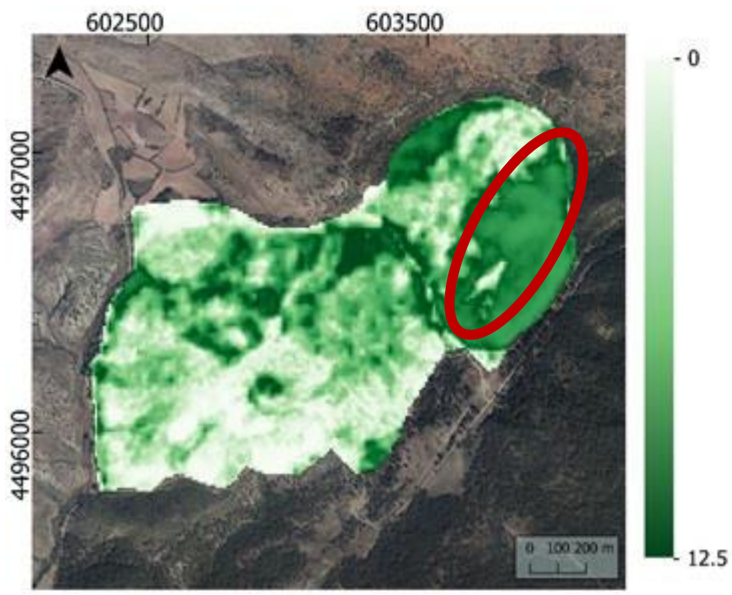


Figure 7. Application of the general algorithm over both a young oak coppice and a mature forest of *Pinus sylvestris* L. without class segmentation in the area of the Chequilla fire, June 2020. Inside the red cercle, mature forest of *Pinus sylvestris* L. with a remarkable underrating of biomass.

GPR gives underrated values of biomass for a homogeneous mature forest of *Pinus sylvestris* L. After a specific algorithm for young oak coppice has been applied over any other forest structure it has not been able to distinguish the species, the structure neither the biomass differences.

4. Discussion

A cost-effective and simple methodology for assessing biomass stock that requires a very low number of forest inventories has been proposed. This methodology applies Gaussian process regression methods over Sentinel-2 surface reflectance and forest inventories data obtained after fieldwork. In this way, carbon stock given as sequestrated $CO_{2eq.}$ can be directly calculated as a function of the total biomass data.

The results of this research allow the evaluation of an innovative and operative method of machine-learning, a non-parametric empirically based method of Gaussian process regressions for biomass assessment at decametric resolution for Sentinel-2. This method has been demonstrated to be an excellent method in terms of cost-effectiveness in a temperate oak coppice forests with shrubs. For an area up to 136 ha 33 samples have been enough to achieve an R^2 of 0.83, that means one sampling plot for each 4 ha of regenerated coppice.

Other methodologies have been purposed for accurate forest inventories, by traditional [7] or even LiDAR means [34,35] but the methodology here introduced reduces the cost and is able to represent more than 80% of the variability of the forest with only one plot each four ha.

This regression method needs no explicit selection of spectral bands and spectral resolutions of the MSI sensor of Sentinel-2 are valid for this purpose. This methodology overpasses the results obtained by traditional regressions between biophysical values and optical indexes. Avoiding the user intervention making an aprioristic selection of the bands allows the GPR to optimize the extraction of the relevant information of each band.

The biomass range in this study lies between 5 and 14 $Mg{*}ha^{-1}$ and the RMSE obtained is low, that shows a high level of determination found. Indeed, an average error of around 20% of total biomass in this coppice forest has been obtained.

This error magnitude is low enough to classify each pixel inside of the corresponding class of fuel models [4,5] and improves the spatial segmentation since representation of the results are accurate at Sentinel-2 pixels with a 10×10 m^2 resolution, while being this approach straightforward and easy to be implemented.

This methodology provides accurate information of biomass and carbon content in agreement with previous studies and can be used as a major improvement of the accuracy for the assessment of carbon sequestration [36].

Once carbon sequestration has been estimated for the whole area of study, and as far as it can be assumed a linear annual growth of young oak coppice until a dominant height of 5 m [37] an average carbon sequestration of 2,094 Mg ha^{-1} $year^{-1}$ for this shrub structure can be assessed, in accordance with previous carbon quantification studies [11].

