Advanced Statistical Analysis of Physiological and Spectral Traits for the Early Detection of *Xylella fastidiosa* in Almond Trees

Laura Teresa Martínez-Marquina [1,†], José Ramón Torres-Martín [2,3,†], José Manuel Velarde-Gestera [3],
Miguel Román-Écija[4], Guillermo León-Ropero[4], José Luis Trapero-Casas[4], Juan Antonio Navas-Cortés[4],
Mihaela I. Chidean [1], and Inmaculada Mora-Jiménez [1]

[1]Department of Signal Theory and Communications, Telematics and Computing Systems
Rey Juan Carlos University, Madrid, Spain
e-mails: `laura.marquina,mihaela.chidean,inmaculada.mora@urjc.es`
[2]International Doctoral School, Rey Juan Carlos University, Madrid, Spain
e-mail: `jr.torres.2023@alumnos.urjc.es`
[3]Department of Research and Innovation, Drimay Consultores S.L., Seville, Spain
e-mails: `josetm, jmvelarde@drimay.es`
[4]Institute for Sustainable Agriculture (IAS), Spanish National Research Council (CSIC), Córdoba, Spain
e-mails: `mromanecija,jleon,jltrapero@ias.csic.es,j.navas@csic.es`
[†]These authors contributed equally to this work

Abstract—Xylella fastidiosa (*Xf*) is one of the most aggressive vascular pathogens threatening woody crops, particularly almond trees, in the Mediterranean region. This paper presents a statistical framework for the early detection of *Xf* infection prior to the onset of visible symptoms, leveraging multitemporal physiological and spectral data collected at the leaf level. The approach integrates measurements from porometry, fluorometry, and spectrometry with a non-parametric bootstrap resampling method to identify traits that differentiate health states and reveal physiological responses linked to disease progression. Results reveal that *Xf*-infected trees, which later develop visible symptoms, exhibit significant differences in median values of both spectral indices and physiological variables compared to healthy and intermediate health groups. Grounded in real field data, this work contributes to data-driven plant health monitoring and precision agriculture, demonstrating the potential of combining physiological and spectral indicators for early, non-invasive diagnosis of vascular diseases in perennial crops. The findings support the development of predictive tools for timely disease detection and management in almond and olive orchards.

Keywords-bootstrap-based non-parametric test; *leaf scorch*; *physiological indices*; *hyperspectral indices*; *precision agriculture*.

I. INTRODUCTION

Almond (*Prunus dulcis*) is a crop of major economic and cultural importance in Mediterranean regions [1]. However, its cultivation is increasingly threatened by pathogens such as the vascular bacterium *Xylella fastidiosa* (*Xf*) [2][3] which has emerged as one of the most devastating pathogens in Europe. The bacterium colonizes and obstruct the xylem vessels, disrupting water transport and leading to symptoms (e.g., leaf scorching, branch dieback or canopy desiccation) and eventually plant death [4][5]. Early detection of the diseases caused by *Xf* is challenging due to the asymptomatic infections and latency of visual symptoms and the systemic nature of the infection, highlighting the need for more sensitive and non-invasive diagnostic approaches.

The need for improved in early detection strategies extend to other vascular pathogens affecting woody crops. For example, the highly virulent defoliating pathotype D of *Verticillium dahliae* (*Vd*) [6] showed early physiological stress signals in infected olive trees, such as changes in canopy temperature, chlorophyll fluorescence, and spectral indices, well before visual symptoms appeared, using airborne hyperspectral and thermal imaging [7]. Subsequent studies [8][9] refined this approach, proving that foliar temperature, chlorophyll content, and pigment-based indices could serve as early indicators of plant disease, with Machine Learning (ML) models to classify disease severity across multiple levels.

Based on prior findings, recent studies have extended the use of remote sensing and ML techniques to differentiate between *Xf* and *Vd* infection in tree host. Although both pathogens share aspects of their pathogenesis that lead to similar visual symptoms, such as canopy desiccation and leaf scorching, they induce distinct physiological responses in host plants. A study in 2021 [10] proposed a three-stage classification approach that combined hyperspectral and thermal traits with ML algorithms to discriminate between *Xf* and *Vd* infections in olive trees, achieving over 90% accuracy for both pathogens. Their findings showed that each pathogen follows divergent physiological pathways: *Xf* is more associated with chlorophyll degradation (mainly captured by the Normalized Phaeophytinization Index, NPQI), anthocyanin accumulation, and changes in photochemical reflectance indices (e.g., normalized Photochemical Reflectance Index, PRI_n), while *Vd* is characterized by alterations in carotenoid content and water stress indicators (e.g., Crop Water Stress Index, CWSI). Building on this work, [2][11] demonstrated that these pathogen-specific spectral signatures remain consistent across both olive and almond trees. In particular, NPQI and Solar-Induced Chlorophyll Fluorescence at 760 nm (SIF@760),

indicator of photosynthetic efficiency and physiological stress, were identified as key traits for detecting *Xf* in olives. In contrast, other indices such as PRI$_n$ and Modified Carotenoid Reflectance Index centered at 700nm (CRI$_{700M}$) were found to be more relevant in almonds, underscoring the importance of host-specific physiological responses in disease detection models. Finally, [12] focused specifically on *Xf* in almond trees, proposing the integration of hyperspectral and thermal imagery with an epidemic spread model. Their method enhanced early detection of asymptomatic *Xf*-infected almond trees, achieving up to 59% accuracy. These results highlight the value of integrating plant physiological traits with spatial epidemiological models for large-scale monitoring.

