Anomaly Detection and Classification in Multispectral Time Series based on Hidden Markov Models

Kareth León-López, *Student Member, IEEE,* Florian Mouret, Henry Arguello, *Senior, IEEE,* Jean-Yves Tourneret, *Fellow, IEEE*

***Abstract*—Monitoring agriculture from satellite remote sensing data, such as multispectral images, has become a powerful tool since it has demonstrated a great potential for providing timely and accurate knowledge of crops. Many studies have been devoted to the early detection of anomalies in time series of multispectral remote sensing images but few of them have considered the identification of the abnormal factors that are damaging the vegetation. This paper presents a framework for anomaly detection and classification exploiting the temporal information contained in remote sensing time series using hidden Markov models (HMM) and machine learning. The anomaly detection part is based on the learning of HMM parameters associated with unlabeled nominal data, that are used to detect abnormal crop parcels referred to as anomalies. The learned HMM are then used in time segments to temporally localize the anomalies affecting the crop parcels. The detected and localized anomalies are finally classified using a supervised classifier, e.g., based on support vector machines. Numerical experiments are conducted on synthetic and real data to evaluate the performance of the proposed algorithm. The real data corresponds to vegetation indices extracted from several multi-temporal Sentinel-2 images of rapeseed crops. The performance of the anomaly detection is evaluated in terms of precision and recall whereas the classification performance is determined using probabilities of correct classification. The proposed approach is compared to standard anomaly detection methods yielding better detection rates with the advantage of allowing anomalies to be localized and characterized.**

***Index Terms*—Anomaly detection, anomaly classification, remote sensing, Hidden Markov models, agricultural monitoring, time series, semi-supervised learning.**

I. Introduction

Multispectral images have been widely used in many studies to explore the vegetation properties of plants through the extraction of vegetation indices [1], [2], [3], [4]. In the last decade, researchers have proposed to use multi-temporal images for several applications including change detection [5] and landcover classification [6], [7], where the challenge is mainly to exploit the redundancy and correlation across the spatial, spectral, and temporal dimensions of the images.

K. León-López and H. Arguello are with the Department of Systems Engineering, Universidad Industrial de Santander, Bucaramanga, Santander, 680002 Colombia (e-mail: kareth.leon@correo.uis.edu.co; henarfu@uis.edu.co). F. Mouret is with TerraNIS SAS and University of Toulouse, IRIT/INP-ENSEEIHT, 31000 Toulouse, France (email: florian.mouret@terranis.fr). J.Y. Tourneret is with the University of Toulouse, IRIT/INP-ENSEEIHT/TeSA, 31071 Toulouse cedex 7, France (e-mail: jean-yves.tourneret@enseeiht.fr).

Hidden Markov models (HMM) are classical tools to analyze time series, allowing temporal correlations to be extracted with the introduction of latent variables interacting with the data [8], [9], [10]. Different works have shown that HMM are valuable tools for modeling the dynamic behavior of crops across time, where the dynamics of vegetation is related to the phenology, chemical nutrients, climatic conditions, or water stress of crops [11], [12]. Some specific tasks for crop analysis based on HMM include crop recognition [13], crop classification [8], [14], and time evolution featuring [12]. In addition, an analysis of the normalized difference vegetation index (NDVI) using the HMM framework is proposed in [11], where the NDVI changes are used to characterize the dynamics of the vegetation during a temporal window.

An important task in crop monitoring is the detection of anomalies that can represent risks for the harvest [3], [15]. Detecting nutrient stresses or drought helps to better understand the management of nutrients and, in turn, leads to reduce cultivation costs and increases crop efficiency [2], [16], [17]. Thus, depending on the kind of detected anomalies, the farmers can take action to reduce the adverse effects of the phenomenon that produces the anomaly response. Anomaly detection (AD) (which includes outlier and novelty detection) is a widely studied problem that relies on the identification of patterns or events that differ from the expected normal behavior of the majority of the data [18]. Some of the most relevant AD techniques include the local outlier factor [19] or its probabilistic version the local outlier probability [20], the isolation forest [21] and the one-class support vector machine (OC-SVM) [22]. Note that when these methods are applied to time series, they do not fully exploit the temporal correlation between consecutive samples.

Some recent works have proposed to include the HMM framework in the AD problem [23], [24]. In particular, in [23], the authors have considered an HMM-based kernel that is used in the traditional OC-SVM method. However, this kernel was defined assuming some specific kind of anomaly, e.g., resulting from mean-value jumps, which is a too strong assumption for crop monitoring. An interesting framework for AD in multivariate time series was proposed in [24], where a set of transformations was used to unify the time series and estimate appropriate features. However, the resulting AD algorithm was trained in a supervised mode, using normal and abnormal labels for the training samples, which are difficult to obtain in most crop monitoring applications. Finally, it

is interesting to mention some other works such as [10], [25], which have demonstrated that crops analyzed at a parcel level from multi-temporal vegetation indices and vegetation phenology provide suitable knowledge of crops across time. However, these studies have not addressed the problem of classifying and identifying possible factors that are damaging the harvest.

This paper introduces a framework for anomaly detection and classification of remote sensing time series based on HMM. The proposed method referred to as AD-HMM learns the normal dynamic behavior of crops using several HMM whose parameters are estimated from nominal data (i.e., data without any anomaly). Abnormal time series are then detected as those being unlikely to have been generated by these HMM. The main advantage of AD-HMM is that the learned HMM can be used for specific time segments of the tested time series, allowing anomalies to be localized during specific time intervals. In a second step, the proposed AD algorithm is complemented by standard classifiers such as SVMs in order to determine the type of detected anomalies.

Figure 1 illustrates the overview of the proposed approach, which is decomposed into four steps, where the small square connecting the inputs is an image preprocessing procedure. In the first step, AD-HMM estimates some HMM by using nominal data. In the second step, the AD algorithm classifies test parcels (parcels that do not belong to the learning set) as normal or abnormal. In a third step, the detected anomalies are localized using a so-called anomaly localization algorithm. The goal of this localization is to determine in which time period the anomaly has occurred. The last step of the proposed methodology classifies anomalies using a supervised classifier, allowing specific kinds of anomalies to be identified, such as crop heterogeneity, early or late growth and senescence (plant degradation) [25]. The anomaly localization and classification steps provide a complementary knowledge to the farmers and producers for making decisions regarding both the time of the vegetation cycle and the crop area.

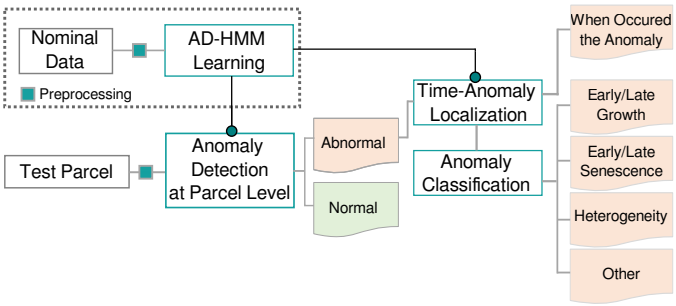

Fig. 1. General overview of the proposed AD and classification approach, where the dotted-square at the top encloses the HMM-based AD referred to as AD-HMM. Based on the learned HMM, AD at the parcel level is used to declare whether a test sample is normal or not, where the AD-HMM considers the temporal structure of the time series associated with the test sample. The detected anomaly is then localized in a specific time interval and finally classified into different types of anomalies that are classically encountered in crops.

