IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 November 2024, accepted 17 December 2024, date of publication 23 December 2024, date of current version 31 December 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3521047

RESEARCH ARTICLE

Physically-Based Detection Algorithm of Buried Archaeological Remains Using Spectral Signatures

ELIAS GRAVANIS AND ATHOS AGAPIOU
Department of Civil Engineering and Geomatics, Cyprus University of Technology, 3603 Limassol, Cyprus

Corresponding author: Elias Gravanis (elias.gravanis@cut.ac.cy)

This work was supported in part by European Union's Horizon Europe Framework Programme (HORIZON-WIDERA-2021-ACCESS-03, Twinning Call) under Grant 101079377; in part by UK Research and Innovation (UKRI) under Project 10050486; and in part by the DEPLOYED Project, an Internal Interdisciplinary Research Programme of Cyprus University of Technology.

ABSTRACT Detection of buried archaeological remains based on identification of archaeological proxies, such as cropmarks, has been widely used. Nevertheless, physically-based models for such archaeological prospection surveys are still missing from the literature. In this work we present a spectral classification criterion procedure for the detection of buried archaeological remains (cropmarks) using remote sensing techniques, in particular top-of-canopy hyperspectral data. The criterion is built by using (1) the radiative transfer model PROSAIL in inverse and forward mode to produce physically-based simulations of spectral signatures of an observed cropmark dataset captured from an artificial test-field, and (2) machine-learning methods (decision trees) to identify the highest importance wavelengths and the associated classification thresholds. This is done by statistically analyzing different simulated dataset size (synthetic hyperspectral image size) and different contents in signatures affected by buried 'remains' relatively to healthy crop signatures. The analysis of the results does indeed allow the formulation of a well-performing criterion, with above 70% detection rate in test synthetic datasets. Our findings show that the physical reduction of the degrees of freedom forming cropmarks plays a significant role in their modelling and successful detection. The underlying hypotheses and issues, as well as the generalizability potential of the method in different conditions are discussed.

INDEX TERMS Archaeological prospection, archaeological proxies, cropmarks, decision trees, hybrid method, hyperspectral image, inverse modelling, machine learning, PROSAIL, radiative transfer models.

I. INTRODUCTION

The detection of buried archaeological remains using remote sensing methods has become an integral part of current archaeo-landscape projects and archaeological prospection surveys [1], [2]. One of the most widely adopted methods for the identification of archaeological proxies is the detection of the so-called "cropmarks", which can -indirectly-indicate the presence of hidden archaeological remains [3], [4], [5]. Cropmarks appear when healthy vegetation, like those of barley or other cereal crops, is cultivated on top of buried archaeological remains. Consequently, crops might either be stressed (negative cropmarks phenomenon) or enhanced (positive cropmarks phenomenon) due to the different context [6]. This observation can be utilized as a first indication (proxy) of the existence of buried remains.

Archaeology is referred to as one of the earliest scientific disciplines that has adopted the use of remote sensing sensors [7]. In the recent past several studies have indicated the promising added value of the detection of cropmarks for different archaeological field surveys. For instance, in Scotland, annual programs of airborne surveys have been conducted by a variety of agencies since 1945 and continued under the auspices of Historic Environment Scotland till today [8], [9]. In general, the use of multispectral and sometimes hyperspectral sensors, sensitive in the visible and near-infrared part of the spectrum, is used for capturing "anomalies" of crops' spectral profiles that are cultivated on top of archaeological remains.

The associate editor coordinating the review of this manuscript and approving it for publication was Xinyu Du.

Indeed, until today a detailed understanding of the formation of cropmarks from a theoretical perspective is still missing in the literature. A better understanding of how cropmarks are formed, and which (physical) parameters affect their development still needs to be addressed by the scientific community. Despite the technological advancements of the space sector (i.e., higher spatial and spectral resolution sensors) and automation in image processing (i.e., deep learning algorithms), the establishment of theoretical models is not met.

To overcome this gap, the study initiates exploring the use of theoretical radiative transfer models, widely used for agriculture and forestry applications, based on a previously obtained ground ''truth'' spectral dataset cube. The findings of this study aim to establish, for the first time, theoretical models and spectral indices that can have regional applicability in the future.

From technical point of view, the radiative transfer model PROSAIL [19], is used to associate the observed signature dataset to a leaf and canopy physical parameters dataset. This way the spectral information is transcribed to biophysical and biochemical information. That is, the degrees of freedom forming the signatures are sought in the physical and geometric characteristics of crop canopy, at least the ones modelled in PROSAIL. The idea is that these degrees of freedom carry the information about the type of spectral signature they are associated with, i.e. affected by the buried 'remains' (denoted by A) or not (denoted by H). The statistics of the physical parameter dataset allows generating simulated physical parameters in such a way that carry the associated spectral signatures can be characterized as A or H beforehand; the signatures are obtained by running PROSAIL forward. Therefore large, synthetic (simulated) datasets of spectral signatures can be produced and analyzed.

In the first steps, our analysis proceeds following a previous publication of the authors [20]. In particular, we processed spectral signatures obtained from an experiment that took place in Alambra test field in Cyprus between 2011–2012 [21]. A small test field, measuring 5 × 5 meters, was constructed to simulate archaeological features below the surface. Ground spectroradiometric measurements were taken during the period October 2011 until April 2012. For each of the sixteen different campaigns organized during this period, spectral signatures were collected over the ''archaeological'' and non-archaeological areas. A calibrated handheld GER 1500 spectroradiometer was used for this purpose, that was able to record the electromagnetic radiation between the visible to the near-infrared spectrum. The signatures were calibrated using a Lambertian spectralon panel.

The signatures for the purposes of this study, are then transformed to the unit interval between their minimum in the deep blue (∼400nm) and their maximum in the near infrared (∼850 nm) and suitably averaged over a dataset representing synthetic signatures over a test field. This essentially defines a new index, which runs over different wavelengths with nanometer resolution. Then machine learning methods (decision tree algorithms) are used to identify the highest importance wavelengths in classifying a given signature as type A or H, according to the definition of type given above, along with the associated one-dimensional decision boundaries (thresholds). Our findings allow us the formulation of a simple and effective classification procedure (criterion) which allows us to identify the type A signatures in 'unknown' synthetic datasets with a success rate above 70%. The simplicity of the procedure rests mainly on the use of one-dimensional decision boundaries of the highest importance wavelengths and not the full multi-dimensional boundaries of the decision trees, and a crude use of the deduced distributions of highest importance wavelengths across the spectrum (being essentially replaced by uniform distributions).

In summary, in this paper we put forward cropmark detection criteria, given in terms of hyperspectral information, which are (1) physically-based, i.e. using a natural reduction of degrees of freedom in a hyperspectral signature, and (2) simply structured, requiring comparison with a threshold curve for a certain index. These features present significant potential for usability and generalizability across various archaeological context and crop – soil conditions.

The paper is organized as follows. Section II describes the construction of physically-based simulation of spectral signatures. Section III presents the decision tree analysis of synthetic analysis of synthetic datasets of different sizes and different content in type A signatures introducing our classification criterion. Section IV presents the assessment applied to the test synthetic dataset of different sizes and type A content. Finally, in section V a general discussion is given summarizing our findings and conclusions as well as pointing out important avenues of progress for future work are drawn.

II. PHYSICALLY-BASED SIMULATION OF SIGNATURES

Radiative transfer models describe the spectral and directional reflectance of plant canopies taking into account the geometric and biochemical leaf characteristics, the canopy architecture properties, background soil characteristics and the direction of observation, based on the physical laws of light transport within vegetation. These properties we will collectively call 'physical parameters' (of leaf and canopy), which we shall list and discuss below. The conceptual and physical basis of radiative transfer modelling of canopy reflectance has been reviewed e.g. in [22].

PROSAIL [19] is a widely applied canopy reflectance model due to its simplicity and strong overall performance, see e.g. [23] and references therein. Important geometric assumptions underlying PROSAIL is that canopy is modelled as an infinite turbid medium i.e., a homogeneous and dense layer of very small reflectors (leaves), and it is one-dimensional, in the sense that -by horizontal homogeneity-radiation transfer occurs in the direction across the canopy thickness [22], [24], as opposed to 3D or other types of radiative transfer modelling, see e.g. [25]. For a thorough comparison of various radiative transfer models see the RAMI exercises [26], [27].

Although PROSAIL has been used in the forward mode, e.g. for sensitivity analysis related to vegetation indices [28], the estimation (retrieval) of physical parameter values from observed reflectance data is the primary application of the model [23] by a variety of inversion techniques. The general problem of the retrieval of the biophysical and biochemical parameters from spectral data involves a variety of methods which have been combined with the physically-based techniques [29]. Common inversion methods applicable to radiative transfer models include: iterative optimization algorithms, Look-Up-Tables (LUTs) and hybrid inversion [30, 23, 31]. The classical optimization algorithms are regarded as computationally demanding and susceptible to being trapped in local minima of the cost function [32]. The widely used LUT-based inversion is computationally faster although it may be affected by the ill-posed-ness of the inverse problem and low estimation accuracy [25]. A variety of methods improving the robustness and accuracy of the LUT-based algorithms have been used, such as incorporating prior knowledge; see e.g. [30] and references therein. A major drawback of the LUT approach is that the cost function is an open issue because the structure of the parameter uncertainties is in general not known [33].

In hybrid inversion [30, 34, 23, 31] one uses LUTs, that is, simulated synthetic data, and a certain method in the realm of machine learning to model the non-trivial map: reflectance signature space → physical parameter space. Vegetation indices (VIs), or other transformations of the spectral data, may also be used in the place of the reflectance signatures, and most often modelling is oriented towards the retrieval of specific biophysical variables, such as leaf chlorophyll content or leaf area index; for a recent review of various approaches and objectives in the context of the hybrid method see e.g. [31]. Machine learning methods employed in the hybrid approach range from relatively simple statistical methods such partial least square regression (PLSR) and decision trees, to more complex methods such as kernel-based algorithms, random forests and ensemble methods, and neural networks; the subject is covered in an already quite extensive and rich literature, see e.g. [30], [31], and [33]. It should be mentioned that the use of the hybrid method (with neural networks combined with PROSAIL) has reached the operational level e.g. through the algorithm implemented in Sentinel Application Platform (SNAP) biophysical processor tool [35]. Additionally, the application of neural networks has progressed well into the deep learning approaches, such as convolution neural networks [36]; interesting recent works include [37], [38].

A. THE GENERAL PROCEDURE

In this work we adopt a pragmatic approach while applying the hybrid method. Given a dataset of observed reflectance signatures but without direct measurements of biochemical and biophysical variables of the crop being available, we use PROSAIL to reduce the degrees of freedom of the problem in a physically plausible way, in order to generate (physically-based) simulations of the original dataset. The physical reduction is a rather necessary step because of the multicollinearity present in the hyperspectral data, a feature which has been very well analyzed and discussed in efforts to statistically retrieve the values of, or model, the physical variables, see e.g. [33] and [39]. It has also been shown that using a statistical method for dimensionality reduction before the radiative model inversion in the context of the hybrid method, improves the accuracy of the physical parameter retrieval [40]. In this work, we shall apply inversion directly to the hyperspectral data.