In this study, we aim to identify early physiological indicators of *Xf* infection in almond trees, prior to the onset of visible symptoms. To this end, we analyse multi-temporal leaf-level data acquired through spectrometry, porometry and fluorometry. Using bootstrap-based statistical testing, we identify the most informative traits for discriminating health states over time. Our goal is to establish a foundation for predictive models that support early and non-invasive diagnosis of *Xf* in almond trees, with potential transferability to other woody crops affected by vascular pathogens.

The paper is organized as follows: Section II introduces the database, detailing the preprocessing steps and the instance labelling procedure. It also presents the statistical test employed to identify relevant variables Section III reports and discusses the results, while Section IV concludes the study and outlines directions for future research.

II. Materials and Methods

In this section, we describe the dataset used in the study and also detail the preprocessing procedure. We present the criteria for instance labelling based on visual symptoms and molecular diagnostics, along with the strategy adopted to ensure temporal consistency in label assignment. Finally, we introduce the statistical framework employed to identify the most informative variables for early infection detection.

A. Dataset. Labelling and Preprocessing

The dataset provided by IAS-CSIC includes data from 96 almond trees located in several commercial almond orchards on Mallorca Island (Balearic Islands, Spain). Each tree, identified by a unique ID, has four records corresponding to different measurement dates in 2024: May 15th, June 5th, June 25th, and July 15th. Measurements were taken from one or two branches per tree, depending on whether the tree showed visual symptoms of the disease in the previous crop season. If symptoms were present in 2023, two branches were measured in 2024; otherwise, only one branch was measured. All measured branches were tested for the presence of *Xf* using a molecular diagnosis based on the quantitative Polymerase Chain Reaction (qPCR) analyses [13].

Leaf-level spectral reflectance data were collected using the PolyPen RP410-UVIS (Photon Systems Instruments, Brno, Czech Republic), a portable spectroradiometer that captures

reflectance across 246 discrete wavelengths ranging from 326.2 nm to 791.8 nm. These data were used to calculate several vegetation indices, including physiological traits such as pigment concentration, leaf structure, and photosynthetic activity, serving as potential indicators of plant stress related to vascular diseases. This approach follows established methodologies developed by the IAS-CSIC team [3][14][15], where spectral traits have proven useful for detecting early plant responses to vascular pathogens.

In addition to spectral data, we incorporated physiological parameters measured with the portable instrument LI-600 porometer/fluorometer (LI-COR Biosciences, Lincoln, NE, USA). This device measures stomatal conductance, leaf temperature, ambient humidity, and steady-state chlorophyll fluorescence (Fs), providing information on plant water status and dynamic photochemical activity. Together, these physiological traits complement spectral data. As a result, 43 variables, hereafter referred to as physiological traits, were preserved from the LI-600 for further analysis.

For the exploratory analysis, missing values were removed on a per-variable basis. Outliers in physiological traits, defined as values exceeding the mean ± 20 times the interquartile range (IQR), were imputed using the mean of the corresponding branch. Additionally, some hyperspectral measurements exhibited anomalous values across the entire spectral range for specific leaves; these were considered unreliable and similarly imputed using the branch-level mean.

All leaf-level observations were labelled according to the health status of their corresponding branch. Four mutually exclusive groups were defined based on a combination of three criteria: (i) visual symptoms observed at the measured branch, (ii) molecular diagnosis via qPCR [13] conducted during the final measurement in July, and (iii) overall health status assessed by using a severity score.

Initially, each branch was assigned a single label per time point, which was then applied to all leaf samples collected from that branch. Importantly, branches from the same tree could receive different labels depending on their individual condition. Thus, the four categories were defined based on expert knowledge and measurements taken in July, as follows: Label 0: Negative qPCR result, no visual symptoms, and no suspicion of disease (severity score = 0); Label 1: Branches with negative qPCR result, no visual symptoms, but suspected disease (severity score > 0); Label 2: Branches with positive qPCR result, but no visual symptoms, regardless of the severity score; Label 3: Branches presenting visual symptoms, regardless of the qPCR result.