II. Proposed Method

This section presents the proposed AD and classification approach, which is summarized in the detailed flowchart depicted in Fig. 2. Note that the gray shaded squares highlight the main steps of the method: (1) image preprocessing yielding features at the parcel level, (2) learning HMM associated with normal parcels referred to as AD-HMM learning, (3) AD at the parcel level (point AD), which includes the localization of anomalies, and (4) anomaly classification. Next subsections describe each procedure following the flowchart of Fig. 2.

A. Image Preprocessing and Feature Extraction

The image preprocessing step requires multi-temporal and multispectral images, and the corresponding parcel boundaries (e.g., resulting from a parcellation database such as the land parcel identification system (LPIS)). A set of temporal vegetation indices (VIs) is extracted from these images. For this study, five vegetation indices derived from the visible, near-infrared (NIR), and short-wave infrared (SWIR) were estimated based on images acquired by the Sentinel-2 sensor[1]. The VIs are summarized in Table I and the corresponding spectral bands are detailed in Table II.

TABLE I
Vegetation indices estimated from multispectral images, where NIR, R, G, SWIR, and Re denotes the near-infrared, red, green, short-wave infrared, and red-edge bands.

Vegetation Index (VI)	Formula
Normalized difference VI	$\text{NDVI} = \frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R}}$
Green-Red VI	$\text{GRVI} = \frac{\text{G} - \text{R}}{\text{G} + \text{R}}$
Normalized difference water Index (SWIR)	$\text{NDWI}_{SWIR} = \frac{\text{NIR} - \text{SWIR}}{\text{NIR} + \text{SWIR}}$
Normalized difference water Index (Green)	$\text{NDWI}_{G} = \frac{\text{G} - \text{NIR}}{\text{G} + \text{NIR}}$
Modified Chlorophyll Absorption Ratio Index using the Optimized Soil Adjusted VI [26]	$\text{MCARI/OSAVI} = \frac{(\text{Re} - \text{IR}) - 0.2(\text{Re} - \text{R})}{(1 + 0.16)\frac{\text{NIR} - \text{R}}{\text{NIR} + \text{R} + 0.16}}$

TABLE II
Spectral bands of the Sentinel-2A multispectral images employed in the VIs Estimation [25].

Spectral band	Band 3 Green	Band 4 Red	Band 5 Red-Edge	Band 8 NIR	Band 11 SWIR
Wavelength (μm)	0,560	0,665	0,705	0,842	1,610
Resolution (m)	10	10	20	10	20

Two statistical indicators, namely the median and interquartile range (IQR), were computed for each temporal VI, where the IQR is defined by the difference between the 75th and 25th percentiles. The motivation for employing statistical indicators for the temporal VIs is that they encompass the mean and dispersion of the VIs with a reduced computational load in the

[1]Sentinel-2 (S2A & S2B) level 2A images were downloaded with a spatial resolution of 10–60 m and a spectral resolution of 13 bands. The theoretical revisit time is 5 days. Images with a cloud coverage greater than 20% were removed from the database.

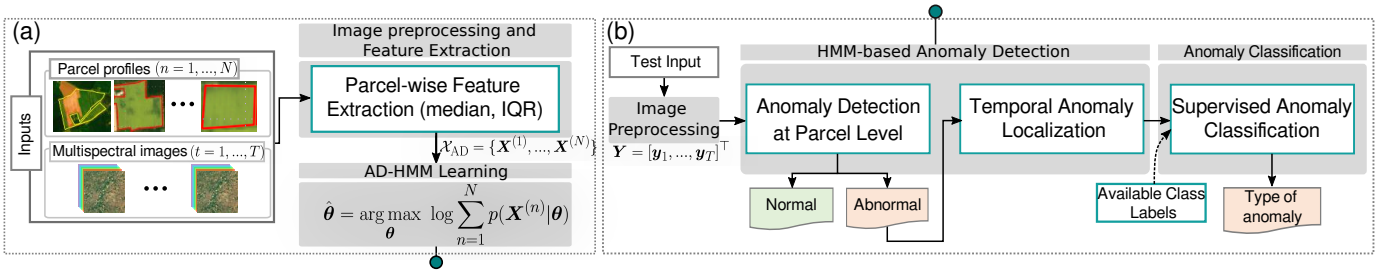

Fig. 2. Flowchart illustrating the main steps and outputs of the proposed approach: (a) Learning step, where multi-temporal/multispectral images and parcel profiles are used to extract features of time series for a given parcel, and (b) Test step. Gray shaded squares indicate the different tasks, namely image preprocessing, AD-HMM learning, AD, anomaly localization and anomaly classification.

data processing. The preprocessing step provides K features for each parcel of the multi-temporal image and for each time instant, i.e., $N \times K \times T$ features, where N is the total number of available parcels and T denotes the number of time instants. To define the extracted time series, let $\boldsymbol{X}^{(n)} = [\boldsymbol{x}_1^{(n)}, ..., \boldsymbol{x}_T^{(n)}]^\top$ denotes all the time series computed for the n-th parcel (with $n = 1, ..., N$), where $\boldsymbol{x}_t^{(n)} \in \mathbb{R}^K$ is the t-th feature vector at time $t \in \{1, ..., T\}$, and K is the number of features. Finally, $\mathcal{X}_{\text{AD}} = \{\boldsymbol{X}^{(1)}, ..., \boldsymbol{X}^{(N)}\}$ is the set of N time series extracted from the normal multispectral images and included in the learning set for AD.

The construction of the set $\mathcal{X}_{\text{AD}}$ is referred to as *parcel-wise feature extraction*, since it extracts statistics from the temporal VIs to build a set of features from the multi-temporal images and their parcel boundaries. The obtained time series are then validated by experts to make sure that they correspond to a normal behavior for AD-HMM learning. Note that the anomalies identified by the expert are excluded from the database and saved (with an anomaly label) in another set denoted as $\mathcal{X}_{\text{AC}}$ for the anomaly classification procedure (see [25] for further details about the database construction).

B. Anomaly Detection based on HMM Learning

1) HMM for Temporal Vegetation Indices: Hidden Markov models (HMM) are doubly stochastic processes defined using an unobservable (hidden) state process, which can be observed via another set of stochastic processes produced by a sequence of observations [9]. HMM allow the characterization of dynamic systems via a set of hidden states $\boldsymbol{s} = \{s_1, ..., s_D\}$ which are inferred from the observations of the system, where D is the number of states in the model. Concisely, an HMM can be formally described by the unknown parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{A}, \boldsymbol{B}\}$, where $\boldsymbol{\pi} \in \mathbb{R}^D$ is the initial probability vector, which defines the initial probabilities of the system to be in the different states; $\boldsymbol{A} \in \mathbb{R}^{D \times D}$ is the transition probability matrix, which relates the state changes of the hidden latent variable; and $\boldsymbol{B}$ is the emission probability matrix, which is the probability of observing a given value in state s.

In particular, given the n-th time series of temporal VIs $\boldsymbol{X}^{(n)}$, the hidden state sequence that reveals a possible state s_i of $\boldsymbol{x}_t^{(n)}$ across time is denoted as $\boldsymbol{Z}^{(n)} = [\boldsymbol{z}_1^{(n)}, ..., \boldsymbol{z}_T^{(n)}]^\top$, with $1 \leq i \leq D$. On the other hand, the entries of the transition probability matrix $\boldsymbol{A}$ are given by $a_{i,j} = P(\boldsymbol{z}_t = s_i | \boldsymbol{z}_{t-1} = s_j)$, which is the probability of transition from a state s_i to the state s_j, for $i, j \in \{1, ..., D\}$. Finally, the entries of the emission probability matrix $\boldsymbol{B}$ are given by $b_{i,t} = P(\boldsymbol{x}_t | \boldsymbol{z}_t = s_i)$, which defines the probability density function of the time-sample $\boldsymbol{x}_t$ at time t given that $\boldsymbol{x}_t$ is in the state s_i. More precisely, in the proposed analysis, the emission probability distribution $\boldsymbol{B}$ is assumed to be a mixture of Gaussian distributions, with M multivariate normal densities. Note that the set of states $\boldsymbol{s}$ is typically related to the changes of the life vegetation cycle [8], [10].