Given the dataset of the 106 observed signatures, PROSAIL model inversion is done by the iterative optimization method using standard Python language [41] optimizers to obtain the group of values of the physical parameters associated with each signature. Stability of the inversion has been checked by perturbing the initial conditions of the estimated physical variables (discussed below) within their adopted range. PROSAIL is implemented using the Python language bindings. Hence, we obtained a dataset of 106 groups of physical parameters. Each group is labelled as A or H, depending on the signature is associated with.

The statistics of this dataset allows us to generate simulations of the A and H groups of physical parameters. By running PROSAIL forwardly we then generate physically-based reflectance signatures. Due to the (relatively small) size of the original dataset, we treat the entire period January-February-March as a single period, therefore values of the physical parameters from anytime in the period may be used to produce any single simulated signature. Therefore, the simulated signatures cannot have a time-label. On the other hand, we know by construction whether any given simulated signature is an A or a H one.

B. PROSAIL INVERSION

The PROSAIL physical parameter values and range used in the inversion process are generic for grain crops [23] and they are given in Table 1. The seven quantities indicated in bold are varied in the given intervals and they form the group of the physical parameters on which our simulation procedure is based. The values chosen for certain fixed parameters and the choice of the estimated parameters rest on their expected importance, see e.g. [19] and [42] and numerical experimentation aiming at plausible distributions of estimated values.

Figure 1 shows the top and lower 10% percentile ordered by the value of Pearson R2 parameter of the fitted curves with respect to the original signatures. It should be noted that the characteristic pattern of the signatures on the right is not exhibited by the signatures below the lower 10% percentile of fit (not shown here), pertaining more to the signatures on the left. The overall performance of the fitting process can be seen in the histograms shown in Figure 2, showing the

TABLE 1. **The values and the ranges of the PROSAIL input parameters used in this study.**

Parameter	Value
Leaf structure index, N [-]	1.5
Leaf chlorophyll content, Cab [μg cm^{-2}]	0 – 100
Leaf carotenoid content, Car [μg cm^{-2}]	0 – 30
Leaf dry matter content, Cm [g cm^{-2}]	0.002
Equivalent water thickness, Cw [cm]	0.02
Brown pigment content, Cbrown [-]	0
Leaf anthocyanin content, Cant [μg cm^{-2}]	0 – 4
Leaf area index, LAI [m^2 m^{-2}]	0.5 – 8
Average leaf angle, lidfa [deg]	30 – 89
Hot-spot size parameter, hspot [m m^{-1}]	0.25 – 0.75
Soil reflectance, psoil [-]	0 – 1.0
Sun zenith angle [deg]	0
View zenith angle [deg]	0
Relative azimuth angle [deg]	0

distribution of values of the root-mean-squared (RMSE) of the reflectance expressed in percentage (on the left) and of the Pearson R2 values (on the right). It is worth to note that these RMSE values are smaller than the radiometric uncertainty of widely used satellite sensors.

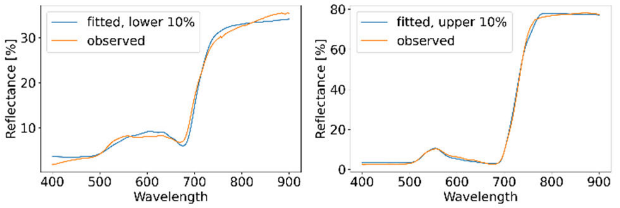

FIGURE 1. **The observed and fitted reflectance curves for upper and lower 10% of the R^2 distribution.**

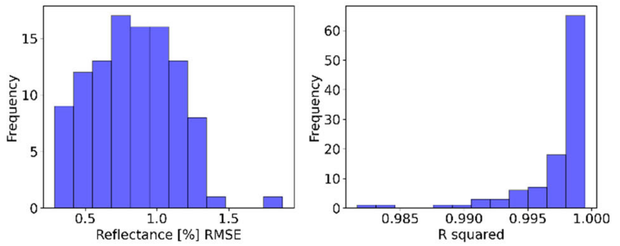

FIGURE 2. **The observed and fitted reflectance curves for upper and lower 10% of the R^2 distribution.**

The estimated parameters are plotted in Figure 3. They are organized according to the date of the associated signature and the label A or H. The mean value of each parameter per day of observation is shown in dashed lines. The first thing one may observe is that there is a noticeable difference between the estimated parameters for the A and H observations, for all parameters except, perhaps, the hot-spot size parameter (hspot). There is therefore an imprint of the type of signature (A or H), which, as we have explained, we wish to utilize in generating simulations of these types of signatures.

The leaf average angle (lidfa) is a canopy structure parameter quantifying the average inclination of the leaves. The leaf average angle is an important parameter from the point of view of PROSAIL inversion, as it affects the estimation of other physical parameters such as the leaf chlorophyll content or the leaf area index [23]. The lidfa curves in Figure 3 indicate erectophile leaves, as expected for barley plants. The downward trend of the curve i.e. the leaves turning toward horizontal, and the steep upward trend in the last interval, are consistent with the growth cycle of the plant. One may note that the leaves of the A observations have quite smaller angles than the H ones. One may also note that lidfa curves trend suggests that this parameter is negatively correlated with Cab and Car, as it will be verified below.

The leaf area index (LAI) is another canopy architecture parameter which is defined as the total one-sided green area of leaves per unit horizontal area of the ground [47]. It is a key parameter for modeling mass and energy exchange between vegetation and the atmosphere, and hence plays an important role in the fields of agriculture (for crop growth and yield), forestry, and ecology, as well as a global climate variable [48], [49]. As opposed to the leaf inclination angle, the leaf area index can be measured by numerous methods, direct (more or less destructive) or indirect methods, see e.g. [50], thereby allowing validation of canopy modelling. For these reasons, LAI estimation is the most primary example in parameter retrieval applications using PROSAIL [23]. The magnitude of estimated LAI values is rather typical for grain crops, see e.g. [48], [51], and [52] and barley in particular, see e.g. [53], [54], [55], and [56]. LAI is expected to increase with time, see e.g. [57], [58], and [59]. The nearly constant trend exhibited by the LAI curves in Figure 3 suggests that we observe essentially the maximum value of LAI in the given phenophases. We also observe that the A observations tend to have lower LAI than the H ones.

The soil reflectance (psoil) parameter (range 0 to 1), which quantifies soil brightness and is used as interpolation parameter between to two (wet and dry) reference soil reflectance spectra, starts at low reflectance and increases with a gentle slope reaching roughly 20% in the final date. This is consistent with the growth stage of the plant and the calendar time as the soil moisture is expected to decrease from mid-winter to mid-spring.

The important parameter hspot quantifies the effect of reflectance enhancement in the case where the direction of observation coincides with the direction of sunlight. This is known as hotspot or opposition effect and may be explained by the fact that, when observations are made from other directions, the scattered light has a higher probability of being scattered away by intervening particles (leaves), see e.g. the discussion in [60], [61], and [62]. Given that all observations in the dataset under consideration were obtained from approximately solar zenith view, the hot spot effect plays a significant role. The hspot is interpreted geometrically as the ratio of leaf (horizontal) size to the canopy height [62]. The barley plant has long but very curved leaves and it is not obvious how to attribute a specific number for the hspot parameter. For these reasons, we let it vary between the values from 0.25 to 0.75 (see Table 1). The result, shown in Figure 3, is an essentially random distribution of hspot

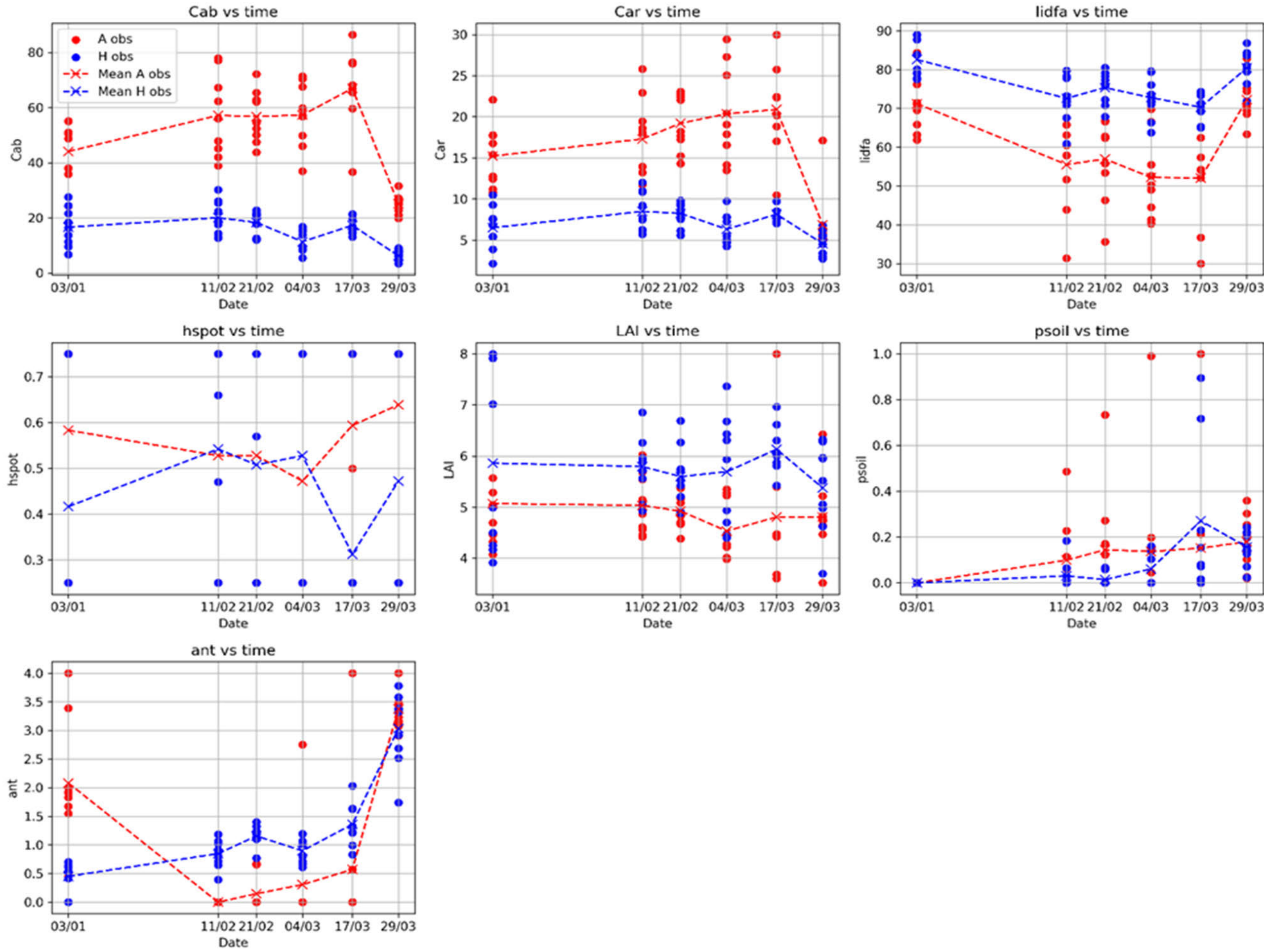

FIGURE 3. **The estimated physical parameters by PROSAIL inversion for the A and H observed signatures group according to the date of observation. Their mean values per day are shown in dashed lines.**

between the extreme values. (It should also be noted that numerical experimentation showed that the distribution of the other physical parameters is not affected by the specific values of the extreme values of hspot.) This statement will be further supported below by the fact that the parameter hspot is essentially uncorrelated with the other parameters and its variation may be attributed to -primarily- random effects of the measuring procedure.