Figure 1 (a) provides a schematic overview of the labelling procedure described above. Each row represents one of the four measurement dates, while each column corresponds to a single tree identified by its ID. In cases where two branches from the same tree were measured, both labels are displayed within the same cell, separated by a diagonal line. The figure uses a colour coding: green for Label 0; yellow for Label 1; orange for Label 2; red for Label 3. Cells without colour indicate not available data. This visual format illustrates the temporal
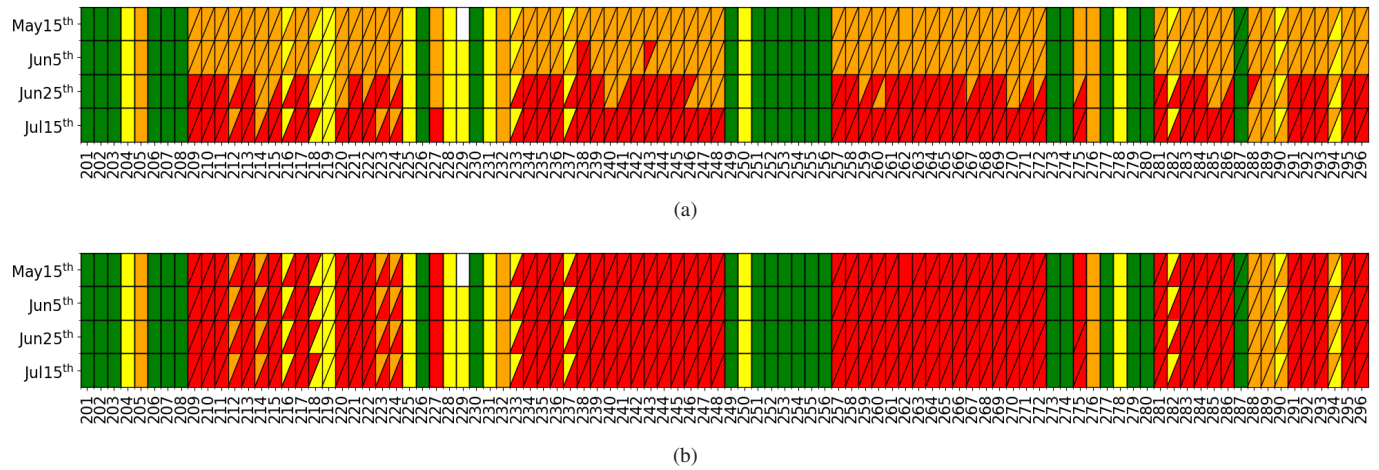
(a)

(b)

Figure 1. Label assignment scheme: (a) initial labelling based on visual symptoms, qPCR results in July, and overall health status based on a severity score; (b) retrospective labelling after backward propagation from the final measurement date (July 15th).

progression of the assigned labels and highlights variability within individual trees.

For subsequent analyses, a retrospective labelling approach was adopted: the label assigned at the final measurement (July 15th) was propagated backward to all earlier observations of the same branch. This approach aligns with the goal of identifying physiological and spectral markers linked to the final health status, ultimately aiming to develop predictive models for early-season detection of *Xf* infection. Figure 1 (b) displays this backward label propagation using the same structure and colour coding as in panel (a), where labels are redefined as follows:

- Label 0: Branches with neither visual symptoms nor suspected disease throughout the monitoring period (severity score = 0) and negative qPCR result in July.
- Label 1: Branches with no visual symptoms over time and negative qPCR result in July, but with suspected disease (severity score > 0).
- Label 2: Branches with no visual symptoms over time but a positive qPCR result in July, indicating *Xf*-infection.
- Label 3: Branches that eventually develop visual symptoms during the monitoring period.

To better illustrate the backpropagation labelling process, let us take tree ID 209 as an example. Neither of its branches showed visible symptoms on May 15th or June 5th, and were therefore initially assigned to Label 2 (Figure 1(a)). However, symptoms became evident on 25th and July 15th, leading to the retrospective assignment of Label 3 to the earlier dates, May 15th and June 5th (Figure 1(b)). This situation is not uncommon: branches ultimately assigned to Label 3 may not exhibit visible symptoms until the third or fourth measurement. Therefore, although they are retrospectively assigned to the most severe health status group from the first measurement, their actual condition may be milder in the early stages.

B. Statistical Analysis. Non-Parametric Difference Test

To identify which of the D variables are most discriminative in differentiating between the defined groups, a **non-parametric**

test for statistical differences was employed. This approach requires the definition of a test statistic, denoted as $T(\cdot)$, which depends on the variable θ_d, where $d = \{1, \ldots, D\}$.

In our context, two test statistics $T(\cdot)$ were used: (i) the difference in medians between two groups, denoted as u and v (Eq. 1), and (ii) the difference in standard deviations between the same groups (Eq. 2).

$$T_1(\theta_d) = \text{median}(\theta_d)_u - \text{median}(\theta_d)_v \quad (1)$$

$$T_2(\theta_d) = \text{std}(\theta_d)_u - \text{std}(\theta_d)_v \quad (2)$$

A statistical comparison between groups was carried out by performing a hypothesis test for each variable. The aim was to determine whether the observed differences between groups for a given variable were statistically significant. The hypotheses for each test were defined as follows:

H_0 : There is no significant difference between groups with respect to θ_d

H_1 : There is a significant difference between groups with respect to θ_d

where H_0 and H_1 denote the *null* and *alternative* hypotheses, respectively.

To robustly estimate the value of the test statistic $T(\theta_d)$, particularly in scenarios with limited sample sizes and unknown population distributions, we employed a non-parametric **Bootstrap** resampling approach [16][17]. This approach involves repeatedly resampling the observed data with replacement to generate an empirical distribution of the statistic. The resulting estimate, $\hat{T}(\theta_d)$, closely approximates the true value, i.e., $\hat{T}(\theta_d) \approx T(\theta_d)$.