The discrimination of anomalies in the AD-based HMM decides between two hypotheses defined as follows

$$H_0 : \text{Absence of anomaly}$$
$$H_1 : \text{Presence of anomaly},$$

where under hypothesis H_1 a given parcel $\boldsymbol{X}^{(n)}$ is supposed to be abnormal whereas it corresponds to a normal behavior under hypothesis H_0. The likelihood of a given parcel is then defined as

$$\begin{aligned} P(\boldsymbol{X}^{(n)}|\boldsymbol{\theta}) &= \sum_{\text{all } \boldsymbol{Z}^{(n)}} P(\boldsymbol{X}^{(n)}|\boldsymbol{Z}^{(n)}, \boldsymbol{\theta}) P(\boldsymbol{Z}^{(n)}, \boldsymbol{\theta}), \\ &= \sum_{\boldsymbol{z}_1^{(n)}, ..., \boldsymbol{z}_T^{(n)}} \pi_{\boldsymbol{z}_1^{(n)}} b_{\boldsymbol{z}_1^{(n)},1} a_{\boldsymbol{z}_1^{(n)}, \boldsymbol{z}_2^{(n)}} ... a_{\boldsymbol{z}_{T-1}^{(n)}, \boldsymbol{z}_T^{(n)}} b_{\boldsymbol{z}_T^{(n)},T}. \end{aligned} \tag{1}$$

To correctly model and learn the temporal structure of the underlying data for AD, the HMM model parameter vector $\boldsymbol{\theta}$ is estimated by maximizing the log-likelihood, i.e.,

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \log \sum_{n=1}^{N} P(\boldsymbol{X}^{(n)}|\boldsymbol{\theta}), \tag{2}$$

where $\hat{\boldsymbol{\theta}}$ is the parameter vector that better explains $\mathcal{X}$. A local optimal solution of Problem (2) can be found via the Baum-Welch algorithm [9].

2) Generating Different HMM-models: The estimator $\hat{\boldsymbol{\theta}}$ defined in (2) is associated with all the parcels contained in the training set $\mathcal{X}_{\text{AD}}$. In order to account for different possible structures in the underlying data, this paper proposes to build several HMM associated with subsets of N_s samples chosen in $\mathcal{X}_{\text{AD}}$, with $N_s \ll N$. These subsets are built using blocks of time series chosen randomly in $\mathcal{X}_{\text{AD}}$, which leads to L HMM models denoted as $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_L\}$, with $\hat{\boldsymbol{\theta}}_\ell = \{\boldsymbol{\pi}^{(\ell)}, \boldsymbol{A}^{(\ell)}, \boldsymbol{B}^{(\ell)}\}$ for $\ell = 1, ..., L$. These subsets of time series will be denoted as $\{\mathcal{X}^\ell\}_{\ell=1}^L$ with $\mathcal{X}^\ell \in \mathbb{R}^{N_s \times K \times T}$ and $\mathcal{X}_{\text{AD}} = \bigcup_{\ell=1}^L \mathcal{X}^\ell$. The choice of parameter L will be discussed

in Section III.

3) AD-HMM Learning Algorithm: Algorithm 1 summarizes the main steps of the proposed AD-HMM learning, corresponding to the second gray box in Fig. 2(a). This algorithm receives the set of time series parcels $\mathcal{X}_{\text{AD}}$, the number N_s of subsets used to learn a single model, the number of the HMM states D, and the number of models to be learned L. Default values resulting from simulations conducted on real images are provided for each parameter. In the initialization step (Line 1), the number of Gaussian distributions M used in the mixtures is estimated directly from the input data. A strategy to estimate this number of Gaussians is to use the number of local maxima in the data histogram and to reduce this number until the algorithm performance decreases significantly. In the next step, a random index selection of N_s time series stacked in $\mathcal{X}^\ell$ is performed, indicating the parcels to be selected from the set $\mathcal{X}_{\text{AD}}$, as detailed in the previous subsection. The HMM model parameters $\boldsymbol{\theta}$ are then randomly initialized [9]. Finally, the HMM model parameters are estimated using the Baum-Welch procedure and stacked into the set $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_L\}$, where $\hat{\boldsymbol{\theta}}_\ell = \{\boldsymbol{\pi}^{(\ell)}, \mathbf{A}^{(\ell)}, \mathbf{B}^{(\ell)}\}$.

Algorithm 1: AD-HMM Learning Procedure

Input: $\mathcal{X}_{\text{AD}}$: Set of time series associated with parcels;
N_s: # of images per model (default $N_s = 100$);
L: # of models to be learned (default $L = 10$);
D : # of states (default $D = 18$);

1 **Initialize:** M: Estimate the number of Gaussian mixtures;
2 **for** $\ell = 1$ **to** L **do**
3 Built $\mathcal{X}^\ell$ by randomly selecting N_s parcels in $\mathcal{X}_{\text{AD}}$;
4 Initialize $\boldsymbol{\theta}_\ell = \{\boldsymbol{\pi}^{(\ell)}, \boldsymbol{A}^{(\ell)}, \boldsymbol{B}^{(\ell)}\}$;
5 $\hat{\boldsymbol{\theta}}_\ell \leftarrow$ BAUM-WELCH$(\mathcal{X}^\ell, \boldsymbol{\theta}_\ell, M, D)$;

Output: L HMM models $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_L\}$ and their subsets of time series $\mathcal{X}^\ell$.

C. HMM-based Anomaly Detection

For the testing part (see Fig. 2(b)), AD is first performed at the parcel level to detect abnormal parcels. The detected anomalies are then localized in time as explained in Section II-C2. The proposed strategy is composed of 1) a *point anomaly detection* step detecting abnormal parcels, and 2) a *contextual anomaly detection* [18] step which allows the starting time of the anomaly to be estimated.

1) AD at the parcel level: The probability that a time series $\boldsymbol{Y} = [\boldsymbol{y}_1, ..., \boldsymbol{y}_T]^\top$ has been generated by the ℓ-th learned model is written as

$$w_\ell = \log P(\boldsymbol{Y}|\hat{\boldsymbol{\theta}}_\ell), \tag{3}$$

for $\ell = 1, ..., L$, where $\boldsymbol{w} = [w_1, ..., w_L]^\top \in \mathbb{R}^L$ is a vector containing the log-probabilities of the test signal with respect to the L HMM models learned using Algorithm 1. Note also that these log-probabilities can be estimated using the forward-algorithm [9], which is summarized in Table III, where $\alpha_{i,t}$ is the forward variable, $a_{i,j}^{(\ell)}$ and $b_{j,t}^{(\ell)}$ are the elements of the matrices $\boldsymbol{A}^{(\ell)}$ and $\boldsymbol{B}^{(\ell)}$ for the ℓ-th model with parameter vector $\hat{\boldsymbol{\theta}}_\ell$. Note that the estimated log-probability in (3) is determined using the procedure in Table III, which sums the forward variables to determine a unique probability for a given parcel. The final AD rule (at the parcel level) is defined as:

$$q = \max_{\ell=1,...,L} w_\ell \underset{H_0}{\overset{H_1}{\lessgtr}} \tau, \tag{4}$$

where $q \in \{0, 1\}$ is the predicted class (with the highest probability) and τ is a threshold related to the probability of false alarm of the test [27]. Looking carefully at the proposed AD rule (4), a tested time series is declared as normal if at least one probability w_ℓ (probability of $\boldsymbol{Y}$ given $\hat{\boldsymbol{\theta}}_\ell$) exceeds the threshold τ. This detection rule is motivated by the fact that the different HMM models capture the possible temporal structures of all normal time series.