C. PHYSICAL PARAMETER STATISTICS

As mentioned above, each signature and hence every group of estimated physical parameters is labeled as A or H. The latter are understood as qualitatively different states of the plant and canopy for which we would like to find a quantitative classification algorithm. Hence, the seven estimated physical parameters (Cab, Car, lidfa, hspot, LAI, psoil, ant) corresponding to A or H should be treated a priori independent variables whose statistics should be quantified. That is, we have totally fourteen variables, denoted by adding an index A or H to the physical parameter symbols, e.g. Cab_A or psoil_H. Recalling that the original dataset of observed signatures consists of 53 A and 53 H signatures, we have to analyze a dataset of 53 fourteen-parameter groups. Calculating the mean value of each parameter across this dataset, we may define the fluctuations of the parameters: fluctuation = parameter value – parameter mean. The 14 × 14 correlation matrix of the parameter fluctuations in the dataset is visually given in Figure 4. The color-bar indicates the magnitude and sign of the correlation between any two variables.

As expected, there is a strong positive correlation between the chlorophyll and carotenoid content of leaves (correlation coefficient 0.92), as is well known in the literature, see e.g. [42]. The strong correlation holds only between same type of observations, A or H. A relatively significant negative correlation (with coefficient –0.48) holds between the leaf chlorophyll content and leaf average angle. Similarly, anthocyanins are also correlated with chlorophyll content with coefficient –0.53 for the A observations and –0.39 for the H observations. These correlations are consistent with our previous observations regarding the time-evolution of parameters shown in Figure 3. Finally, the hot-spot parameter shows a low correlation with the other parameters roughly at 0.1 on average. Hence, we shall treat this parameter as varying independently of the other.

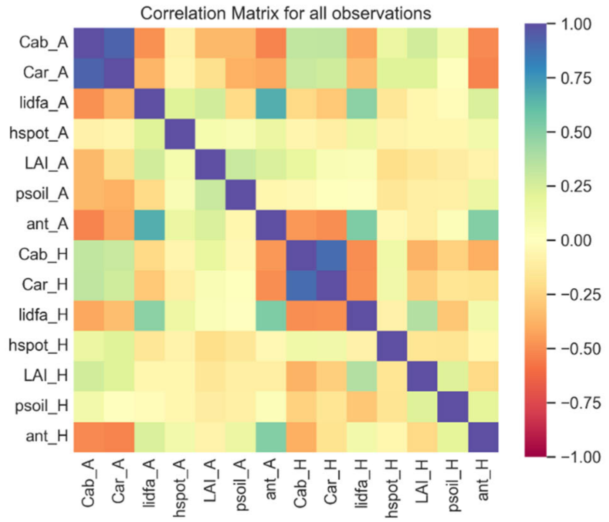

FIGURE 4. **Chromatic representation of the correlation between the 14 estimated parameters (seven physical parameters for each of A and H observations).**

A different overview of the physical parameters' dataset is given by the histogram representations of Figure 5, where we plot seven-bin histograms of the values of each physical parameter for the A and H observations. We observe that the values for the chlorophyll content, carotenoid content, average leaf angle and leaf area index are more or less symmetrically distributed around their mean; one, also observes the higher variance in the values of the A observations relatively to the H observations, a fact not clearly visible in the timeline graphs of Figure 3. On the other hand, the soil reflectance coefficient and anthocyanins are skewed to the low values. Finally, the hot-spot parameter appears to vary randomly between the extreme values 0.25 and 0.75 with nearly equal probability (and as mentioned above, practically uncorrelated with the other parameters). For completeness, the mean and standard deviation values for all parameters expect the hspot ones are explicitly given in Table 2.

D. PHYSICALLY-BASED SIMULATIONS OF REFLECTANCE SIGNATURES

The information contained in the histograms and the correlation matrix is now used to model the statistics of the dataset in order to generate simulations of the physical parameters corresponding to A and H events. The procedure employed is realized by standard routines of Python 3.9 and may be described as follows: (1) The (empirical) correlation matrix is Cholesky decomposed. (2) Generate a number of normal multivariate simulations of the twelve variables given in Table 2 using the mean values and empirical correlation matrix. The hspot parameters are simulated separately as we discuss below. (3) For each variable, the generated data are sorted (ranked) and transformed to the unit interval [0,1]. The scaled data are regarded as (unit) uniform random variables. (4) The cumulative distribution functions associated with the empirical histograms are then used (by percentile matching) to generate simulations of physical parameters. These steps were implemented in code using the functions of Python 3.9. The hspot parameters (hspot_A and hspot_H) are modelled and simulated as independent binomial variables with values 0.25 and 0.75 with equal probabilities of occurrence. Hence, we obtain a synthetic dataset of physical parameter simulations. Finally, we run PROSAIL forward to obtain a synthetic dataset of simulated reflectance signatures. Figure 6 shows the flowchart of the computations for the generation of physically-based simulations of signatures.

III. DECISION TREES AND CLASSIFICATION CRITERIA

A. DEFINITION OF AN INDEX

Given this assumption, we generate samples of simulated signatures, which we think of as synthetic hyperspectral 'images' i.e. a (synthetic) data cube. Any element of the sample we may think of as a 'pixel'. (These pixels will not have a spatial meaning.) Therefore, each pixel carries a simulated signature $\rho(\lambda)$. The entire image is described by a single function $\rho(\text{pixel}, \lambda)$. We generate a simulation batch, that is, a number of simulations of the synthetic hyperspectral image.

First, we rescale as we did in a previous publication [20] the signature of each pixel (independently) to produce a function $\rho_{\text{rescaled}}(\text{pixel}, \lambda)$ which takes values in the interval [0,1]. The rescaled reflectance is defined by

$$\rho_{\text{rescaled}}(\text{pixel}, \lambda) = \frac{\rho(\text{pixel}, \lambda) - \min(\rho(\text{pixel},\lambda))}{\max(\rho(\text{pixel},\lambda)) - \min(\rho(\text{pixel},\lambda))} \tag{1}$$

For each pixel, this function has a minimum value 0 (near deep blue) and a maximum value 1 (in the NIR). In order to visualize the effect of rescaling consider a synthetic image of 1000 pixels, containing 15% of pixels A. In Figure 7 we plot the histograms of values for the measured and rescaled reflectances at the wavelength $\lambda = 571$ nm, which is a simple useful choice for classification purposes [20]. One observes that rescaling reduces significantly the overlap between the A and H observations facilitating classification.

Now for each pixel we construct our index for every wavelength defined by

$$\text{index}(\text{pixel}, \lambda) = \left\langle \frac{1}{\rho_{\text{rescaled}}(\text{pixel}, \lambda)} \right\rangle \rho_{\text{rescaled}}(\text{pixel}, \lambda) \tag{2}$$

where the symbol $<\ldots>$ denotes an average over the entire image and is defined by (3):

$$\left\langle \frac{1}{\rho_{\text{rescaled}}(\text{pixel}, \lambda)} \right\rangle = \frac{1}{N-1} \sum_{\text{all pixels} \neq \text{pixel}} \left(\frac{1}{\rho_{\text{rescaled}}(\text{pixel}, \lambda)} \right) \tag{3}$$

where N is the size of the sample (image). This average depends on the wavelength and (weakly) on the pixel.

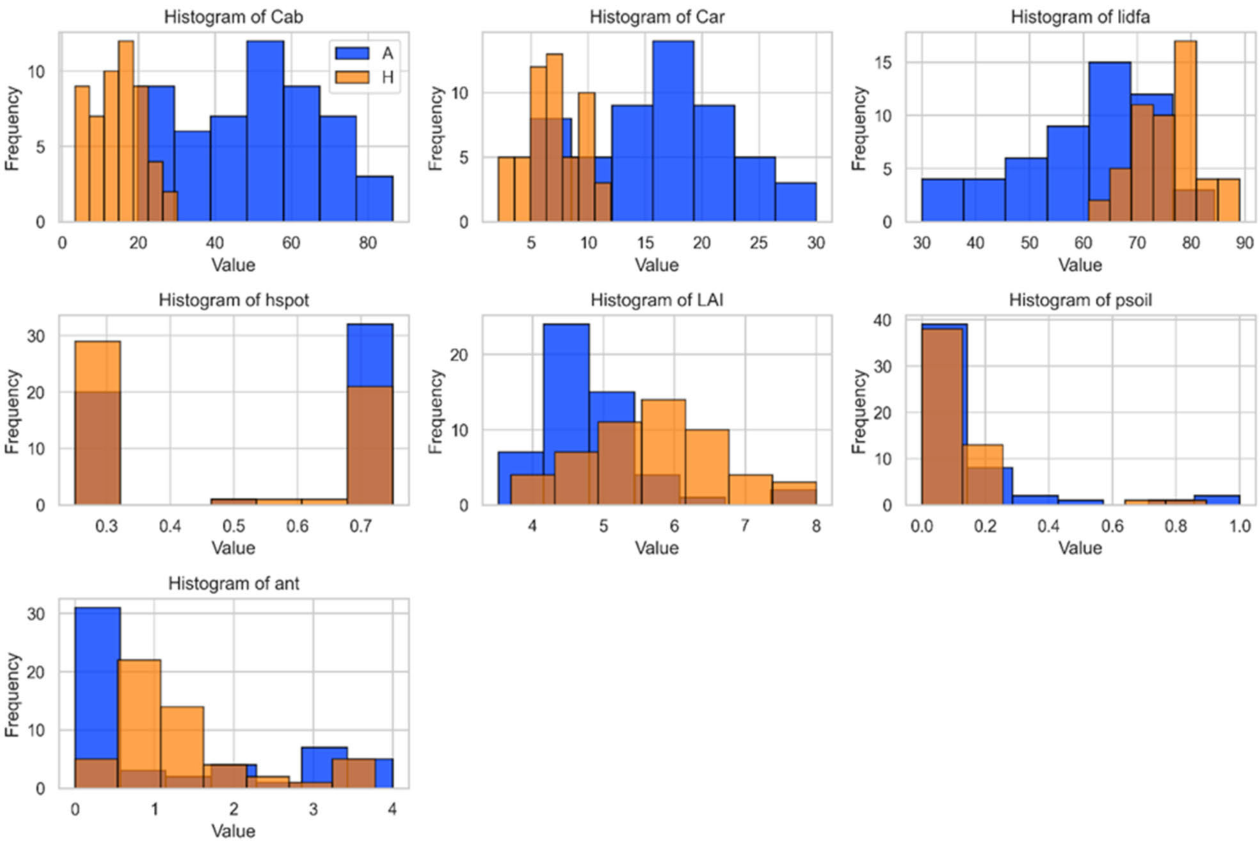

FIGURE 5. Histograms of the seven physical parameters estimated value for the A and H observations.

TABLE 2. Mean values and standard deviations of the physical parameters (in units given in Table 1).