The bootstrap procedure involves generating multiple *re-samples* B^* of size m by randomly drawing observations from the original sample B of size n, with $m \leq n$. In this study, when comparing two groups, m was set to the size of the minority group to ensure balanced resampling. Importantly, resampling was performed *with replacement*, allowing individual observations to appear more than once

within a given B^*. This process was repeated N times, yielding a collection of N bootstrap estimates for the test statistic $T(\theta_d)$, denoted as $\{\hat{T}_1^*(\theta_d), \ldots, \hat{T}_N^*(\theta_d)\}$. These estimates form an empirical approximation of the sampling distribution of the statistic. Choosing a sufficiently large number of resamples N is essential to reduce variability in the bootstrap estimates and to obtain stable and reliable statistical inference.

Finally, to evaluate the statistical significance of the estimated test statistic $\hat{T}(\theta_d)$, we computed its empirical Confidence Interval (CI) based on the bootstrap distribution. The CI defines a range within which the true value of $T(\theta_d)$ is expected to lie with a specified probability, serving as a criterion for accepting or rejecting the null hypothesis. To construct the CI, a significance level α is defined, typically set to 0.05, which corresponds to a $100(1 - \alpha)\%$ confidence level. This implies that the true value of $T(\theta_d)$ is expected to fall within the interval with probability approximately $1 - \alpha$. Accordingly, the hypothesis test can be reformulated in terms of the CI as follows:

$$\begin{cases} H_0 : 0 \in CI(T(\theta_d)_{uv}) \\ H_1 : 0 \notin CI(T(\theta_d)_{uv}) \end{cases}$$

In this framework, if the CI does not include zero, there is sufficient evidence to reject H_0 in favour of H_1, indicating a statistically significant difference between the groups.

Based on this framework, the statistical test was applied at the leaf level. Each bootstrap instance corresponded to data from an individual leaf, allowing the analysis to capture within-branch variability in physiological and spectral traits. This procedure was performed independently for each of the D variables and across all four measurement time points. By treating each leaf as a distinct observation, the analysis achieves the highest granularity, enabling the detection of subtle group differences at the most localized scale.

III. Results and Discussion

The non-parametric bootstrap test was applied across all variables, including vegetation indices from the literature [3] and a broad set of physiological traits, to assess group-level differences. For each variable and time point, we evaluated differences in median values between health groups and also examined differences in variability (standard deviation) to capture physiological heterogeneity linked to disease progression.

To streamline the presentation, Figure 2 highlights hyperspectral and physiological traits that at least in two measurement dates showed significant differences between the two most contrasting health states, Label 0 (asymptomatic and qPCR-negative) and Label 3 (symptomatic). These comparisons are shown as the first CI for each variable. Additional CIs represent comparisons involving intermediate labels, which were also examined to evaluate their potential for early detection. Green CI indicate statistically significant differences, while red intervals indicate non-significant ones. The black dot indicates the median of each bootstrap distribution. While the Label 0 vs. Label 3 comparison was expected to yield the most pronounced differences, when comparing groups with intermediate labels

(1 and 2), relevant differences also emerged. This reinforces the potential utility of these variables for early detection of physiological and spectral changes prior to symptom onset. Table I lists the variables whose CIs are displayed in Figure 2, along with their corresponding descriptions. Spectral indices ($MCARI_2$ to RGI) were computed from leaf-level hyperspectral reflectance data, while physiological traits (GSW to T_{leaf}) were obtained from direct measurements with the porometer/fluorometer.

A. Hyperspectral Indices

The results in Figure 2 (a) indicate that several spectral indices consistently differ in median values between the healthiest leaves (Labels 0 and 1) and those showing advanced symptoms (Label 3), particularly during the early measurements in May 15th and June 5th. This suggests that certain spectral indices may serve as early indicators of physiological disruption before visible symptoms appear. In contrast, differences in standard deviation were generally less conclusive (data not shown due to space constraints). The most notable variability differences were observed at the second time point (June 5th), particularly between the most severely affected groups (Labels 2 and 3). In some cases these differences persisted into later stages of the season, including late June and July, suggesting progressive physiological divergence as the infection advanced.

The indices DCabxc and TCARI, both associated with chlorophyll content, showed consistent differences between leaves labelled as 0 and those leaves labelled as 3 across all measurement dates. Additionally, both indices occasionally showed significant differences when comparing leaves labelled as 1 or 2 to those labelled as 3. Notably, TCARI has previously been identified as one of the most discriminative indices for detecting *Vd* symptoms in olive trees at both early and advanced stages, due to its sensitivity to chlorophyll degradation [9]. Similarly, in

TABLE I. SUMMARY OF VARIABLES.