TABLE III
ESTIMATION OF LOG-PROBABILITIES FOR A TEST SIGNAL.

Forward-procedure for the $\ell = 1, ..., L$ models	
1) $\alpha_{i,1} = \pi_i^{(\ell)} b_{i,t}^{(\ell)}$	(Initialization)
2) $\alpha_{i,t+1} = \big[\sum_{j=1}^{D} \alpha_t \; a_{i,j}^{(\ell)}\big] b_{j,t}^{(\ell)}$	(Induction)
3) $\log P(\boldsymbol{Y}|\hat{\boldsymbol{\theta}}_\ell) = \sum_{j=1}^{D} \alpha_{i,T}$	(Ending)

2) Anomaly localization via segmentation: When a tested time series $\boldsymbol{Y}$ has been declared as abnormal in (4) (i.e., when $q = 1$), it goes into the second step devoted to anomaly localization. In this step, the HMM models $\hat{\boldsymbol{\Theta}} = \{\hat{\boldsymbol{\theta}}_1, ..., \hat{\boldsymbol{\theta}}_L\}$ determined using Algorithm 1 are used on time segments $\big[t_{\rho_1}, t_{\rho_2}\big] = \{t \,|\, t_{\rho_1} \leq t \leq t_{\rho_2}\}$ (instead of analyzing the complete time series) to determine the starting point of the anomaly in the time series. Consider the forward variable $\alpha_{i,t-1}$ at time $t-1$ in its scaled version defined as $\tilde{\alpha}_{i,t-1} = \alpha_{i,t-1} / \sum_{i=1}^{D} \alpha_{i,t-1}$, where $i = 1, ..., D$, and D is the number of HMM states. The probability of having $\boldsymbol{Y}$ generated by the model $\hat{\boldsymbol{\theta}}$ at time t can be written it terms of $\alpha_{i,t-1}$ as follows

$$u_t = 1/\big(\textstyle\sum_{i=1}^{D} \tilde{\alpha}_{i,t-1} a_{i,j} b_{i,t-1}\big), \tag{5}$$

where u_t depends on the scaled forward variable $\tilde{\alpha}_{i,t-1}$, the transition probability $a_{i,j}$ and the emission probability at time $t-1$. Note that this expression for u_t results from the first-order Markov chain rule, which assumes that the current state (at time t) depends only on its predecessor state (at time $t-1$) [9]. Note also that the forward variable $\alpha_{i,t-1}$ is used in its scaled version $\tilde{\alpha}_{i,t-1}$ to avoid overflow. Indeed, this variable relies on the sum of a large number of terms, as shown in the induction step of the forward procedure. The log-likelihood of the time series in the time segment $[t_{\rho_1}, t_{\rho_2}]$ is defined as

$$\log P(\boldsymbol{y}_{t_{\rho_1}}, ..., \boldsymbol{y}_{t_{\rho_2}} | \boldsymbol{\pi}, \boldsymbol{A}, b_{i,[t_{\rho_1}-1, t_{\rho_2}-1]}) = -\sum_{t=t_{\rho_1}}^{t_{\rho_2}} u_t, \tag{6}$$

where u_t has been defined in (5), and $b_{i,[t_{\rho_1}-1, t_{\rho_2}-1]}$ is the emission probability for the time interval $\big[t_{\rho_1}, t_{\rho_2}\big]$, for $i = 1, ..., D$. As a result, the probability that a time series $[\boldsymbol{y}_{t_{\rho_1}}, ..., \boldsymbol{y}_{t_{\rho_2}}]$ has been generated by the ℓ-th learned model

on the segment $[t_{\rho 1}, t_{\rho 2}]$ can be computed as

$$\boldsymbol{w}_{\ell,[t_{\rho_1},t_{\rho_2}]} = \log P\Big(\boldsymbol{y}_{t_{\rho_1}}, ..., \boldsymbol{y}_{t_{\rho_2}} | \boldsymbol{\pi}^{(\ell)}, \mathbf{A}^{(\ell)}, b^{(\ell)}_{i,[t_{\rho_1}-1,t_{\rho_2}-1]}\Big), \tag{7}$$

where $\boldsymbol{w}_{\ell,[t_{\rho_1},t_{\rho_2}]}$ is the vector containing the log-probabilities of the different models in the time interval $[t_{\rho_1}, t_{\rho_2}]$.

The anomaly localization studied in this paper considers four time intervals associated with different crop life cycles to identify potential anomalies. These intervals are referred to as *Growing*, *Flowering*, *Adult-phase*, and *Senescence*, as in [11], [13]. The hypothesis test to localize the anomaly in a given cycle is defined as

$$\boldsymbol{w}_{\text{cycle}} = \max_{\ell=1,...,L} \boldsymbol{w}_{\ell,[t_{\rho_i},t_{\rho_j}]} \underset{H_0}{\overset{H_1}{\lessgtr}} \tau_{\text{cycle}}, \tag{8}$$

where cycle $\in$ {*Growing, Flowering, Adult-phase, Senescence*}, τ_{cycle} is a threshold depending on a pre-defined probability of false alarm for a given cycle, t_{ρ_i}, t_{ρ_j} are the beginning and the end of the cycle, and $\boldsymbol{w}_{\text{cycle}}$ is a vector containing the highest probabilities for each cycle given the tested time series. Note that (4) defines the parcel detection rule whereas (8) localizes the anomaly in one of the pre-defined cycles.

D. Anomaly Classification

A final step can be included in the analysis to identify the anomaly that has affected the parcel $\boldsymbol{Y}$ using a supervised classifier. After an anomaly has been localized using the steps displayed in Fig. 2(b), we propose to classify the detected anomaly into one of the C classes defined by the user and corresponding to the possible types of anomalies affecting the analyzed crop.

The set of features used for the classification is composed of the detected time series, which are introduced as column-features in the input matrix. More precisely, the feature matrix for training the classifier is of the form $\boldsymbol{X}_{\text{AC}} = [\boldsymbol{x}_1, ..., \boldsymbol{x}_R]$, where $\boldsymbol{x}_r = [x_{r,1}, ..., x_{r,KT}]^\top$ is a vector containing the KT features extracted from the r-th time series at all time instants. It is important to highlight that each time series $\boldsymbol{x}_r$ selected for training the classifier contains an abnormal time series from $\mathcal{X}_{\text{AC}}$ with the corresponding label, denoted by $v_r \in \{1, ..., C\}$, where C is the total number of classes. In the testing part, the classifier generates for each time series $\boldsymbol{y} = [y_1, ..., y_{KT}]^\top$ a label $v_{\boldsymbol{y}} \in \{1, ..., C\}$ indicating the class of the anomaly $\boldsymbol{y}$.

III. Simulations results

The performance of the proposed methodology is evaluated on both simulated and real data[2]. AD is evaluated in terms of precision, recall, and area under the precision-vs-recall curve (AUC), whereas the probabilities of correct classification are used for anomaly classification[3]. The precision, and recall are defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP + FN}},$$

where TP, TN, FP, and FN are the numbers of true positives, true negatives, false positives, and false negatives[4]. On the other hand, the probability of correct classification for class c (denoted as P_c) is defined as:

$$P_c = \frac{1}{R_c} \sum_{r=1}^{R_c} \delta(v_r, \hat{v}_r), \tag{8}$$

where R_c is the total number of training vectors of the class $c \in \{1, ..., C\}$, v_r and $\hat{v}_r$ are the true and estimated labels of the r-th training vector of class c and $\delta(\cdot)$ is an indicator such that $\delta(v_r, \hat{v}_r) = 1$, if $v_r = \hat{v}_r$, and zero otherwise.