In this study area of 10×10 km^2, it was clearly shown that NDVI values are very similar for both burnt and recovered areas and other not disturbed forest parts. Consequently, the application of GPR trained for a specific forest structure over other forest types must be avoided as it will create confusion in the assessment of any biophysical variable since reflectance values are quite similar for both vegetation groups and only a clear segmentation of the target areas could resolve this flaw.

Although some spectral methods for land use and land classification after a forest fire have been proposed for a segmentation of the areas [3], it is recommended the use of ground-truth maps to apply algorithms over the same forest structures. According to [11,12] segmentation of the classes is necessary to avoid errors in the biomass assessment, as reflectance values are not valid to segment forest structures in narrow areas.

Moreover, to take advantage of this methodology for biomass assessment, land segmentation is necessary for improving its accuracy because if any error of land classification is done, GPR will under- or overestimate the results as shown in an experiment over 14,8 ha of *Pinus sylvestris* L.

5. Conclusions

As a conclusion, the combination of Sentinel-2 imagery, Gaussian process regressions, a reduced database and story segmentation can be considered as a suitable method for shrubland biomass assessment and can be an essential tool for affordable and adaptive management at local scale by avoiding long and expensive forest inventories.

This methodology has been tested over a specific shrubland structure and further studies could be done focused in any other type of shrubs and pole structures. In addition, the effect of larger spatial scales should be analysed to better understand the effect of the density of plots in the result.

Author Contributions: Conceptualization, D.V.V.; methodology, D.V.V., and V.L.A.; software, D.F.; formal analysis, D.V.V. and D.F.; investigation, D.V.V.; data curation, R.R.P. and D.V.V.; validation, E.C.A.; visualization, D.V.V.; writing—original draft preparation, D.V.V. and R.R.P.; writing—review and editing, D.V.V. R.R.P., D.F. and V.L.A.; supervision, J.V.O.V.; funding acquisition, V.L.A and R.R.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Interreg SUDOE REMAS project – grant number SOE3/P4/E0954 – “Greenhouse gas emissions risk management in forest fires” (2019-2022), co-financed by the Interreg Sudoe Program through the European Regional Development Fund (ERDF).

Acknowledgments: Authors acknowledge Helena Van den Berg for her technical support.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Navarrete-Poyatos, M.A.; Navarro-Cerrillo, R.M.; Lara-Gómez, M.A.; Duque-Lazo, J.; Varo, M.d.l.A.; Palacios Rodriguez, G. Assessment of the Carbon Stock in Pine Plantations in Southern Spain through ALS Data and K-Nearest Neighbor Algorithm Based Models. *Geosciences* **2019**, *9*, 442. https://doi.org/10.3390/geosciences9100442
2. González-Díaz, P.; Ruiz-Benito, P.; Gosalbez Ruiz, J.; Chamorro, G.; Zavala, M.A. A Multifactorial Approach to Value Supporting Ecosystem Services in Spanish Forests and Its Implications in a Warming World. *Sustainability* **2019**, *11*, 358. https://doi.org/10.3390/su11020358
3. Navarro-Cerrillo, R.M.; Blanco-Oyonarte, P. Estimation of above-ground biomass in shrubland ecosystems of southern Spain. *Forest Syst*. **2006**, *15*, 197–207. http://dx.doi.org/10.5424/srf/2006152-00964
4. Scott, J.H.; Burgan, R.E. **2005**. *Standard fire behavior fuel models: a comprehensive set for use with Rothermel's surface fire spread model*. Gen. Tech. Rep. RMRS-GTR-153. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. https://doi.org/10.2737/RMRS-GTR-153
5. Quílez, R.; Chinchilla, S. Caracterización de modelos de combustible en la provincia de Valencia. *FLAMMA* **2013**, 3 (2): 31-40.
6. Serrada, R.; Garrachón, S. Distribución de biomasa en algunas repoblaciones de pino carrasco de la provincia de Guadalajara. *Cuad. Soc. Esp. Cien. For.* **2000**, *10*, 139-146.
7. DGCN. *III Inventario Forestal Nacional Español 1996-2006*. Editor, Dirección General de Conservación de la Naturaleza. Ministerio de Medio Ambiente: Madrid, Spain, 2006.
8. Aranha, J.; Enes, T.; Calvão, A.; Viana, H. Shrub Biomass Estimates in Former Burnt Areas Using Sentinel 2 Images Processing and Classification. *Forests* **2020**, *11*, 555. https://doi.org/10.3390/f11050555
9. Pasalodos-Tato, M.; Ruiz-Peinado, R.; del Río, M. et al. . Shrub biomass accumulation and growth rate models to quantify carbon stocks and fluxes for the Mediterranean region. *Eur J Forest Res* **2015**, 134, 537–553. https://doi.org/10.1007/s10342-015-0870-6
10. Montero, G.; López-Leiva, C.; Ruiz-Peinado, R.; López-Senespleda, E.; Onrubia, R.; Pasalodos-Tato, M. Producción de biomasa y fijación de carbono por los matorrales españoles y por el horizonte orgánico superficial de los suelos forestales. Ministerio de Agricultura, Pesca y Alimentación: Madrid, Spain. 2020.
11. Dong, J.; Kaufmann, R.K.; Myneni, R.B.; Tucker, C.J.; Kauppi, P.; Liski, J.; Buermann, W.; Alexeyev, V.; Hughes, M.K. Remote sensing estimates of boreal and temperate forest woody biomass: Carbon pools, sources, and sinks. *Remote Sensing of Environment* **2003**, *84*, pp. 393-410. https://doi.org/10.1016/S0034-4257(02)00130-X