	Cab_A	Car_A	lidfa_A	LAI_A	psoil_A	ant_A	Cab_H	Car_H	lidfa_H	LAI_H	psoil_H	ant_H
mean	50.92	16.56	60.19	4.86	0.12	1.09	14.92	7.02	75.77	5.74	0.09	1.29
st. dev	17.13	6.49	12.61	0.85	0.23	1.46	6.54	2.28	6.31	0.99	0.16	0.90

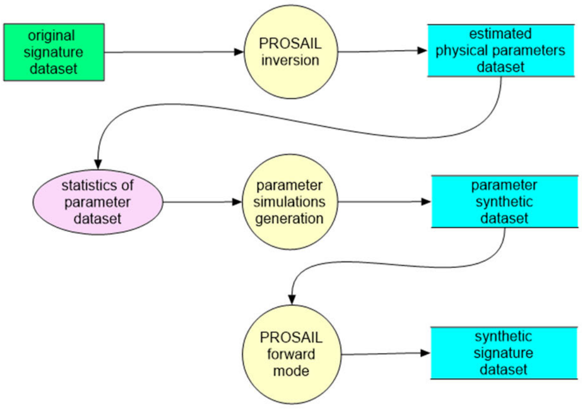

FIGURE 6. Flowchart of computation for generating physically-based synthetic signature.

Essentially, the definition of the index is a way to introduce a dimensionless signature using the mean value of the inverse signature over the image as a scale (that is, as a unit).

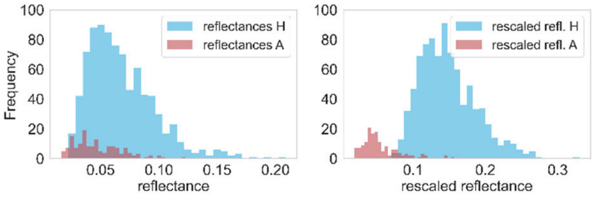

FIGURE 7. Histograms of the values of reflectance (left) and rescaled reflectance (right) for 571 nm.

B. DECISION TREES AND CLASSIFICATION CRITERIA

On each synthetic hyperspectral image, which is a sample of signatures, we apply machine learning, and in particular, decision trees [63], [64], to find the decision boundary separating the A and H signatures. The simplicity of the decision tree algorithm, and its natural affinity with classification problems, makes it a most suitable method for our purposes i.e. the formulation of explicit and simple classification criteria. A priori, the decision boundary is a 500-dimensional surface in the space of the index values for the wavelengths 400 nm to 900 nm. As before [20] we find that there is almost always a single feature (wavelength or the corresponding index value) that dominates. Therefore, we shall treat the decision boundary as 0-dimensional, although with a lower degree of success than the actual higher dimensional boundary. That is, we may think in terms of a simple threshold for the index at the dominant wavelength that separates the A and H signatures. Such an approach allows us indeed to formulate classification criteria which are explicit as well as simple. More formally, on a given image we find a classification criterion of the following form: If, for any given pixel on the image, we have that

$$\text{index}(\text{pixel}, \lambda = \text{dominant wavelength}) < \text{threshold} \quad (4)$$

then this pixel is a (simulated) A observation. Details of the application of the decision trees algorithm in this problem will be given shortly. First, we explain further the specifics of the simulations we shall run and analyze.

The dominant wavelength (and the associated threshold) is not the same for all images in the simulation batch. Therefore, we obtain a distribution of dominant wavelengths and associated values of threshold. Of course, certain nm will repeatedly arise and some of those will arise more frequently. Also, for a given wavelength arising as dominant, the threshold value changes from image to image. To visualize this information, we start by presenting unique value histograms (i.e. the bin width is exactly 1 nm) and the associated mean value of the threshold for each bin (wavelength). Figure 8 shows the flowchart of this analysis.

To generate a specific synthetic image two additional parameters are required: its size, that is, the number of pixels (simulated signatures) and the number or the percentage of these pixels which are type A. In order to acquire an understanding of the influence of these parameters on the classification analysis at the basic level of 1 nm bands, we investigate two cases of the image size, 100 and 500 pixels, containing 2, 5, 10 and 20 % pixels of type A. The simulation batch contains 5000 synthetic images for case of the 100 pixels image, and 1000 synthetic images for the 500 pixels image. For ease of reference, we summarize the image and simulation batch size in Table 3.

Now, for each case of image size and content in A pixels, we create an image twice the size, that is, containing twice as many A pixels, using each half as a training and validation set, respectively. Using gini impurity as the default measure of the quality of tree splits in Python scikit-learn decision tree classifier, and maximum node depth equal to fixed 4, beyond which the models do not further improve, we run decision tree classification for every image in a simulation batch (the size of the batch is given in Table 3).

An understanding of the performance of these decision tree models in the various cases can be acquired by plotting the histograms of classification score values for an entire batch. In Figure 9 we present the histograms of the recall and precision score values, for both image sizes and for 'few' (e.g. 5) and 'many' (e.g. 50) A pixels. From these results we conclude that the scores values are always fairly high but get lower as the proportion of the A pixels in an image gets higher. For example, one observes that both precision and recall are really high in the large image case with few A pixels, while the scores spread to lower values in the case of many A pixels in a small image.

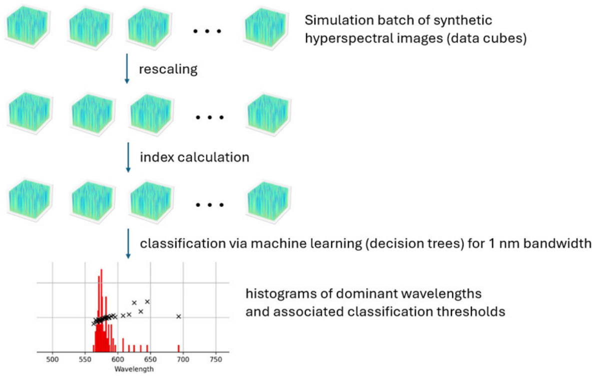

FIGURE 8. **Steps of analysis of the synthetic image dataset to deduce classification information (highest importance wavelengths and associated thresholds).**

TABLE 3. **Number of simulations for each image size.**

Image size	Simulations
100	5000
500	1000

From the obtained decision tree models, we are interested in keeping only minimal information: as we explained above, we keep only the wavelength and the corresponding threshold value (for the index given by equation 2) of highest importance. The models allow us to see that these dominant wavelengths, arising from the first split, are indeed of high importance: The distribution of feature importance values in the simulation batches corresponding also the results of Figure 9 are presented in Figure 10. The results show that importance is essentially always higher than 70%, with median values above 80%. We also see, same as with the score value distributions (Figure 9) that dominant feature importance is more widely distributed as the proportion of the A pixels in the image gets higher.

We may now summarize our procedure of feature and threshold extraction given a simulation batch: (1) for every image in the batch, which is an index (eq. 2) data cube, we run a decision tree algorithm to extract the highest importance (dominant) wavelength and the associated 0-dimensional decision boundary (threshold); (2) across the batch, we summarize these threshold values for every dominant feature occurrence. The results of this analysis are shown in the Figures 11 and 12 for the 100 and 500 pixels image, respectively. In these figures, the primary vertical axis corresponds to the relative frequency (counts fraction) of appearance of given wavelength as dominant. The secondary axis corresponds to the mean threshold value for the events (images)

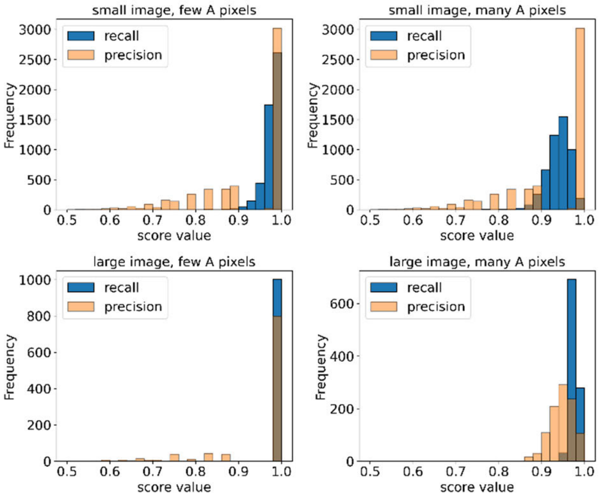

FIGURE 9. Distributions of classification score values in the indicated cases.

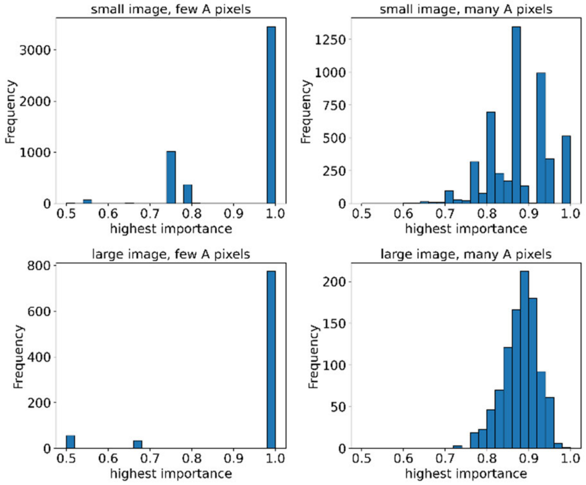

FIGURE 10. Distributions of dominant wavelength importance values in the indicated cases.

for the given wavelength as dominant. For the case of the smaller image of 100 pixels, all histograms in Figure 11 indicate that primarily two wavelengths, 571 and 693 nm, arise as dominant. Moreover, for low values of A of pixels, the 693 nm wavelength is more important for classification than the 571 nm, but that is reversed as the number of A pixels increases. The threshold for both these important wavelengths increases with the number of A pixels, remaining always below the value 1.0. We also observe that the threshold has a peculiar behavior in the neighborhood of 670 – 680 nm. This is due to the physical characteristic of the reflectance signatures attaining a local minimum and exhibiting a very low value at those wavelengths in the red-edge part of the spectrum (see Figure 1). Figure 12 tells a similar story for the 500 pixels image. When the A pixels are a few, these two wavelengths are the important ones. We observe though that, same as with the 100 pixels image, other wavelengths visibly arise as dominant in the vicinity of 571 and 693 nm as the percentage of A pixels increases. Moreover, one finds similarities upon comparing e.g. the 10% A histogram for the 100 pixels image with the 2% A histogram for the 500 pixels image, which both contain the same number of A pixels (10 pixels). This suggests that the absolute number of A pixels, and not their percentage in the image, shapes the frequency pattern. This is indeed the case, as we shall explicitly show below.

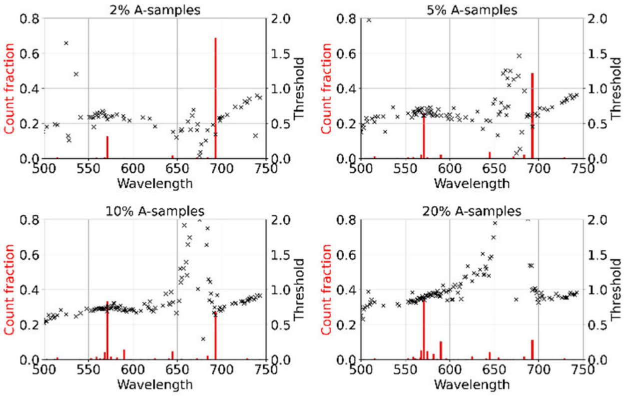

FIGURE 11. Distribution of dominant wavelengths for the 100 pixels image containing different percentages of A pixels.