A. Synthetic data

Synthetic data are first considered to evaluate the AD-HMM approach for point and contextual AD. For point AD, the proposed approach is compared against three state-of-the-art algorithms: the OC-SVM with a Gaussian (RBF) kernel, the isolation forest (IF), and the hidden-Markov AD (HMAD) [23] with a linear kernel. Note that OC-SVM and IF are considered here for novelty detection, using both training and testing phases (denoted as OC-SVM-N and IF-N), for a fair comparison with the proposed approach. The parameters of each algorithm were adjusted by cross validation leading to 1000 isolation trees and a sub-sampling ratio of 256 for IF, an outlier ratio of $\nu = 0.1$ for HMAD, and an outlier fraction of $\nu = 0.1$ for OC-SVM. Note that the value of the kernel bandwidth γ used in for the RBF kernel was estimated using the median of the pairwise Euclidean distances between vectors of the feature set (i.e, using Jaakkola's heuristic [28]).

Synthetic time series were generated as zero-mean Gaussian sequences with unit variance for the nominal data, i.e., with mean $\mu = 0$ and variance $\sigma^2 = 1$. Mean-value jumps (first scenario) and variance changes (second scenario) were then introduced as anomalies in the testing sequences to evaluate two anomaly scenarios. Specifically, the number of nominal time series generated for the learning phase was $N = 1000$ with $T = 300$ temporal samples. For the testing phase, a set of 500 sequences was generated, with $10\%, 20\%, 30\%$, and 50% of abnormal time series with anomaly blocks having 90 samples. Fig. 3 compares the different AD algorithms in terms of average AUC (the AUC values were obtained from 10 Monte Carlo runs for each probability of anomalies). Fig. 4 shows examples of AUC curves obtained when there are 30% of anomalies in the testing set. In general, the proposed AD-HMM reaches comparable results to those obtained with the OC-SVM algorithm. In the case of variance changes, Fig. 3 shows that the HMAD performance decreases for large percentage of anomalies, reflecting the fact that this algorithm was designed for anomalies characterized by mean value jumps. An additional controlled experiment was conducted to evaluate the robustness of the AD algorithms to possible anomalies present in the training set, with percentages of anomalies of $5\%, 10\%, 20\%, 30\%$, and 50%. The results displayed in Fig. 5 show that the proposed method seems to be more robust to

[2]We used the HMM toolbox available online at https://www.cs.ubc.ca/~murphyk/Software/HMM/hmm.html.

[3]The higher the value of the metric, the better the detection or classification.

[4]The detection threshold was determined using the point of the AUC curve located *the closest to the ideal point* $(0, 1)$.

the presence of anomalies in the training set, when compared to the state-of-the-art.

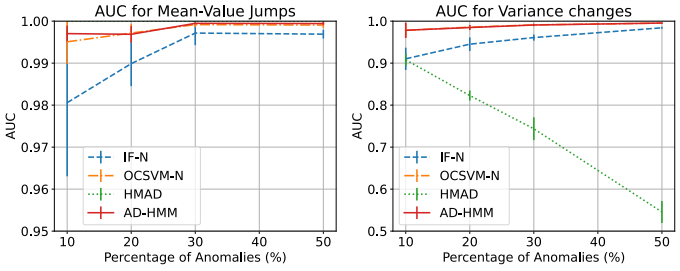

Fig. 3. AD performance in terms of area under the precision-vs-recall curve (AUC) for different percentages of anomalies in the **testing** set.

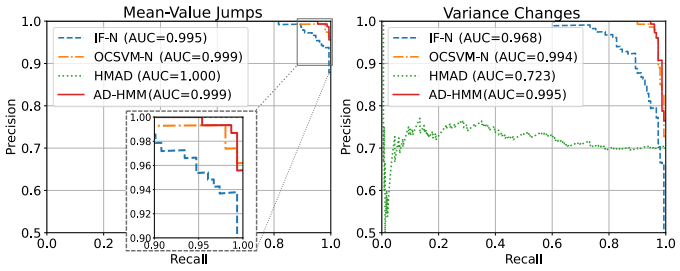

Fig. 4. Illustration of the precision-vs-recall curves for each AD algorithm when the percentage of mean-value jumps and variance changes anomalies in the testing set is 30%.

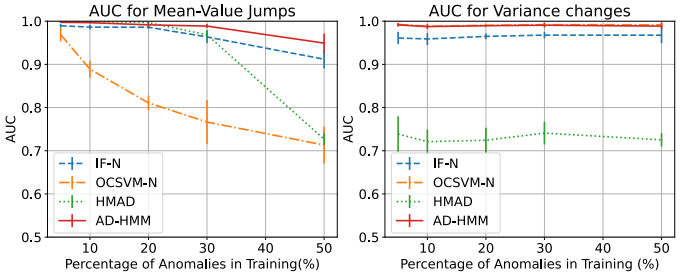

Fig. 5. AD performance in terms of AUC for different percentages of anomalies in the **training** set when the testing set has 30% of anomalies.

In order to evaluate the performance of the proposed anomaly localization (contextual AD) on data with a controlled ground truth, we simulated anomaly cycles by considering time segments of 60 time samples, leading to 5 artificial cycles or intervals. The time series detected as abnormal after the proposed AD-HMM approach were then analyzed to localize the detected anomalies. Note that the synthetic anomalies were generated according to the two scenarios corresponding to mean-value jumps with amplitude $\mu = 1.2$ and variance changes with a ratio of variances equal to 1.5. These scenarios correspond to examples of typical anomalies that can affect crop parcels due to growing or heterogeneity problems. Fig. 6 summarizes the AUC, precision, and recall obtained on each time interval using these synthetic data considering the two scenarios.

As a complement, Fig. 7 displays some anomaly localization results obtained for time series subjected to anomalies defined

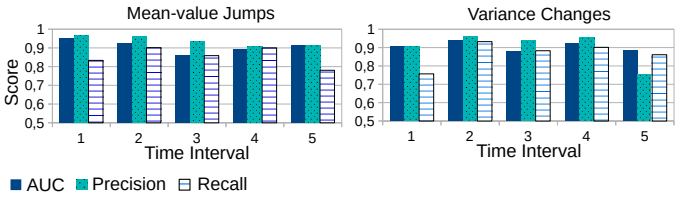

Fig. 6. Summary of the time anomaly localization (contextual AD) performance on each synthetic time interval.

by mean-value jumps (top) or variance changes (bottom). Note that the abnormal time series are represented using a solid line and that the $\star$ labels indicate the samples affected by the anomaly (ground truth). The time intervals considered by the algorithm are located between the dotted vertical lines and those detected as abnormal using the algorithm are shown using the shaded orange boxes. These results illustrate the anomaly localization performance of the proposed method.