12. Muukkonen, P.; Heiskanen, J. Estimating biomass for boreal forest using ASTER satellite data combined with standwise forest inventory data. *Remote Sensing of Environment* **2005**, *99*, 464-447. https://doi.org/10.1016/j.rse.2005.09.011
13. Kumar, L.; Sinha, P.; Taylor, S.; Alqurashi, A.F. Review of the use of remote sensing for biomass estimation to support renewable energy generation. *J. of Applied Remote Sensing* **2015**, *9(1)*, 097696. https://doi: 10.1117/1.JRS.9.097696.
14. Galidaki, G.; Zianis, D.; Gitas, I.; Radoglou, K.; Karathanassi, V.; Tsakiri-Strati, M.; Woodhouse, I.; Mallinis, G. Vegetation biomass estimation with remote sensing: focus on forest and other wooded land over the Mediterranean ecosystem. *International Journal of Remote Sensing* **2017**, *38:7*, 1940-1966. https://doi:10.1080/01431161.2016.1266113.
15. Verrelst, J.; Camps-Valls, G.; Muñoz-Marí, J.; Rivera, J.P.; Veroustraete, F.; Clevers, J.G.P.W.; J. Moreno, J. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – A review. *ISPRS Journal of Photogrammetry and Remote Sensing* **2015**, *10*, 273-290. https://doi.org/10.1016/j.isprsjprs.2015.05.005
16. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Macé, D.; Pragnère, A.; Myneni, R.; Knyazikhin, Y.; Wang, L. Retrieval of canopy biophysical variables from bi-directional reflectance. Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment* **2003,** *84*, 1−15. https://doi.org/10.1016/S0034-4257(02)00035-4
17. Verrelst, J.; Alonso, L.; Camps-Valls, G.; Delegido, J.; Moreno, J. (2012). Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Transactions on Geoscience and Remote Sensing* **2012**, *50(5)*, 1832-1843. doi: 10.1109/TGRS.2011.2168962.4
18. Wang,T.; Xiao, Z.; Liu, Z. Performance evaluation of machine learning methods for leaf area index retrieval from time-series MODIS reflectance data. *Sensors* **2017**, *17(1)*, 81. doi: 10.3390/s17010081.
19. Rasmussen C.E. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*; Bousquet O., von Luxburg U., Rätsch G., Eds. Springer: Berlin Heidelberg. Germany, 2004; Lecture Notes in Computer Science, vol 3176. https://doi.org/10.1007/978-3-540-28650-9_4
20. Ruiz de la Torre, J. Matorrales. In: *Tratado del Medio Natural.* Universidad Politécnica de Madrid: Madrid, Spain, 1981; Volume II, pp. 501-541.
21. Copernicus Open Access Hub: https://scihub.copernicus.eu/. Last visited: April 2022.
22. Canfield, R. H. Application of the line interception method in sampling range vegetation. *Jour. For.* **1941**, *39*, 388-394.
23. Silva, V.; Catry, F.X.; Fernandes, P.M.; Rego, F.C.; Bugalho, M.N. Trade-offs between fire hazard reduction and conservation in a Natura 2000 shrub–grassland mosaic. *Appl Veg Sci*. **2016**, *23*, 39– 52. https://doi: 10.1111/avsc.12463
24. Robert, C. P.; Casella, G. Monte Carlo Statistical Methods. Springer: New York, USA, 2004.
25. Ruiz-Peinado, R. Dpto. de Dinámica y Gestión Forestal. Centro de Investigación Forestal (INIA, CSIC). Madrid. Spain. Personal communication, 2021.