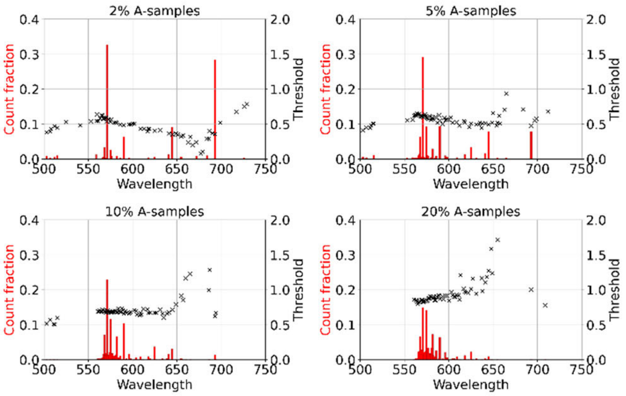

FIGURE 12. Distribution of dominant wavelengths for the 500 pixels image containing different percentages of A pixels.

The fact that many different wavelengths arise as dominant, even with somewhat low but not insignificant counts, suggests that it is more pertinent to think in terms of suitable bands with widths fairly larger than 1 nm. Then we may study in more detail the influence of the number of the A pixels on the classification criteria, defined by the dominant wavelength (band) and the associated mean threshold. That will provide the means by which one may discover the A pixels in unknown image, or more precisely, assess the ability of the formulated classification criteria to identify the A pixels in a given synthetic hyperspectral image.

In the light of the previous observations, we empirically aggregate the wavelengths into bands around the two important wavelengths, which we designate as ‘visible band’ and ‘red edge band’ and given in Table 4:

TABLE 4. **Defined band range.**

band	wavelengths (nm)
visible	550 – 650
red edge	680– 700

Then, the counts are added for all wavelengths within each band and the associated thresholds are averaged uniformly within that band. It is important to emphasize that that decision trees algorithms are not run again in a multispectral fashion. This was attempted and failed to give any meaningful results, thereby showing that multispectral information is too aggregate for our purposes. Instead, the classification analysis should be done entirely at the level of hyperspectral resolution, and then smooth out the final results by aggregating into bands.

In terms of these bands, we turn now to quantify in detail the influence on the classification of having different numbers and percentages of A pixels present in an image. Figures 13 and 14 summarize the results of our analysis for the images of 100 and 500 pixels, respectively. In particular, both figures show: (1) the fraction of cases in the simulation batch where the each band in Table 4 as dominant (counts fraction), presented in vertical bars and labelled in the primary vertical axis of the figures as function of the number of A pixels in the synthetic hyperspectral images, and (2) the associated mean threshold for the given spectral band and for every given number of A pixels in the images, presented as lines with bullet markers. In both Figures 13 and 14 the horizontal axis refers to the number of A pixels in the images, which corresponds to different percentage of A pixels in the two different image sizes. We observe that counts fraction is nearly the same for the two image sizes, when understood as a function of number of A pixels. We also observe for very low numbers of A pixels the band 680 nm – 700 nm (which we have called the red edge band) arises far more frequently than the visible band (550 nm – 650 nm), and the visible band becomes more important for 7 or more A pixels in the image for both image sizes, and gradually becomes the dominant spectral band. We also observe that both bands together are dominant in the 85% or more of the synthetic images in the simulation batch.

This is so because other bands play a role in the classification by small percentages. Nonetheless, we shall regard the chosen two-band system, whose members apparently behave in a complimentary way, as adequately covering the dominant wavelengths for our purposes. We observed that threshold curves are quite different for the two image sizes. For the small images (Figure 13) the curves exhibit larger values than the threshold curves for the larger images (Figure 14). Additionally, the curves for the smaller images are nearly straight lines, and they have a specific point where they cross each other.

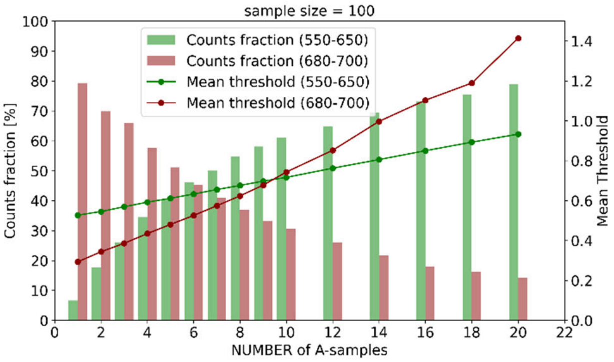

FIGURE 13. **Mean value of the threshold and percentage (counts fraction) of dominant wavelengths in the visible band (green color) and red-edge band (red color) of Table 4 for the 100 pixels image as a function of the number of A pixels in the image.**

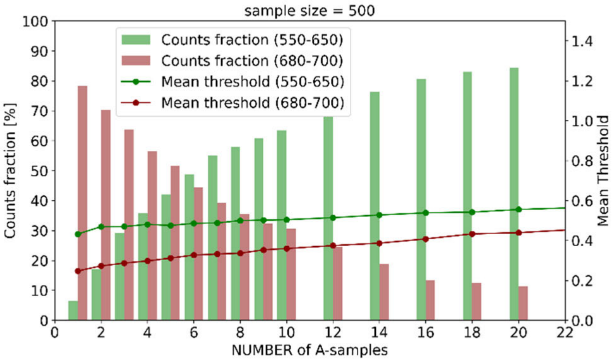

FIGURE 14. **Mean value of the threshold and percentage (counts fraction) of dominant wavelengths in the visible band (green color) and red-edge band (red color) of Table 4 for the 500 pixels image as a function of the number of A pixels in the image.**

The behavior of the threshold curves is clarified in Figure 15 where we plot the mean thresholds alone against the percentage of A pixels in the image for both spectral bands and images sizes. We observe that, for each band, the curves for the small and larger image (given in continuous and dashed lines, respectively) do not differ significantly when plotted against the percentage variable. The color coding for the spectral bands is the same as above.

A difference arises with the curve for the red edge band for the larger image, which behaves erratically beyond 8%, which corresponds to 40 A pixels in the image. The reason for that is the low count fraction, that is a smaller sample to calculate the mean threshold from, for those high numbers of A pixels. This becomes clear by introducing an 'overall mean threshold' curve, shown in blue line with cross markers, which is the mean value thresholds over both bands (where V = visible band and RE = red-edge band, as defined in Table 4), (5), as shown at the bottom of the next page, in order to provide a single number for the threshold that takes into account the statistical relative weight of each spectral band. We observe immediately that the overall mean threshold curve for the larger image coincides with visible band curve

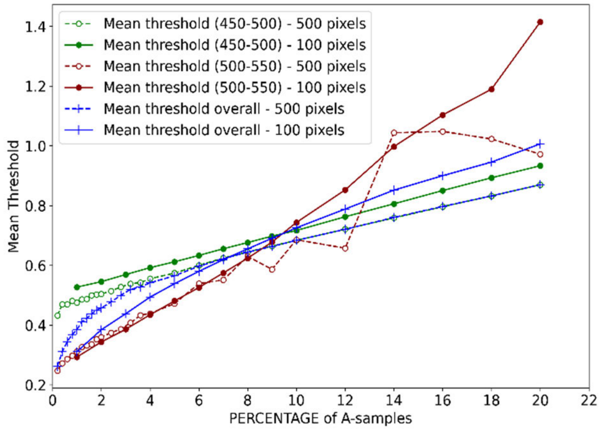

FIGURE 15. **Mean value of the threshold and percentage (counts fraction) of dominant wavelengths in the visible band (green color) and red-edge band (red color) of Table 4 for the 500 pixels image as a function of the number of A pixels in the image.**

beyond 8%, which means that the count fraction of the red edge band is negligible for those percentages. We additionally observed that the weighted threshold curve is strongly curve in the low percentages, as opposed to the curve for smaller image. Moreover, the latter curve does not exhibit the asymptotic coincidence with visible band curve in this range of percentages (up to 20%). Overall, we conclude that (1) the image size does have a significant effect on the overall thresholds, (2) nonetheless, when counts and thresholds are looked at with respect to the pertinent variables, i.e. number and percentage of A pixels, respectively, the effect of image size exhibits a certain predictability. In particular, only the threshold curve of the visible band (550 nm – 650 nm) appears to be visibly affected: that curve appears to be well approximated by a straight line, whose slope and intercept change by the image size. In what follows, the overall mean threshold will be regarded as defining the classification criteria. The performance of these criteria is assessed by applying them to test synthetic images in the next section.

IV. ASSESSMENT OF CLASSIFICATION CRITERIA

A. A SIMPLE CLASSIFICATION CRITERION

We would like now to turn the mean threshold curves over both bands, shown in blue in Figure 15, into practical classification criteria, and assess their performance.

To this end, consider a synthetic hyperspectral image of 100 or 500 pixels, generated for testing purposes, and we shall call it a test image. That is, the type (H or A) of each one of its pixels is known, but this information will not be used at first. Then, we proceed as follows.

- Step 1. For each pixel, we calculate the average value of the function index(pixel, λ), defined in equation (2), over the both the visible (V) and red-edge bands (RE), defined in Table 4 as the intervals 550-650 nm and 680-700 nm, respectively. This mean value will be denoted as index(pixel):

$$\text{index(pixel)} = \frac{1}{120} \sum_{\lambda \in V \cup RE} \text{index(pixel},\lambda) \qquad (6)$$

  These bands contain a total of 120 wavelengths, hence the factor in equation (6).
- Step 2. We order the numbers index(pixel) in an increasing order, keeping track of the pixel each of these numbers is associated with.
- Step 3. We plot the ordered values of the index, that is, essentially a rank-order plot, together with mean threshold overall curve (blue curves in Figure 15) corresponding to image size we have chosen. Given that the number of the A pixels in the test image is regarded as unknown at this stage, the horizontal axis is a rank-order label.
- Step 4. We find where the rank-order curve of the test image crosses first the (theoretical) overall mean threshold curve (to the nearest integer value of the horizontal axis). We introduce the following empirical criterion: We regard all points of the rank-order curve up and including the crossing point as being associated with A pixels. Put differently, we regard all index values below the theoretical curve as being associated with A pixels.

Given that the theoretical (overall mean threshold) curves have been constructed beforehand for any given image size, it is straightforward to apply this procedure to any given test image and identify certain pixels as A type. Also, given that we keep track of A or H type of pixels, we can assess the performance of the proposed criterion.

The flow of work that implements our methodology, once the physically-based simulations datasets have been produced, is present in Figure 16.

B. QUANTITATIVE ASSESSMENT OF THE PROPOSED CRITERION

We start off by generating test images for both image size of 100 and 500 pixels, with different numbers of A pixels. To obtain representative results, we generate small batches of 20 test images for each case of image size and number of A pixels, in order to be able to average fluctuating quantities, such as the number of predicted A pixels, for images of the same size and A pixels content.