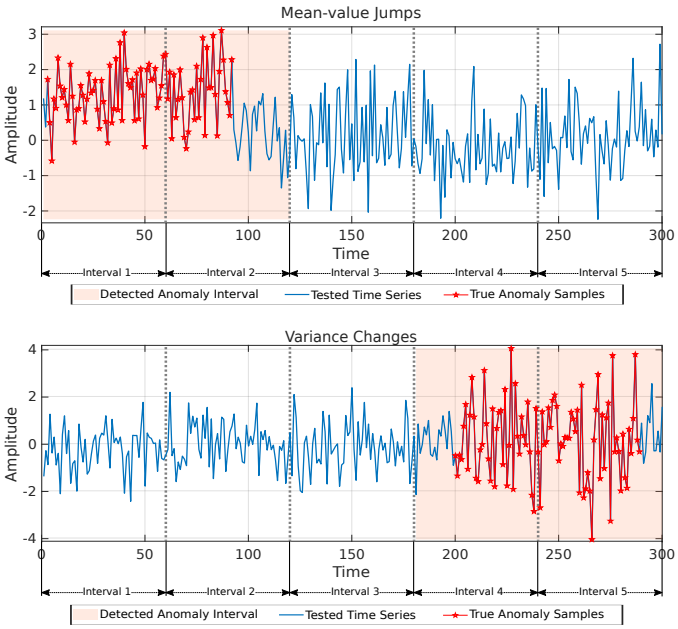

Fig. 7. Examples of anomaly localization for synthetic data subjected to mean-value jumps (top) and variance changes (bottom). The solid line represents the tested time series and the $\star$ line indicates the time samples affected by the anomaly. The time intervals considered in the analysis are delimited by the gray dotted vertical lines and the segments detected as abnormal are highlighted with a shaded orange box.

B. Experiments on real data: multispectral rapeseed images

1) Study Area: The research site is located in Beauce, North of France. This site contains a lot of crop fields such as rapeseed and wheat. A set of 13 multispectral Sentinel-2 images was selected between October 2017 and June 2018. The dataset was processed by the THEIA land data center to level 2A. A set of 2218 rapeseed parcels was extracted from the images with the support of agronomists, as illustrated in Fig. 2(a). The resulting dataset was analyzed with the aim of labeling part of the data for evaluation purposes. The anomalies found in the data were related to heterogeneity and

vegetation phenology problems (which are popular agronomic anomalies), and boundary errors and wrong shape (also-called non-agronomic anomalies). Note that the labeling done in [25] was used in this work.

The partitioning of the dataset for AD was randomly performed as follows: $N = 500$ normal parcels for AD-HMM learning and the remaining parcels (1021 abnormal parcels and 697 normal parcels) for AD-HMM testing. The test parcels detected as anomalies were then considered for the classification experiments. Based on the anomalies found in the database, the anomaly classification step considers the following classes: late growth, early/late senescence and other, where "other" is a class containing all the anomalies that do not belong to the other classes. These anomalies are non-agronomic anomalies (such as errors in parcel boundaries, wrong crop type, or shadow perturbations produced by clouds) or agronomic anomalies affecting a very small number of parcels (such as early growth, early flowering, and crop heterogeneity).

For illustration purposes, Fig. 8 displays (a) the expected temporal profile of the NDVI median with anomalies related to problems in the vegetation phenology such as early/late growth, early/late senescence, and early flowering, and (b) the distribution of a set of 500 nominal and abnormal NDVI medians. As can be seen in the histograms and in the zoom, the normal and abnormal data have different distributions, allowing anomalies to be detected. Note that the different vegetation cycles for the rapeseed crops indicated in the top of Fig. 8(a) are located between the vertical gray dotted lines and were selected based on [2], [25].

2) Analysis of vegetation indices: To analyze the impact of using different time series of VIs for AD, Fig. 9 compares the performance of the proposed AD-HMM algorithm using the median and IQR of different combinations of VIs introduced in Table I. More precisely, Fig. 9 shows the AUC values obtained for different VI combinations after averaging the results of 10 Monte Carlo runs. Note that all the VIs were scaled such that each column of the feature matrix take its values in the interval $(0, 1)$ (minimum-maximum scaling). One can observe that the performance of AD-HMM is very similar when using different VI combinations. Therefore, the remaining analyses will be performed with NDVI only.

The overall performance of the proposed AD-HMM algorithm depends on the parameters N_s, L, and D, that need to be adjusted. The number of states was varied in the set $\{3, 9, 12, 15, 18\}$ whereas the values of N_s and L were chosen in the set $\{10, 25, 50, 100, 200\}$. For $D = 18$, the averaged AUC metrics vary in the interval $[0.80, 0.83]$, and the best performance for all the VIs was obtained when $N_s = 100$ and $L = 10$. These values were selected in the rest of the analysis, in particular to display Fig. 9.

3) Anomaly detection results: The experiments conducted on real data using the proposed AD approach were compared to different algorithms including IF-N, OC-SVM-N, and HMAD. All the algorithms were run using the NDVI features. The parameters of the algorithms were set by cross validation. For OC-SVM-N, the outlier ratio was set to $\nu = 0.1$, and the kernel parameter for the RBF kernel was estimated using

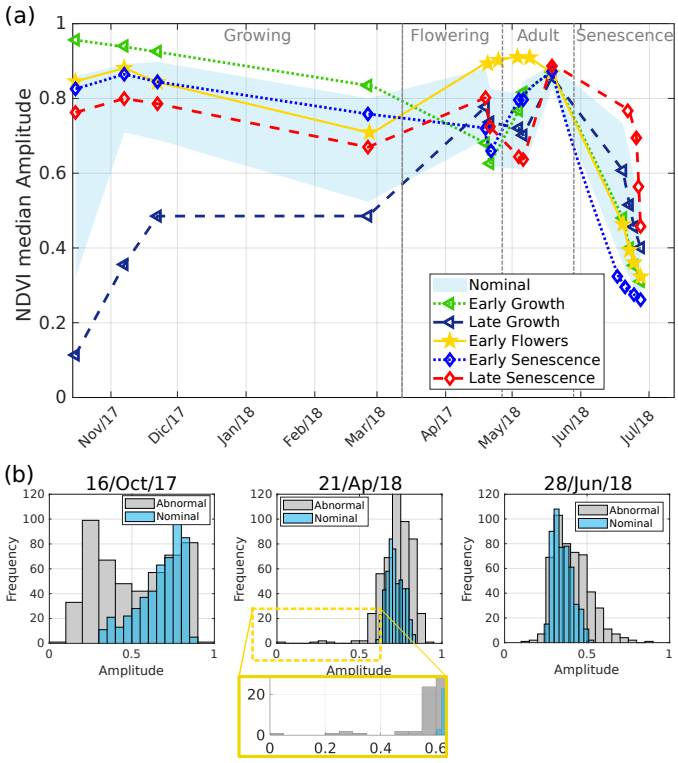

Fig. 8. Temporal profiles and distribution of nominal (blue) and abnormal data. a) Five typical time series profiles for agronomic anomalies are shown, where the shaded blue section corresponds to the normal time series. b) Histogram of 500 time series of normal (blue) and abnormal (gray) NDVI median for three dates, which illustrates how the distribution of abnormal data deviates respect to the nominal data, leading potential anomalies to be detected by the proposed approach.

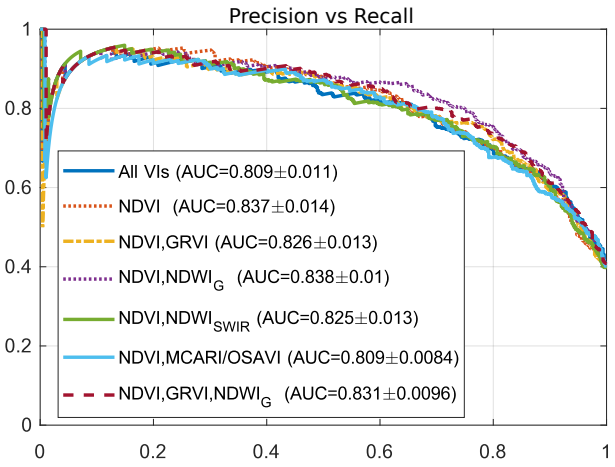

Fig. 9. Performance of the AD-HMM detection using the median and IQR of different temporal vegetation indices.