26. Ibañez J.J.; Vayreda, J.; Gracia, C. Metodología complementaria al Inventario Forestal Nacional en Catalunya. In: El inventario Forestal Nacional. Elemento clave para la gestión forestal sostenible; Bravo, F; del Río M.; del Peso, C., Eds. Fundación General de la Universidad de Valladolid: Valladolid, Spain, 2002; pp. 67-77.
27. Montero, G.; Pasalodos, M.; Montoto, R.; Lopez-Senespleda, E.; Onrubia, R.; Bravo-Oviedo, A.; Ruiz-Peinado, R. Contenido de Carbono en la biomasa de las principales especies de matorral y arbustedos de España. In Actas del VI Congreso Forestal Español. Sociedad Española de Ciencias Forestales Ed. Vitoria-Gasteiz, 10-14 junio 2013.
28. Camps-Valls, L.; Gómez-Chova, J.; Muñoz-Marí, J.; Vila-Francés, J.; Amorós, J.; del Valle-Tascon, S.; Calpe-Maravilla, J. Biophysical parameter estimation with adaptive Gaussian Processes. Proceedings of the Geoscience and Remote Sensing Symposium, Cape Town, South Africa, 2009; IEEE International, IGARSS 2009, Volume: 4.
29. Neal, R.M. Bayesian learning for neural networks. Springer: New York, USA, 1996.
30. MacKay, D. J. C. Introduction to Gaussian processes. In *Neural Networks and Machine Learning*; C. M. Bishop, Ed.; Kluwer Academic Press, 1998; pp. 133–166.
31. Rasmussen C.E, Nickisch, H. Gaussian process regression and classification Toolbox version 4.2 for GNU Octave 3.2.x and Matlab 7.x. Released under the FreeBSD License. The code and associated documentation are available from http://gaussianprocess.org/gpml/code. Last visited, 4th April 2022
32. Harper, W.V. Reduced major axis regression: Teaching alternatives to least squares. In sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, AZ, USA, 13–18 July 2014; Makar, K., De Sousa, B., Gould, R., Eds.; International Statistical Institute: Voorburg, The Netherlands, 2014; pp. 1–4.
33. Vinué-Visús, D. Coll-Aliaga, E.; Oliver-Villanueva, J.-V. Remote Sensing Assessment of Biomass Evolution Depending on Forest Management. *Proceedings* **2019**, *1, 3*. https://doi:10.3390/proceedings2019019003.
34. Ortuño-Pérez, S.F..; Núñez-Martí, M. V.; Mauro, F.; Gamazo, M. V; Martín-Fernández, S.; García-Abril, A. Estimación de los costes de inventario forestal con y sin información LIDAR. In Montes: servicios y desarrollo rural. Proceedings of the VI Congreso Forestal Español. 10/06/2013 - 14/06/2013, Vitoria- Gasteiz, Spain. pp. 1-10.
35. Vinué, D., Camacho, F., Fuster, B. Above-ground biomass and biophysical variables estimation in Mediterranean forest using terrestrial laser scanner. A case of study in Liria (Spain). In Proceedings of the Fifth Recent Adv. Quant. Remote Sens.; José A. Sobrino, Ed.; Universitat de València, Spain, 2017.
36. Montero, G.; Ruiz-Peinado, R.; Muñoz, M. Producción de Biomasa y Fijación de CO_2 por los Bosques Españoles; Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Ministerio de Educación y Ciencia: Madrid, Spain, 2005.
37. Adame, P.; Cañellas, I.; Roig, S.; del Río, M. Modelling dominant height growth and site index curves for rebollo oak (*Quercus pyrenaica* Willd.). *Ann. For. Sci*. **2006**, *63, 929-940*. https://doi.org/10.1051/forest:2006076