Figures 17 and 18 show the rank-order curves for the index of the test images together with the theoretical mean threshold curves for the two image sizes of 100 and 500 pixels, respectively. The different rank-order curves (in dashed lines) correspond to different numbers of A pixels in the image.

$$\text{overall mean threshold} = \frac{(\text{threshold for V}) \times (\text{counts of V}) + (\text{threshold for RE}) \times (\text{counts of RE})}{(\text{counts of V}) + (\text{counts of RE})} \qquad (5)$$

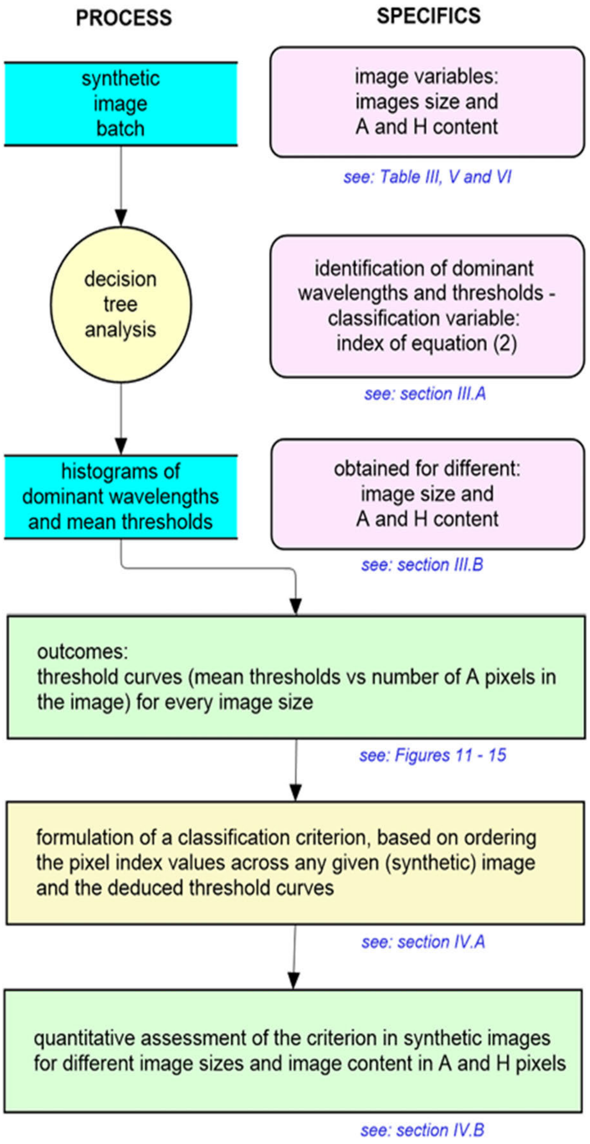

FIGURE 16. **Flowchart of the proposed methodology once the synthetic datasets are produced.**

Specifically, we plot the average of the rank-order curves of the 20-image batch for each given number of A pixels. As suggested above, the point where each one of these curves (to the nearest integer number of pixels) the theoretical curves (blue lines) correspond to the predicted number of A pixels in the test image: all points up and including the crossing point are labelled as A pixels. These plots illustrate in a visual manner the way the synthetic hyperspectral images index values behave. The crossing, or better the nearest point, between the rank-order and theoretical curves can also be found algorithmically. Table 5 shows these results for the 100 pixels image case. The second column shows the number of predicted A pixels according to the crossing point, for every given number of A pixels in the image, while the third column is percentage of the predicted to the actual number of A. The fourth column provides an assessment score: The mean number of pixels predicted as A which are actually A, for the different cases of numbers of A pixels in the image. The fifth column shows the proportion of the correctly predicted (detected) to the actual number of A. The second column shows that, as the image contains more A pixels, the criterion has a decreasing rate of success, albeit fairly high. On the other hand, as it is shown in the fourth column of Table 5, it does have a consistently fairly high degree of success in correctly identifying the A pixels. For example, for 20 A pixels in the image of 100 pixels, the method successfully identifies 14.8 of them on average, that is 75%.

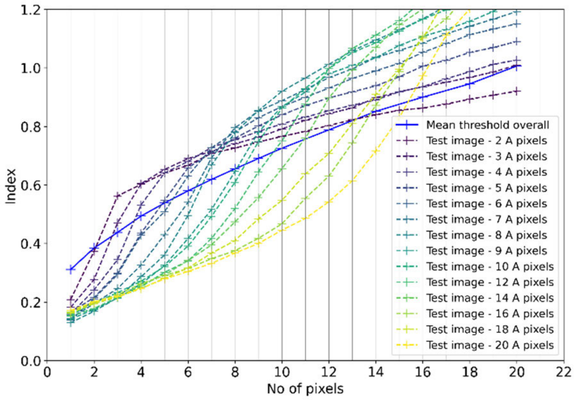

FIGURE 17. **Rank-order curves (dashed lines) for the 100 pixels images with 2 to 20 type A pixels plotted together with the (theoretical) overall mean threshold curve.**

The results of Table 5 show that there are limits to the performance of the method. That can be attributed, in part, to the assumption that we have looked for a one-dimensional decision boundary in the decision-tree classification process, i.e. a mere threshold for a single dominant wavelength every time, while the actual decision boundary is certainly higher dimensional. Another reason is that the averaging of thresholds over each band is done in a uniform manner for all different numbers of A pixels, and not using the detailed information of the distributions of the dominant wavelengths (such as the ones shown in Figures 11 and 12) within the bands, which are different for different numbers of A pixels.

In Figure 18 the rank-order curves refer to cases of test images (20-image batches) with a range from 2 to 50 A pixels in the 500 pixels of each image. The rank-order curves for these larger images exhibit a similar pattern between themselves as well as relative to the theoretical curve (overall mean threshold curve) of the 500 pixels image. We observe, for example, that the 2 A rank-order curve crosses for a second time the theoretical curve at 55 pixels, i.e. at 11% of the image size, while in the case of the 100 pixels image, the 2 A rank-order curve crosses for a second time the theoretical curve at 13 pixels, i.e. 13% of the image size.

Table 6 summarizes the predicted and mean value of the correctly predicted A pixels for each case from the 2 to 50 A pixels in the 500 pixels images. As in the case of smaller image, we observe that the rate of successful identification of the A pixels is consistently quite high, in the neighborhood of 75%. A common feature of the performance of

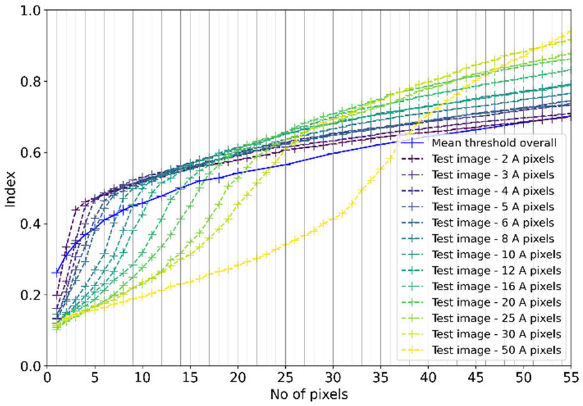

FIGURE 18. Rank-order curves (dashed lines) for the 500 pixels images with 2 to 50 type A pixels plotted together with the (theoretical) overall mean threshold curve.

TABLE 5. Assessment results for the 100 pixels test image.

No of A	No of predicted A	Prediction rate (%)	Mean of detected A	Detection rate (%)
2	2	100.0	1.6	80.0
3	3	100.0	2.4	80.0
4	4	100.0	3.2	80.0
5	5	100.0	4.3	86.0
6	5	83.3	4.6	76.7
7	6	85.7	5.4	77.1
8	7	87.5	6.2	77.5
9	8	88.9	7.2	80.0
10	8	80.0	7.4	74.0
12	10	83.3	9.5	79.2
14	11	78.6	10.4	74.3
16	14	87.5	13.4	83.8
18	13	72.2	12.6	70.0
20	15	75.0	14.8	74.0

TABLE 6. Assessment results for the 500 pixels test image.

No of A	No of predicted A	Prediction rate (%)	Mean of detected A	Detection rate (%)
2	2	100.0	1.4	70.0
3	3	100.0	2.4	80.0
4	4	100.0	3.2	80.0
5	5	100.0	3.6	72.0
6	5	83.3	4.4	73.3
8	7	87.5	6.4	80.0
10	9	90.0	7.8	78.0
12	11	91.7	9.6	80.0
16	14	87.5	12.6	78.8
20	17	85.0	15.5	77.5
25	21	84.0	19.1	76.4
30	24	80.0	22.6	75.3
50	37	74.0	36.6	73.2

the classification criterion for both image sizes, is that the detection rate decreases slowly with the number of A pixels in the image, but in a way that for realistic proportions of A pixels in the image remains relatively high.

V. SUMMARY AND CONCLUSION

In this work we extend previous research of the authors which proposed an algorithm to identify the imprint on hyperspectral signatures of crops from buried archaeological remains (cropmarks) analyzing a dataset of roughly a hundred measured signatures from an artificial test site. In that work, we introduced a new index and classification (identification) criteria for cropmarks, analyzing the dataset through machine learning (decision-trees), by which one-dimensional decision boundaries (thresholds) for pertinent wavelength ranges were found.

The present work aims at introducing a simple classification, expressed in terms of spectral information based on hyperspectral data, that could be easily transferred, tested and assessed in different conditions (crop type, soil properties, climatic conditions, period of observation, etc) in the future. This can have a major impact on existing archaeological prospection practices as it will establish, for the first time, regional remote sensing physically-based models for the detection of cropmarks (and therefore, archaeological proxies). To this end, we extend the previously obtained results in several ways.

- We enlarge the dataset by physically-based simulations of the observed signatures. This is done by using canopy radiative transfer model (PROSAIL) inversion to reduce the observed signatures to a few biophysical and biochemical parameter values. Inversion was done using iterative optimization by standard Python language optimizers. The statistics of the physical parameters dataset for stressed (A) and healthy (H) signatures (that is, marginal p.d.f.s and covariance matrix) allow generating -arbitrarily many- simulations of this dataset. Running PROSAIL in the forward mode we may generate any desired number of physically-based simulations, whose type, A or H, is known. Given that no ground-truth data were available for validation of our PROSAIL inversion results, our analysis should be understood as a plausible physical reduction of degrees of freedom underlying the observed signatures.
- Designating, in a rather descriptive manner, a simulated hyperspectral signature as a 'pixel' and a collection of them as a 'synthetic image', we investigate the influence of (1) the image size, and (2) the number of type A pixels present in it, on the threshold values that allow classifying a given pixel as A or H. This is done by following the steps of the previous work, applying decision tree algorithms to identify dominant wavelengths and the associated one-dimensional decision boundaries (thresholds). This analysis produces a suitable range of dominant wavelengths, 550 – 650 nm and 680 – 700 nm, over which the mean value of threshold is found and recorded for given image size and number of type A pixels in it. For a given image size, these results form a (theoretical) threshold curve against the number of A pixels. This is done for image sizes of 100 and 500 pixels for a range of different numbers of A pixels.
- A classification criterion is proposed that allows the identification of type A pixels in any given synthetic image (a physically-based simulation called a test

image) whose A pixels are previously unknown. This is done by calculating for every pixel the mean value of our index over all wavelengths in the ranges 550 – 650 nm and 680 – 700 nm, putting these values in an increasing order (rank-order curve), and finding the crossing point of this curve with the theoretical threshold curve, for the given image size. Then we regard all pixels below and including the crossing point as type pixels. An assessment analysis for both image sizes shows that this criterion successfully identifies the A pixels with a rate above 70%.