Jaakkola's heuristic [28]. The IF-N algorithm was run using 1000 isolation trees and a sub-sampling ratio of 256. The outlier fraction used in HMAD was set to $\nu = 0.3$. Fig. 10 summarizes the performance of the different algorithms, where the proposed approach obtains slightly better results against OC-SVM-N and IF-N. The poor performance obtained with HMAD is probably due to the fact that the anomalies affecting crop parcels are not limited to changes in the mean of NDVI.

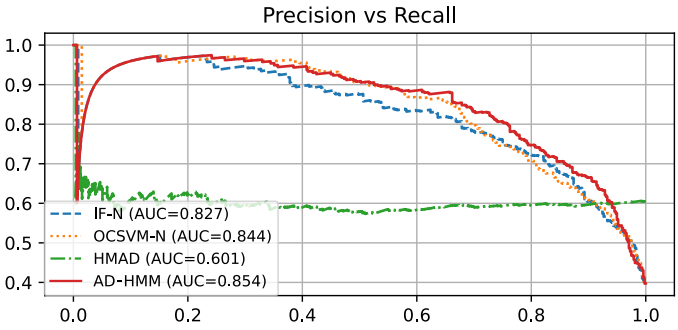

Fig. 10. Performance of different AD methods to detect abnormal parcels in the real dataset.

4) Time Anomaly Localization: The signals detected as abnormal in the previous step were then analyzed to localize the anomalies affecting the crop. For an easier interpretation, the acquisition dates presented in Fig. 8 were transformed into integer values following the cycle time intervals of rapeseed crops as follows: *Growing*$= \{t|1 \leq t \leq 4\}$, *Flowering*$= \{t|5 \leq t \leq 6\}$, *Adult*$= \{t|7 \leq t \leq 9\}$, and *Senescence*$= \{t|10 \leq t \leq 13\}$.

Fig. 11 shows the results obtained for three time series with growth and wrong type problems, where the cycle detected by the proposed AD-HMM is indicated in the top of each figure as "Cycle Detected" whereas the class of the anomaly is referred to as "True Class". The lattice box on the plots in Fig. 11 highlights the detected cycle. Note that based on the learned models, the proposed approach can estimate when the temporal structure deviates from the nominal behavior, even, for subtle deviations as shown in the plot of the middle for *Late Senescence* problems.

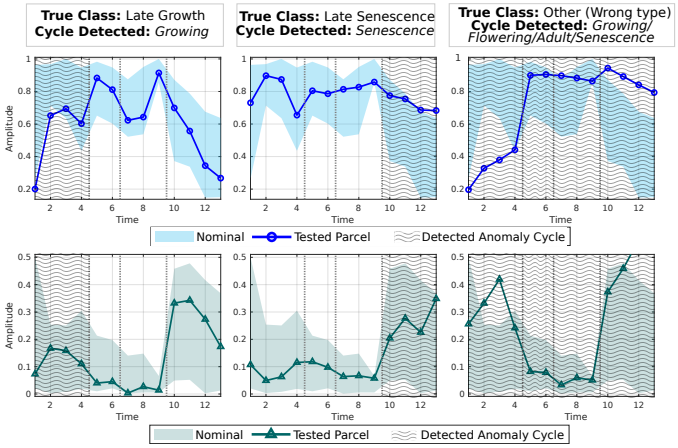

Fig. 11. Time anomaly localization for three tested parcels of rapeseed crops affected by different anomalies. Each plot displays the median (top) and IQR (bottom) of the NDVI features. The box in the top indicates the class of the anomaly and the detected cycle. The lattice box highlights the detected cycle.

C. Anomaly Classification After AD and Localization

The last step of the proposed algorithm classifies some classes of anomalies detected in the rapeseed crops. To evaluate the classification performance on the available samples, the leave-one-out cross-validation (LOOC) strategy was considered. LOOC consists in leaving one vector out of the database, training the classifier with all the remaining samples, testing the classifier with the vector removed from the database and repeating these operations R times, where R is the size of the database. This strategy was selected given the few number of training samples available for anomaly classification. The classifiers considered in this section were the random forest (RF) algorithm with 100 trees, the k-nearest neighbor (k-NN) classifier with $k = 3$, and the support vector machine algorithm with linear (SVM-LN) and Gaussian (SVM-RBF) kernels. The multi-class strategy used in the SVM-based classifiers was based on the One-Against-One voting strategy [29]. In addition, the synthetic minority over-sampling technique (SMOTE) was used to oversample the training set to mitigate the unbalanced nature of the dataset [30].

TABLE IV
PROBABILITIES OF CORRECT CLASSIFICATION (P_c) COMPUTED USING LEAVE-ONE-OUT CROSS-VALIDATION FOR THE DIFFERENT CLASSIFIERS

#	Classes	# Samples	SVM-LN	SVM-RBF	KNN	RF
1	Late Growth	179	0,79 (0,01)	**0,82** (0,01)	0,80 (0,01)	0,81 (0,02)
2	Early Senescence	51	0,92 (0,03)	**0,93** (0,03)	0,84 (0,03)	0,72 (0,01)
3	Late Senescence	29	0,96 (0,01)	**1,00** (0,00)	0,90 (0,03)	0,71 (0,05)
4	Other	352	0,58 (0,01)	0,60 (0,01)	0,61 (0,01)	**0,76** (0,01)
Average Pc			0,81 (0,01)	**0,84** (0,01)	0,79 (0,02)	0,75 (0,02)

TABLE V
CONFUSION MATRIX FOR THE SVM-RBF CLASSIFIER

		Predicted Class			
	Classes	Late Growth	Early Senescence	Late Senescence	Other
True Class	Late Growth	**146**	6	14	13
	Early Senescence	1	**46**	0	3
	Late Senescence	0	0	**29**	0
	Other	54	59	28	**211**

Table IV shows the the estimated probabilities of correct classification for the different classifiers (where the highest values are highlighted in bold and the corresponding standard deviations are indicated into brackets). As can be observed, the highest classification performance is obtained from the SVM-RBF classifier, where the resulting confusion matrix is shown in Table V. Note that the class *other*, which contains anomalies such as heterogeneity, wrong crop type, errors in parcel boundaries, and shadow perturbations, allows us to be close to a real scenario where anomalies that cannot be explained by abnormal plant growing are often present. It is important to mention here that in the rapeseed crops of this study, those classes (wrong crop type, heterogeneity, and shadow perturbations) affect either the whole time series or some time intervals in a random way, yielding anomalies located in any time interval. This lack of structured patterns increases the complexity of the classification, which explains the relatively poor classification performance obtained for this class. Additional information resulting from other data, e.g., from synthetic aperture radar images, might be considered to improve the classification performance. This work is currently under investigation.

IV. Conclusions

This paper studied a method for detecting, localizing and classifying anomalies that affect agricultural crops based on hidden Markov models (HMM) and machine learning. The proposed anomaly detection based on HMM (AD-HMM) exploited the temporal structure of time series of vegetation indices extracted from multispectral images to perform both point (parcel-wise) and contextual (temporal-wise) anomaly detection. The proposed method also allowed the detected anomalies to be temporally localized and classified into predefined classes, information that is valuable for crop monitoring. A comparison with classical anomaly detection algorithms, in terms of precision and recall, provided very promising results. An interesting property of the proposed anomaly detection algorithm is its capacity of localizing and classifying the anomalies located within each time series by exploiting the previously learned HMM models.