| Variable	Description
$MCARI_2$	Modified Chlorophyll Absorption Index
MSAVI	Modified Soil-Adjusted Vegetation Index
TVI	Triangular Vegetation Index
$MTVI_1$ $MTVI_2$	Modified Triangular Vegetation Indices
CTR_1	Carter Index
DCabxc	Reflectance Band Ratio Index
TCARI	Transformed Chlorophyll Absorption in Reflectance Index
T_O	TCARI / Optimized Soil-Adjusted Vegetation Index
VOG_1 VOG_2	Vogelmann Indices
$PRIM_3$	Photochemical Reflectance Index
G	Greenness Index
RGI	Red-Green Index
GSW	Stomatal conductance
GTW	Total conductance
$E_{apparent}$	Transpiration
VP_{leaf}	Leaf vapor pressure
$H2O_{leaf}$	Leaf H2O mole fraction
Fs	Minimum fluorescence in light
Fm'	Maximum fluorescence in light
$PhiPS_2$	Quantum efficiency in light
ETR	Electron transport rate
T_{leaf}	Calculated leaf temperature
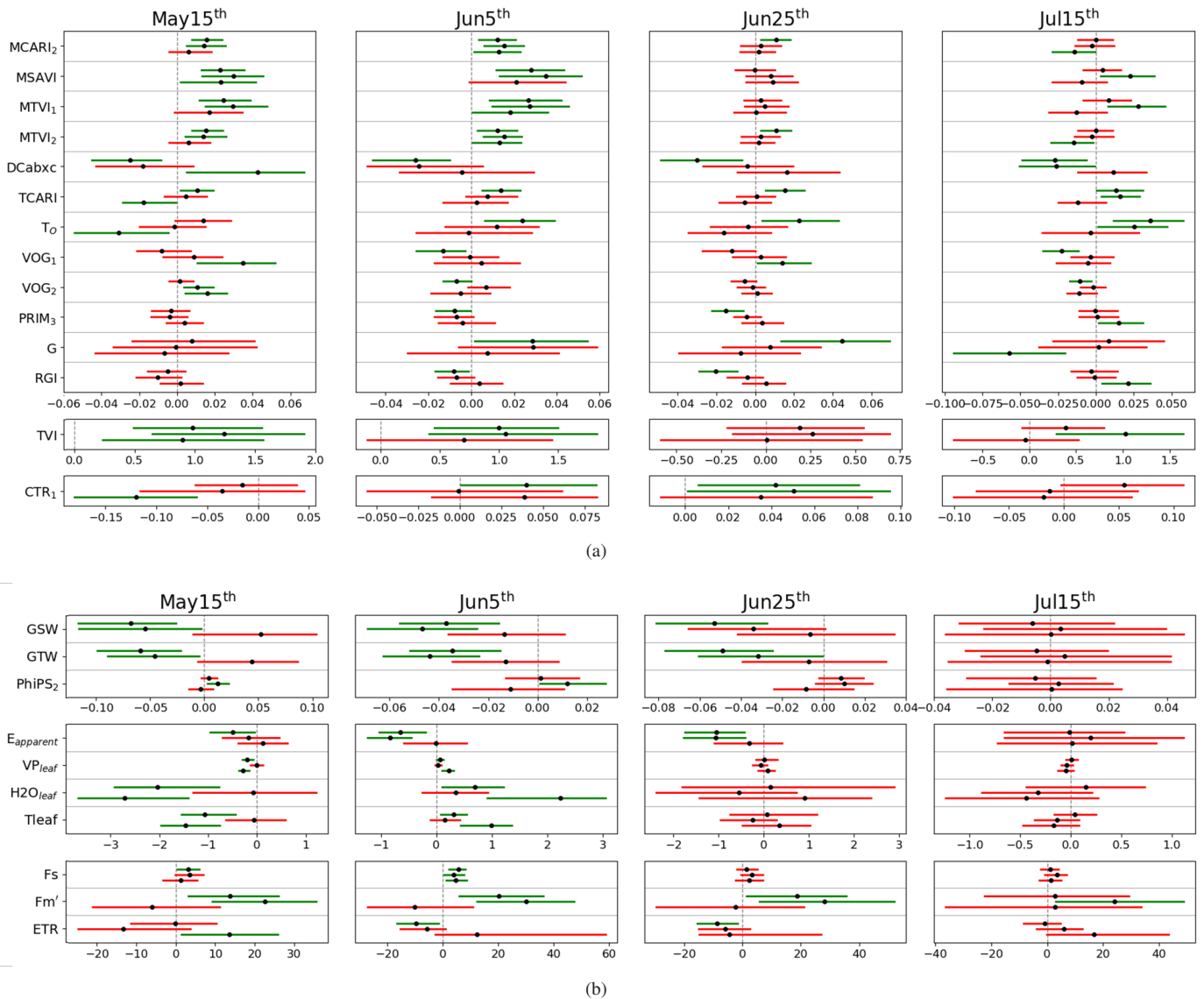
Figure 2. CI of bootstrap distributions for median differences between labelled groups pairs across time points: (a) spectral indices; (b) physiological traits. Each variable includes three CI, corresponding to comparisons between Label 0 vs. 3, Label 1 vs. 3, and Label 2 vs. 3, shown from top to bottom.

almond trees infected with *Xf*, TCARI also played a key role in remote sensing-based classification models, forming part of the index subset that most effectively distinguished symptomatic from asymptomatic trees [12]. The same study also highlighted the relevance of structural indices, such as MCARI, which are linked to leaf nutritional status. A progressive decline in MCARI values was reported as symptom severity increased, consistent with reduced nitrogen content in *Xf*-infected leaves. Our analysis aligns with these findings, showing statistically significant differences in MCARI between leaves from label group 0 and 3 during the first three measurement dates, and between Labels 1 and 3 during the first two. In all cases, the distribution shifted to the right, indicating lower MCARI values in the most affected group. In contrast, differences in standard deviation between groups were negligible, suggesting that overall variability in MCARI remained relatively stable

regardless of *Xf* infection status.