Within the context of this work, a main limitation of the proposed procedure is that the decision boundary has been reduced to one dimension, i.e. identifying the wavelength with the highest importance and its threshold, ignoring other wavelengths with much lower but not negligible importances. Nonetheless, this simplification appears to work rather well, and it allows us to identify the stressed signatures by very simple rules: one merely needs a threshold curve for a given image size. Given that the aim is to deduce criteria generalizable to different conditions, this simplicity might indeed be an advantage.

The last observation justifies our specific use of decision-trees, and machine learning in general, in the present work. Regarding interpretability, decision-trees are classified as white box models, because the algorithm is intuitively intelligible, and the feature importance is explicitly calculable. Even in the context of these algorithms, a higher dimensional decision boundary is not easy to describe explicitly, and it is far more susceptible to being case specific. These observations hold more strongly for lower transparency algorithms, such as random forests or artificial neural networks. From the standpoint of machine learning, the aim of the present analysis may be viewed as putting forward an objective for the advanced algorithms: one should search for threshold curves for a certain index, as explained in previous sections. One should bear in mind that, in the cropmark detection research, the original (ground-truth) datasets are never quite large enough for brute force application of machine learning algorithms to naturally produce generalizable results. In the present work we have explored the characteristics of the original dataset by physically-based simulations. The synthetic datasets obtained, are rich enough in information for finding simple detection criteria, but not obviously rich enough to usefully apply powerful classification algorithms.

An aspect of the implementation of the method that requires some discussion is the way we average the thresholds values over the empirically chosen bands. This is done in a uniform manner within each band, for all different numbers of A pixels. Another choice would be to use the theoretical histograms of the dominant wavelengths distributions as filters. These histograms are different for different numbers of A pixels. Using a uniform filter, the only information used in the calculation of the overall mean value of the threshold is the total number of counts corresponding to each band. This choice, which is done for the sake of simplicity and potentially greater applicability of the proposed classification criterion, led indeed to a simple but effective method.

Another important point is that the classification analysis and the statement of the criterion is done in terms of spectral information. Given that, there is always a set of physical parameters associated with each pixel (physically-based simulation of signature), both the analysis and the criterion statement could very well be done in terms of the physical parameters. The difference is that while the signatures are measurable quantities, by various means, the PROSAIL parameters are not all measurable, and in most cases, there is no information even about the measurable ones. Hence, as mentioned above, they are merely treated here as a physically-based way to reduce the degrees of freedom governing the hyperspectral data. Nonetheless, it should be noted that specific physical information, such as known ranges and correlations of parameters, as well as consistency with the growth stage of the crop, has been taken into account in the inversion process.

Finally, we may note that our analysis and results are based on the index introduced in the previous work of the authors mentioned above, which is a fraction of differences, and amounts to transforming an observed or simulated signature in the unit interval in the wavelengths between the deep blue and the near-infrared. This is presumably the reason why the near-infrared does not arise as a significant range of wavelengths in our results. It should be clear that other indices, of similar or entirely different kind, could be used and be also at least as effective. Such investigations will be left for future work. Moreover, part of the definition of the index, involves an averaging over the entire image. This part of the definition is similar to defining and working with dimensionless quantities instead of dimensionful ones, which in general allow obtaining results with enormous generality. At the same time, this averaging over the image tacitly assumes a certain degree of homogeneity present in the pixels of the image, e.g. same crop, as was the case with the studied test field.

Assessment of generalizability of the proposed methodology involves studying the sensitivity of the threshold curves to different archaeological contexts and crop-soil conditions, perhaps refining also the band filter definitions. will necessarily be left for future work. The fact that the different conditions are encoded in a few physical parameters is rather reassuring. The detailed work needed for this assessment will necessarily be left for future research.

ACKNOWLEDGMENT

(Elias Gravanis and Athos Agapiou contributed equally to this work.)

REFERENCES

[1] L. Luo, X. Wang, H. Guo, X. Jia, and A. Fan, “Earth observation in archaeology: A brief review,” *Int. J. Appl. Earth Observ. Geoinf.*, vol. 116, Feb. 2023, Art. no. 103169.

[2] R. Opitz and J. Herrmann, “Recent trends and long-standing problems in archaeological remote sensing,” *J. Comput. Appl. Archaeol.*, vol. 1, no. 1, pp. 19–41, May 2018.

[3] G. Verhoeven, "Are we there yet? A review and assessment of archaeological passive airborne optical imaging approaches in the light of landscape archaeology," *Geosciences*, vol. 7, no. 3, p. 86, Sep. 2017.

[4] M. Zingaro, G. Scicchitano, and D. Capolongo, "The innovative growth of space archaeology: A brief overview of concepts and approaches in detection, monitoring, and promotion of the archaeological heritage," *Remote Sens.*, vol. 15, no. 12, p. 3049, Jun. 2023.

[5] A. Agapiou, A. Hegyi, and A. Stavilă, "Observations of archaeological proxies through phenological analysis over the megafort of Csanádpalota-Juhász T. Tanya in Hungary using sentinel-2 images," *Remote Sens.*, vol. 15, no. 2, p. 464, Jan. 2023.

[6] N. Abate, A. Elfadaly, N. Masini, and R. Lasaponara, "Multitemporal 2016–2018 sentinel-2 data enhancement for landscape archaeology: The case study of the foggia province, southern Italy," *Remote Sens.*, vol. 12, no. 8, p. 1309, Apr. 2020.

[7] M. Barber, *A History of Aerial Photography and Archaeology: Mata Hari's Glass Eye and Other Storiesm*, 1st ed. London, U.K.: Historic England, 2011.

[8] D. C. Cowley, "Creating the cropmark archaeological record in East Lothian, Southeast Scotland," in *Prehistory Without Borders: Prehistoric Archaeology of the Tyne-Forth Region*. New York, NY, USA: Oxford, 2016, pp. 59–70.

[9] Ł. Banaszek, D. C. Cowley, and M. Middleton, "Towards national archaeological mapping. Assessing source data and methodology—A case study from Scotland," *Geosciences*, vol. 8, no. 8, p. 272, Jul. 2018.

[10] S. Campana, "'Total archaeology' to reduce the need for rescue archaeology: The BREBEMI project (Italy)," *Remote Sens. Archaeological Heritage Manage.*, vol. 2011, pp. 33–41, Apr. 2011.

[11] C. Gaffney, V. Gaffney, W. Neubauer, E. Baldwin, H. Chapman, P. Garwood, H. Moulden, T. Sparrow, R. Bates, K. Löcker, A. Hinterleitner, I. Trinks, E. Nau, T. Zitz, S. Floery, G. Verhoeven, and M. Doneus, "The stonehenge hidden landscapes project," *Archaeological Prospection*, vol. 19, no. 2, pp. 147–155, Apr. 2012.

[12] X. Zhang, C. Leng, Y. Hong, Z. Pei, I. Cheng, and A. Basu, "Multimodal remote sensing image registration methods and advancements: A survey," *Remote Sens.*, vol. 13, no. 24, p. 5128, Dec. 2021.

[13] T. Kalaycı and A. Sarris, "Multi-sensor geomagnetic prospection: A case study from neolithic thessaly, Greece," *Remote Sens.*, vol. 8, no. 11, p. 966, Nov. 2016.

[14] A. Agapiou and A. Sarris, "Beyond GIS layering: Challenging the (Re)use and fusion of archaeological prospection data based on Bayesian neural networks (BNN)," *Remote Sens.*, vol. 10, no. 11, p. 1762, Nov. 2018.

[15] Z. Czajlik, M. Árvai, J. Mészáros, B. Nagy, L. Rupnik, and L. Pásztor, "Cropmarks in aerial archaeology: New lessons from an old story," *Remote Sens.*, vol. 13, no. 6, p. 1126, Mar. 2021.

[16] M. Gojda and M. Hejcman, "Cropmarks in main field crops enable the identification of a wide spectrum of buried features on archaeological sites in central Europe," *J. Archaeological Sci.*, vol. 39, no. 6, pp. 1655–1664, Jun. 2012.

[17] B. Molloy, D. Jovanović, C. Bruyere, M. Estanqueiro, M. Birclin, L. Milašinović, A. Šalamon, K. Penezić, C. B. Ramsey, and D. Grosman, "Resilience, innovation and collapse of settlement networks in later bronze age Europe: New survey data from the southern carpathian basin," *PLoS ONE*, vol. 18, no. 11, Nov. 2023, Art. no. e0288750.

[18] B. H. Menze and J. A. Ur, "Mapping patterns of long-term settlement in northern Mesopotamia at a large scale," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 14, pp. E778–E787, Apr. 2012.

[19] S. Jacquemoud, W. Verhoef, F. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, C. François, and S. L. Ustin, "PROSPECT + SAIL models: A review of use for vegetation characterization," *Remote Sens. Environ.*, vol. 2009, pp. 56–66, Jul. 2009.

[20] A. Agapiou and E. Gravanis, "A machine-learning-assisted classification algorithm for the detection of archaeological proxies (Cropmarks) based on reflectance signatures," *Remote Sens.*, vol. 16, no. 10, p. 1705, May 2024.

[21] A. Agapiou, "Development of a novel methodology for the detection of buried archaeological remains using remote sensing techniques," Ph.D. thesis, Dept. Civil Eng. Geomatics, Cyprus Univ. Technol., Limassol, Cyprus, May 8, 2013. [Online]. Available: https://hdl.handle.net/20.500.14279/877

[22] R. B. Myneni, J. Ross, and G. Asrar, "A review on the theory of photon transport in leaf canopies," *Agricult. Forest Meteorol.*, vol. 45, nos. 1–2, pp. 1–153, Feb. 1989.

[23] K. Berger, C. Atzberger, M. Danner, G. D'Urso, W. Mauser, F. Vuolo, and T. Hank, "Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study," *Remote Sens.*, vol. 10, no. 1, p. 85, Jan. 2018.

[24] W. Verhoef, L. Jia, Q. Xiao, and Z. Su, "Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1808–1822, Jun. 2007.

[25] W. A. Dorigo, R. Zurita-Milla, A. J. W. de Wit, J. Brazile, R. Singh, and M. E. Schaepman, "A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 9, no. 2, pp. 165–193, May 2007.

[26] B. Pinty et al., "Radiation transfer model intercomparison (RAMI) exercise: Results from the second phase," *J. Geophys. Res., Atmos.*, vol. 109, no. D6, pp. 11937–11956, Mar. 2004.

[27] J.-L. Widlowski et al., "The fourth phase of the radiative transfer model intercomparison (RAMI) exercise: Actual canopy scenarios and conformity testing," *Remote Sens. Environ.*, vol. 169, pp. 418–437, Nov. 2015.

[28] N. H. Broge and E. Leblanc, "Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density," *Remote Sens. Environ.*, vol. 76, no. 2, pp. 156–172, May 2001.