Further investigation should be conducted to evaluate the interest of the proposed approach for detecting anomalies in other kinds of crops to characterize their dynamic behavior. Another interesting further work is the extension of the proposed AD-HMM to time series of vegetation indices estimated from multiple remote sensing sources, e.g., extracted from synthetic aperture radar (SAR) images or vegetation optical depth (VOD) retrievals derived from microwave sensors (these vegetation indices have been used in phenology studies in [4]). Finally, it could also be interesting to investigate the application of the proposed approach to features estimated from other kinds of sensors such as compressive multitemporal/multispectral sensors [31], [32], which acquire the images using a compressed format.

Acknowledgments

This work was supported by Minciencias through the agreement 910-2019-Program ECOS Nord-Grant 8598 entitled "Fusión de imágenes multitemporales ópticas y de radar de apertura sintética y su aplicación en detección de anomalías en cítricos".

References

[1] F. Baret and G. Guyot, "Potentials and limits of vegetation indices for LAI and APAR assessment," *Remote Sensing of Environment*, vol. 35, no. 2, pp. 161–173, Feb-Mar. 1991.

[2] A. Veloso, S. Mermoz, A. Bouvet, T. Le Toan, M. Planells, J.-F. Dejoux, and E. Ceschia, "Understanding the temporal behavior of crops using Sentinel-1 and Sentinel-2-like data for agricultural applications," *Remote Sensing of Environment*, vol. 199, pp. 415–426, Sept. 2017.

[3] A. J. Peters, E. A. Walter-Shea, L. Ji, A. Vina, M. Hayes, and M. D. Svoboda, "Drought monitoring with NDVI-based standardized vegetation index," *Photogrammetric Engineering and Remote sensing*, vol. 68, no. 1, pp. 71–75, Jan. 2002.

[4] L. Zeng, B. D. Wardlow, D. Xiang, S. Hu, and D. Li, "A review of vegetation phenological metrics extraction using time-series, multispectral satellite data," *Remote Sensing of Environment*, vol. 237, p. 111511, Feb. 2020.

[5] J. Prendes, M. Chabert, F. Pascal, A. Giros, and J.-Y. Tourneret, "A Bayesian nonparametric model coupled with a Markov random field for change detection in heterogeneous remote sensing images," *SIAM Journal on Imaging Sciences*, vol. 9, no. 4, pp. 1889–1921, Nov. 2016.

[6] C. Gómez, J. C. White, and M. A. Wulder, "Optical remotely sensed time series data for land cover classification:A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 116, pp. 55–72, Jun. 2016.

[7] M. A. García, H. Moutahir, G. M. Casady, S. Bautista, and F. Rodríguez, "Using hidden Markov models for land surface phenology: An evaluation across a range of land cover types in southeast Spain," *Remote Sensing*, vol. 11, no. 5, p. 507, Mar. 2019.

[8] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "A hidden Markov models approach for crop classification: Linking crop phenology to time series of multi-sensor remote sensing data," *Remote Sensing Letters*, vol. 7, no. 4, pp. 3633–3650, Mar. 2015.

[9] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, Feb. 1989.

[10] S. Siachalou, G. Mallinis, and M. Tsakiri-Strati, "Analysis of time-series spectral index data to enhance crop identification over a mediterranean rural landscape," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1508–1512, Sept. 2017.

[11] N. Viovy and G. Saint, "Hidden Markov models applied to vegetation dynamics analysis using satellite remote sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 32, no. 4, pp. 906–917, Jul. 1994.

[12] Y. Shen, L. Wu, L. Di, G. Yu, H. Tang, G. Yu, and Y. Shao, "Hidden Markov models for real-time estimation of corn progress stages using MODIS and meteorological data," *Remote Sensing*, vol. 5, no. 4, pp. 1734–1753, Apr. 2013.

[13] P. B. Leite Cerqueira, R. Q. Feitosa, A. R. Formaggio, G. A. O. P. da Costa, K. Pakzad, and I. D. Sanches, "Hidden Markov models for crop recognition in remote sensing image sequences," *Pattern Recognition Letters*, vol. 32, no. 1, pp. 19–26, Jan. 2011.

[14] Y. Yuan, Y. Meng, L. Lin, H. Sahli, A. Yue, J. Chen, Z. Zhao, Y. Kong, and D. He, "Continuous change detection and classification using hidden Markov model: A case study for monitoring urban encroachment onto farmland in Beijing," *Remote Sensing*, vol. 7, no. 11, pp. 15 318–15 339, Nov. 2015.

[15] M. Meroni, D. Fasbender, F. Rembold, C. Atzberger, and A. Klisch, "Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options," *Remote Sensing of Environment*, vol. 221, pp. 508–521, Feb. 2019.

[16] C. Atzberger, "Advances in remote sensing of agriculture: Context description, existing operational monitoring systems and major information needs," *Remote Sensing*, vol. 5, no. 2, pp. 949–981, Feb. 2013.

[17] H. Vargas, A. Camacho, and H. Arguello, "Spectral unmixing approach in hyperspectral remote sensing: a tool for oil palm mapping," *TecnoLógicas*, vol. 22, no. 45, pp. 131–145, May-Aug. 2019.

[18] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, Jul. 2009.

[19] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. on Management of Data*, Dallas, TX, USA, May. 2000, pp. 93–104.

[20] H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek, "LoOP: local outlier probabilities," in *Proc 18th ACM Conf. Inform. Knowl. Manage. (CIKM '09)*, Hong Kong, China, Nov. 2009, p. 1649–1652.

[21] F. T. Liu, K. M. Ting, and Z.-H. Zhou, "Isolation forest," in *Proc. Int. Conf. on Data Mining*, Pisa, Italy, Dec. 2008, pp. 413–422.

[22] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[23] N. Görnitz, M. Braun, and M. Kloft, "Hidden Markov anomaly detection," in *International Conference on Machine Learning*, Lille, France, Jan. 2015, pp. 1833–1842.

[24] J. Li, W. Pedrycz, and I. Jamal, "Multivariate time series anomaly detection: A framework of hidden Markov models," *Applied Soft Computing*, vol. 60, pp. 229–240, Nov. 2017.

[25] F. Mouret, M. Albughdadi, S. Duthoit, D. Kouamé, G. Rieu, and J.-Y. Tourneret, "Detecting anomalous crop development with multispectral and SAR time series using unsupervised outlier detection at the parcel-level: application to wheat and rapeseed crops," arXiv, Sept. 2020. [Online]. Available: https://arxiv.org/abs/2004.08431

[26] C. S. T. Daughtry, C. L. Walthall, M. S. Kim, E. B. de Colstoun, and J. E. McMurtrey-III, "Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance," *Remote sensing of Environment*, vol. 74, no. 2, pp. 229–239, Nov. 2000.

[27] S. M. Kay, *Fundamentals of statistical signal processing*, ser. Prentice Hall signal processing series. Upper Saddle River, NJ: Prentice Hall PTR, 1993.

[28] T. Jaakkola, M. Diekhans, and D. Haussler, "Using the Fisher kernel method to detect remote protein homologies." in *Proc. Int. Conf. on Intelligent Systems for Molecular Biology*, Heidelberg, Germany, Aug. 1999, pp. 149–158.

[29] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[30] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, p. 321–357, Jun. 2002.
[31] J. M. Ramirez and H. Arguello, "Spectral image classification from multi-sensor compressive measurements," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 1, pp. 626–636, Jan. 2020.
[32] H. Vargas and H. Arguello, "A low-rank model for compressive spectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9888–9899, Dec. 2019.