Other structural indices, including TVI, $MTVI_1$, and $MTVI_2$, also showed significant median differences between leaves labelled as 0 and 1 and those labelled as 3, particularly during the first two measurement dates. In particular, when comparing label groups 1 and 3 at the second time point, TVI and $MTVI_1$ exhibited significant differences in dispersion, indicating increased variability in the physiological response of *Xf*-infected trees. Previous studies [9] have incorporated both TVI and $MTVI_1$ into discriminant analysis frameworks to classify *Vd* infection severity in olive trees. These indices significantly contributed to distinguish symptomatic from asymptomatic trees and to finer severity stratification. Further support is provided in [7], which identified these indices as part of a core set of structural traits frequently used in early disease detection models, likely due to their sensitivity to subtle canopy

structure changes associated with early pathogen infection.

Studies by [7] and [9] also identified Optimized Soil-Adjusted Vegetation Index (OSAVI), part of the soil-adjusted vegetation index family, as informative for plant health assessment. In our analysis (see Figure 2), we present results for MSAVI, which, like OSAVI, is derived from Normalized Difference Vegetation Index (NDVI) and uses the same spectral bands. However, MSAVI is often preferred in environments with sparse or discontinuous canopy cover and strong soil background influence, conditions typical of almond orchards. In our work, MSAVI consistently showed differences between the symptomatic reference group (Label 3) and the other label groups, particularly during the earlier measurement dates. This pattern supports the potential of soil-adjusted vegetation indices as effective early indicators of canopy-level physiological changes associated with pathogen infection.

Other notable pigment-related indices included T_O (TCARI/OSAVI), which estimates chlorophyll content while minimizing the soil reflectance effects, and the Vogelmann indices VOG_1 and VOG_2, which are sensitive to chlorophyll content and based on red-edge spectral bands. T_O has consistently been identified as one of the most responsive indices for detecting Xf infection in both olive trees [2][18] and almond trees [12]. In our analysis, T_O showed significant differences between leaves labeled as 0 and 3 during the June and July measurements, with higher values in leaves from healthy branches, indicating greater chlorophyll content. Similarly, the Vogelmann indices, previously reported for Xf detection in olive [2][18] and almond trees [12], showed differences between Labels 0 and 3 across several dates. Differences were also observed between intermediate groups (Labels 1 or 2) and group 3, highlighting the sensitivity of these indices to progressive chlorophyll degradation. Notably, VOG_1 showed significant differences in variability (standard deviation) only between the most severely affected groups (Labels 2 and 3) during early time points, with variability remaining relatively stable later in the season.

Figure 2 also includes the xanthophyll-related index $PRIM_3$, which, along with the G and RGI indices, demonstrated significant differences between leaves labelled 0 and those labelled 3 on June 5th and 25th. Furthermore, G and RGI also showed significant differences between groups 2 and 3 in the July measurement, highlighting their potential to detect progressive stages of Xf infection. The PRI and its variants, such as the normalized PRI_m, are well known for their effectiveness in identifying plant stress symptoms, including those caused by Xf [2]. These indices are closely linked to the xanthophyll cycle, which reflects changes in photosynthetic activity and efficiency under stress conditions [12].

Overall, these results confirm that several spectral indices, particularly those associated with pigment content and vegetation structure, can effectively discriminate between healthy and Xf–infected trees, even in asymptomatic stages. The strongest differences were observed between fully healthy leaves (Label 0) and infected and symptomatic leaves (Label 3). Significant differences among intermediate groups (Labels 1 and 2) further

support the potential of these indices for early detection and large-scale monitoring of Xf infection.

B. Physiological plant traits

Figure 2 (b) shows the CI for physiological plant traits derived from porometer/fluorometer median differences between the healthiest (Label 0) and most affected (Label 3) trees, as well as the intermediate labelled groups against Label 3. Similar to spectral indices, certain physiological traits consistently differ between these groups, particularly in earlier measurement dates. This pattern suggests that these variables may be sensitive to early physiological alterations caused by Xf infection, even before visible symptoms appear. However, by the final measurement in July, these differences tend to diminish likely due to a general increase in thermal and water stress affecting all trees. This is supported by the relevance of temperature-based indices like the Crop Water Stress Index (CWSI) in Xf detection [7][9]–[12]. Under such uniform stress conditions, physiological responses among trees may converge, making it harder to distinguish between healthy and infected trees. It is plausible that trees labelled as 0 could recover post-stress, while those labelled as 3 may not, due to their compromised physiological state. Unlike these indices, the dispersion of variables across groups shows significant differences, particularly between Labels 0 and 3 in May and June. This highlights the potential of porometer/fluorometer-derived traits as early and sensitive indicators of Xf infection, though their discriminatory power may decrease under widespread environmental stress later in the season [10][11].