[29] F. Baret and S. Buis, "Estimating canopy characteristics from remote sensing observations: Review of methods and associated problems," in *Advances in Land Remote Sensing*, S. Liang, Ed. Dordrecht, The Netherlands: Springer, 2008, doi: 10.1007/978-1-4020-6450-0_7.

[30] J. Verrelst, G. Camps-Valls, J. Muñoz-Marí, J. P. Rivera, F. Veroustraete, J. G. P. W. Clevers, and J. Moreno, "Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties—A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 108, pp. 273–290, Oct. 2015.

[31] A. Abdelbaki and T. Udelhoven, "A review of hybrid approaches for quantitative assessment of crop traits using optical remote sensing: Research trends and future directions," *Remote Sens.*, vol. 14, no. 15, p. 3515, Jul. 2022.

[32] D. S. Kimes, Y. Knyazikhin, J. L. Privette, A. A. Abuelgasim, and F. Gao, "Inversion methods for physically-based models," *Remote Sens. Rev.*, vol. 18, nos. 2–4, pp. 381–439, Sep. 2000.

[33] J. Verrelst, Z. Malenovský, C. Van der Tol, G. Camps-Valls, J.-P. Gastellu-Etchegorry, P. Lewis, P. North, and J. Moreno, "Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods," *Surv. Geophys.*, vol. 40, no. 3, pp. 589–629, May 2019.

[34] L. Liang, L. Di, L. Zhang, M. Deng, Z. Qin, S. Zhao, and H. Lin, "Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method," *Remote Sens. Environ.*, vol. 165, pp. 123–134, Aug. 2015.

[35] M. Weiss and F. Baret. (2019). *S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER*. [Online]. Available: https://step.esa.int/docs/extra/ATBD_S2ToolBox_L2B_V1.1.pdf

[36] T. Kattenborn, J. Leitloff, F. Schiefer, and S. Hinz, "Review on convolutional neural networks (CNN) in vegetation remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 173, pp. 24–49, Mar. 2021.

[37] Y. Zhong, X. Hu, C. Luo, X. Wang, J. Zhao, and L. Zhang, "WHU-hi: UAV-borne hyperspectral with high spatial resolution (H_2) benchmark datasets and classifier for precise crop identification based on deep convolutional neural network with CRF," *Remote Sens. Environ.*, vol. 250, Dec. 2020, Art. no. 112012.

[38] S. Bhadra, V. Sagan, S. Sarkar, M. Braud, T. C. Mockler, and A. L. Eveland, "PROSAIL-net: A transfer learning-based dual stream neural network to estimate leaf chlorophyll and leaf angle of crops from UAV hyperspectral images," *ISPRS J. Photogramm. Remote Sens.*, vol. 210, pp. 1–24, Apr. 2024.

[39] C. Atzberger, M. Guérif, F. Baret, and W. Werner, "Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat," *Comput. Electron. Agricult.*, vol. 73, no. 2, pp. 165–173, Aug. 2010.

[40] J. P. Rivera-Caicedo, J. Verrelst, J. Muñoz-Marí, G. Camps-Valls, and J. Moreno, "Hyperspectral dimensionality reduction for biophysical variable statistical retrieval," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 88–101, Oct. 2017.

[41] *Python Programming Language*. Accessed: Mar. 14, 2024. [Online]. Available: https://www.python.org/

[42] M. Danner, K. Berger, M. Wocher, W. Mauser, and T. Hank, "Fitted PROSAIL parameterization of leaf inclinations, water content and Brown pigment content for winter wheat and maize canopies," *Remote Sens.*, vol. 11, no. 10, p. 1150, May 2019.

[43] H. K. Lichtenthaler, "Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes," in *Methods in Enzymology*, vol. 148. New York, NY, USA: Academic, 1987, pp. 350–382.

[44] F. Baret, B. Andrieu, and G. Guyot, "A simple model for leaf optical properties in visible and near-infrared: Application to the analysis of spectral shifts determinism," in *Applications of Chlorophyll Fluorescence in Photosynthesis Research, Stress Physiology, Hydrobiology and Remote Sensing*. Dordrecht, The Netherlands: Springer, Jun. 1988, pp. 345–351.

[45] J.-B. Féret, C. François, A. Gitelson, G. P. Asner, K. M. Barry, C. Panigada, A. D. Richardson, and S. Jacquemoud, "Optimizing spectral indices and chemometric analysis of leaf chemical properties using radiative transfer modeling," *Remote Sens. Environ.*, vol. 115, no. 10, pp. 2742–2750, Oct. 2011.

[46] M. Landi, M. Tattini, and K. S. Gould, "Multiple functional roles of anthocyanins in plant-environment interactions," *Environ. Experim. Botany*, vol. 119, pp. 4–17, Nov. 2015.

[47] N. J. J. Breda, "Ground-based measurements of leaf area index: A review of methods, instruments and current controversies," *J. Experim. Botany*, vol. 54, no. 392, pp. 2403–2417, Sep. 2003.

[48] G. P. Asner, J. M. O. Scurlock, and J. A. Hicke, "Global synthesis of leaf area index observations: Implications for ecological and remote sensing studies," *Global Ecology Biogeography*, vol. 12, no. 3, pp. 191–205, May 2003.

[49] H. Fang, F. Baret, S. Plummer, and G. Schaepman-Strub, "An overview of global leaf area index (LAI): Methods, products, validation, and applications," *Rev. Geophys.*, vol. 57, no. 3, pp. 739–799, Sep. 2019.

[50] G. Zheng and L. M. Moskal, "Retrieving leaf area index (LAI) using remote sensing: Theories, methods and sensors," *Sensors*, vol. 9, no. 4, pp. 2719–2745, Apr. 2009.

[51] M. C. González-Sanpedro, T. Le Toan, J. Moreno, L. Kergoat, and E. Rubio, "Seasonal variations of leaf area index of agricultural fields retrieved from Landsat data," *Remote Sens. Environ.*, vol. 112, no. 3, pp. 810–824, Mar. 2008.

[52] T. Dong, J. Liu, J. Shang, B. Qian, B. Ma, J. M. Kovacs, D. Walters, X. Jiao, X. Geng, and Y. Shi, "Assessment of red-edge vegetation indices for crop leaf area index estimation," *Remote Sens. Environ.*, vol. 222, pp. 133–143, Mar. 2019.

[53] J. Borzuchowski and K. Schulz, "Retrieval of leaf area index (LAI) and soil water content (WC) using hyperspectral remote sensing under controlled glass house conditions for spring barley and sugar beet," *Remote Sens.*, vol. 2, no. 7, pp. 1702–1721, Jul. 2010.

[54] G. Fontanelli, S. Paloscia, M. Zribi, and A. Chahbi, "Sensitivity analysis of X-band SAR to wheat and barley leaf area index in the merguellil basin," *Remote Sens. Lett.*, vol. 4, no. 11, pp. 1107–1116, Nov. 2013.

[55] Y. Afrasiabian, H. Noory, A. Mokhtari, M. R. Nikoo, F. Pourshakouri, and P. Haghighatmehr, "Effects of spatial, temporal, and spectral resolutions on the estimation of wheat and barley leaf area index using multi- and hyper-spectral data (case study: Karaj, Iran)," *Precis. Agricult.*, vol. 22, no. 3, pp. 660–688, Jun. 2021.

[56] P. Rosso, C. Nendel, N. Gilardi, C. Udroiu, and F. Chlebowski, "Processing of remote sensing information to retrieve leaf area index in barley: A comparison of methods," *Precis. Agricult.*, vol. 23, no. 4, pp. 1449–1472, Aug. 2022.

[57] L. Xun, P. Wang, L. Li, L. Wang, and Q. Kong, "Identifying crop planting areas using Fourier-transformed feature of time series MODIS leaf area index and sparse-representation-based classification in the north China plain," *Int. J. Remote Sens.*, vol. 40, nos. 5–6, pp. 2034–2052, Mar. 2019.

[58] X. Zhou, P. Wang, K. Tansey, S. Zhang, H. Li, and H. Tian, "Reconstruction of time series leaf area index for improving wheat yield estimates at field scales by fusion of sentinel-2, -3 and MODIS imagery," *Comput. Electron. Agricult.*, vol. 177, Oct. 2020, Art. no. 105692.

[59] Y. Liu, L. Su, Q. Wang, J. Zhang, Y. Shan, and M. Deng, "Chapter six—Comprehensive and quantitative analysis of growth characteristics of winter wheat in China based on growing degree days," in *Advances in Agronomy*, vol. 159, D. L. Sparks, Ed. Academic, 2020, pp. 237–273.

[60] A. Kuusk, "The hot spot effect of a uniform vegetative cover," *Soviet J. Remote Sens.*, vol. 3, no. 4, pp. 645–658, 1985.

[61] B. Hapke, "Bidirectional reflectance spectroscopy: 4. the extinction coefficient and the opposition effect," *Icarus*, vol. 67, no. 2, pp. 264–280, 1986.

[62] W. Verhoef, "Theory of radiative transfer models applied in optical remote sensing of vegetation canopies," Ph.D. thesis, Fac. Geo-Inf. Sci. Earth Observ., Wageningen Univ. Res., Wageningen, The Netherlands, 1998.

[63] S. B. Kotsiantis, "Decision trees: A recent overview," *Artif. Intell. Rev.*, vol. 39, no. 4, pp. 261–283, Apr. 2013.

[64] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: An updated survey," *Artif. Intell. Rev.*, vol. 56, no. 5, pp. 4765–4800, May 2023.

ELIAS GRAVANIS received the bachelor's degree in physics from the University of Athens, Greece, in 1997, the master's degree from Columbia University, in 2000, and the Ph.D. degree in mathematical physics from the University of London (King's College London), in 2006. He is currently a Special Teaching Stuff with the Civil Engineering and Geomatics Department, Cyprus University of Technology. Since 2010, he has been with the Department of Civil Engineering and Geomatics, Cyprus University of Technology, on various projects in fluid mechanics, theory, and modeling of turbulence and groundwater hydrology. He also works on research projects in the context of geomechanics, geotechnical applications, and fluid transport in porous media, including geological CO2 sequestration. His research interests include inverse methods in hydrogeological problems, utilizing stochastic inversion techniques, and in the context of critical groundwater reserves assessment and management. More recently, he has been working on projects involving machine learning and remote sensing. Additionally, he teaches mathematics, physics, hydraulics, hydrology, and water resources management with the Department of Civil Engineering and Geomatics.

ATHOS AGAPIOU received the Diploma degree from the School of Rural and Surveying Engineering, National Technical University of Athens, in 2005, the M.Sc. degree in geoinformatics from the National Technical University of Athens, in 2007, the M.A. degree from the Department of History and Archaeology, University of Cyprus, in 2010, and the Ph.D. degree from Cyprus University of Technology, in 2013. He has been an Assistant Professor with the Department of Civil Engineering and Geomatics, Cyprus University of Technology, since July 2021. During the last years, he has been involved in several research projects at a European (H2020), regional (JPI-CH; Interreg), and local level (Research and Innovation Foundation, A.G. Leventis). His scientific and research interests include earth observation and geoinformatics intended for archaeological research and cultural heritage landscapes. More information regarding running projects and publications can be found on the webpage of the Earth Observation Cultural Heritage Laboratory (https://web.cut.ac.cy/eocult/).

• • •