Conductance variables (GSW and GTW) exhibit significant differences in median values between the healthiest leaves (Labels 0 or 1) and the most affected ones (Label 3) at early stages, linking leaf health to conductance [12]. Interestingly, both conductance and transpiration, estimated by the $E_{apparent}$ variable, are higher in Label 3 leaves (potentially infected but still asymptomatic) than in healthy ones. This contrasts with reported findings that Xf infection typically reduces conductance and transpiration due to xylem blockage and impaired water transport [3]. This discrepancy may reflect a transitional infection phase, where physiological decline is not yet fully established or is masked by compensatory mechanisms [3]. Among the most affected groups (Labels 2 and 3), median values of conductance and transpiration (i.e., $E_{apparent}$) do not differ clearly. While conductance shows greater dispersion in May and June measurements than in the last measurement, transpiration variability does not differ significantly. These patterns suggest that early or asymptomatic stages of infection can produce variable physiological responses that may not always align with the typical decline reported in advanced stages [3][19].

The VP_{leaf} variable shows significant differences between leaves labelled 0 and 3, as well as between groups 2 and 3 in early stages. In May, the most affected leaves (Label 3) have higher CI values for vapour pressure, but this trend reverses in June. Vapour pressure deficit is recognized as an early indicator

of water stress caused by *Xf* infection [7][12][18]. Similarly, $H2O_{leaf}$ exhibit similar group differences as VP_{leaf}.

For fluorescence parameters, the healthiest leaves showed significantly higher median values of minimum (F_s) and maximum (F_m') fluorescence early in the season, with all group comparisons becoming significant by June. This aligns with previous findings that fluorescence is a sensitive indicator of *Xf* infection, typically decreasing as photosynthetic activity declines [12]. However, these differences fade as visual symptoms appear later in the season, and variability in fluorescence remains inconclusive. Another fluorescence-related variable, $PhiPS_2$, only showed significant differences between groups 1 and 3 during the first two measurement dates. In contrast, ETR values were higher in Label 3 leaves compared to Label 0 during the second and third time points.

Finally, leaf temperature T_{leaf} was significantly higher in group 3 compared to groups 0 and 2 in June, consistent with previous findings [2]. However, this trend reverses in the second measurement date, with T_{leaf} being higher in groups 0 and 2 than in group 3. At more advanced infection stages, significant differences in T_{leaf} are no longer observed between groups.

IV. Conclusion and Future Work

The results directly support the main objective of the study of identifying early physiological indicators of *Xf* infection in almond trees, prior to the onset of visible symptoms. Multi-temporal leaf-level data, analyzed through non-parametric bootstrap testing, consistently revealed that both spectral and physiological traits can detect infection at early stages. Specifically, hyperspectral indices linked to pigment content (e.g., TCARI, DCabxc, MCARI) and structural traits (e.g., TVI, $MTVI_1$, MSAVI), along with physiological variables such as stomatal conductance, transpiration, vapour pressure, fluorescence, and leaf temperature, consistently distinguished healthy from infected trees in early stages.

These findings may suggest that physiological stress responses precede visible symptom development and can be captured through targeted measurements. Moreover, they align with recent evidence indicating that remote sensing traits, particularly fluorescence and thermal signals, enable accurate, large-scale, and early detection of *Xf*, supporting their use in disease monitoring and management. Interestingly, though differences were most pronounced in May and June, they tended to diminish by July. This likely reflects a general increase in thermal and water stress affecting all trees, which may mask infection-specific responses. It is plausible that healthy trees (Label 0) may recover after the stress period, whereas infected trees (Label 3) likely remain physiologically compromised. Although our analysis focused on variables showing statistically significant differences between the two most contrasting health states (Labels 0 and 3) across at least two time points, relevant differences also emerged among intermediate groups. This reinforces the diagnostic potential of the selected indicators and their applicability to early-stage detection.

This study represents an initial step toward developing a comprehensive framework for early *Xf* detection using spectral data acquired via proximal sensors. A limitation of this study is the small size in intermediate health groups (Labels 1 and 2), which may reduce the statistical power in comparisons involving these categories. To address this, future work will aim to improve dataset balance by increasing representation in these groups.

Given the complexity of *Xf* infections, further validation is essential. This includes defining threshold values for vegetation indices that reliably capture early physiological alterations. Additional research should refine trait selection, account for environmental variability, and validate findings across different crops and landscapes. Future efforts should also focus on integrating the identified physiological and spectral indicators into operational remote sensing workflows, as well as ML and spatial epidemiological models, to improve diagnostic accuracy and scalability. Expanding the use of multispectral and thermal sensors, along with optimized band selection, will be essential for developing cost-effective, real-time tools for disease surveillance and management.

References

[1] E. Spada *et al.*, "Economic Sustainability Assessment of Mediterranean Crops: A Comparative Life Cycle Costing Analysis", *Frontiers in Sustainable Food Systems*, vol. 6, p. 1 004 065, 2022.

[2] P. J. Zarco-Tejada *et al.*, "Divergent Abiotic Spectral Pathways Unravel Pathogen Stress Signals Across Species", *Nature Communications*, vol. 12, no. 1, p. 6088, 2021.

[3] P. J. Zarco-Tejada *et al.*, "Previsual Symptoms of Xylella Fastidiosa Infection Revealed in Spectral Plant-Trait Alterations", *Nature Plants*, vol. 4, no. 7, pp. 432–439, 2018.

[4] J. Rapicavoli, B. Ingel, B. Blanco-Ulate, D. Cantu, and C. Roper, "Xylella Fastidiosa: an Examination of a Re-Emerging Plant Pathogen", *Molecular Plant Pathology*, vol. 19, no. 4, pp. 786–800, 2018.

[5] K. Schneider *et al.*, "Impact of Xylella Fastidiosa Subspecies Pauca in European Olives", *Proceedings of the National Academy of Sciences*, vol. 117, no. 17, pp. 9250–9259, 2020.

[6] J. A. Navas-Cortés *et al.*, "Spatiotemporal Analysis of Spread of Infections by Verticillium Dahliae Pathotypes within a High Tree Density Olive Orchard in Southern Spain", *Phytopathology*, vol. 98, no. 2, pp. 167–180, 2008.

[7] R. Calderón, J. A. Navas-Cortés, C. Lucena, and P. J. Zarco-Tejada, "High-Resolution Airborne Hyperspectral and Thermal Imagery for Early Detection of Verticillium Wilt of Olive using Fluorescence, Temperature and Narrow-Band Spectral Indices", *Remote Sensing of Environment*, vol. 139, pp. 231–245, 2013.

[8] R. Calderon, C. Lucena, J. L. Trapero-Casas, P. J. Zarco-Tejada, and J. A. Navas-Cortes, "Soil Temperature Determines the Reaction of Olive Cultivars to Verticillium Dahliae Pathotypes", *PLoS One*, vol. 9, no. 10, e110664, 2014.

[9] R. Calderón, J. A. Navas-Cortés, and P. J. Zarco-Tejada, "Early Detection and Quantification of Verticillium Wilt in Olive using Hyperspectral and Thermal Imagery over Large Areas", *Remote Sensing*, vol. 7, no. 5, pp. 5584–5610, 2015.

[10] T. Poblete *et al.*, "Discriminating Xylella Fastidiosa from Verticillium Dahliae Infections in Olive Trees using Thermal- and Hyperspectral-Based Plant Traits", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 179, pp. 133–144, 2021.

[11] T. Poblete *et al.*, "Detection of Symptoms Induced by Vascular Plant Pathogens in Tree Crops using High-Resolution Satellite Data: Modelling and Assessment with Airborne Hyperspectral Imagery", *Remote Sensing of Environment*, vol. 295, p. 113 698, 2023.

[12] C. Camino *et al.*, "Detection of Xylella Fastidiosa in Almond Orchards by Synergic Use of an Epidemic Spread Model and Remotely Sensed Plant Traits", *Remote Sensing of Environment*, vol. 260, p. 112 420, 2021.

[13] P. Bonants *et al.*, "Development and Evaluation of a Triplex TaqMan Assay and Next-Generation Sequence Analysis for Improved Detection of Xylella in Plant Material", *Plant Disease*, vol. 103, no. 4, pp. 645–655, 2019.

[14] R. Calderón, M. Montes-Borrego, B. Landa, and P. Zarco-Tejada, "Detection of Downy Mildew of Opium Poppy using High-Resolution Multi-Spectral and Thermal Imagery Acquired with an Unmanned Aerial Vehicle", *Precision Agriculture*, vol. 15, no. 6, pp. 639–661, 2014.

[15] R. Calderón Madrid, J. A. Navas Cortés, C. Lucena León, and P. J. Zarco-Tejada, "High-Resolution Hyperspectral and Thermal Imagery Acquired from UAV Platforms for Early Detection of Verticillium Wilt using Fluorescence, Temperature and Narrow-Band Indices", in *Proc of the Workshop on UAV-based Remote Sensing Methods for Monitoring Vegetation*, Transregional Collaborative Research Centre 32, 2013.

[16] B. Efron and R. J. Tibshirani, *An introduction to the bootstrap*. Chapman and Hall/CRC, 1994.

[17] C. Figuera *et al.*, "Nonparametric Model Comparison and Uncertainty Evaluation for Signal Strength Indoor Location", *IEEE Transactions on Mobile Computing*, vol. 8, no. 9, pp. 1250–1264, 2009.

[18] T. Poblete *et al.*, "Detection of Xylella Fastidiosa Infection Symptoms with Airborne Multispectral and Thermal Imagery: Assessing Bandset Reduction Performance from Hyperspectral Analysis", *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 27–40, 2020.

[19] A. J. McElrone, J. L. Sherald, and I. N. Forseth, "Interactive effects of water stress and xylem-limited bacterial infection on the water relations of a host vine", *Journal of Experimental Botany*, vol. 54, no. 381, pp. 419–430, 2003.