Remote Sensing of Environment 282 (2022) 113281

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Detecting *Xylella fastidiosa* in a machine learning framework using Vcmax and leaf biochemistry quantified with airborne hyperspectral imagery

C. Camino [a,*], K. Araño [a], J.A. Berni [b], H. Dierkes [a], J.L. Trapero-Casas [b], G. León-Ropero [b], M. Montes-Borrego [b], M. Roman-Écija [b], M.P. Velasco-Amo [b], B.B. Landa [b], J.A. Navas-Cortes [b], P. S.A. Beck [a]

[a] European Commission (EC), Joint Research Centre (JRC), Ispra, Italy
[b] Instituto de Agricultura Sostenible (IAS), Consejo Superior de Investigaciones Científicas (CSIC), Córdoba, Spain

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
Hyperspectral
Vcmax
Nitrogen
Xylella fastidiosa
Machine learning

ABSTRACT

The bacterium *Xylella fastidiosa* (*Xf*) is a plant pathogen that can block the flow of water and nutrients through the xylem. *Xf* symptoms may be confounded with generic water stress responses. Here, we assessed changes in biochemical, biophysical and photosynthetic traits, inferred using biophysical models, in *Xf*-affected almond orchards under rainfed and irrigated conditions on the Island of Majorca (Balearic Islands, Spain). Recent research has demonstrated the early detection of *Xf*-infections by monitoring spectral changes associated with pigments, canopy structural traits, fluorescence emission and transpiration. Nevertheless, there is still a need to make further progress in monitoring physiological processes (e.g., photosynthesis rate) to be able to efficiently detect when *Xf*-infection causes subtle spectral changes in photosynthesis. This paper explores the ability of parsimonious machine learning (ML) algorithms to detect *Xf*-infected trees operationally, when considering a proxy of photosynthetic capacity, namely the maximum carboxylation rate (V_{cmax}), along with carbon-based constituents (CBC, including lignin), and leaf biochemical traits and tree-crown temperature (T_c) as an indicator of transpiration rates. The ML framework proposed here reduced the uncertainties associated with the extraction of reflectance spectra and temperature from individual tree crowns using high-resolution hyperspectral and thermal images. We showed that the relative importance of V_{cmax} and leaf biochemical constituents (e.g., CBC) in the ML model for the detection of *Xf* at early stages of development were intrinsically associated with the water and nutritional conditions of almond trees. Overall, the functional traits that were most consistently altered by *Xf*-infection were V_{cmax}, pigments, CBC, and T_c, and, particularly in rainfed-trees, anthocyanins, and T_c. The parsimonious ML model for *Xf* detection yielded accuracies exceeding 90% (kappa = 0.80). This study brings progress in the development of an operational ML framework for the detection of *Xf* outbreaks based on plant traits related to photosynthetic capacity, plant biochemistry and structural decay parameters.

1. Introduction

Increasing global trade and travel in a changing climate has led several plant pests to spread to multiple continents and cause major agricultural, environmental, and socio-economic impacts (EFSA (European Food Safety Authority), 2019a). *Xylella fastidiosa* (*Xf*) is considered one of the most dangerous plant pathogenic bacteria worldwide and can infect >600 plant species (EFSA, 2021). Up to date, there is currently no control measure available to eliminate *Xf* from diseased plants in open field conditions once they became infected (EFSA, 2019b). *Xf* causes significant yield losses in numerous crops in the Americas (where it is native), and more recently in Europe and Israel. *Xf* is known to cause severe direct damage to major crops including almonds, citrus, grapevines, olives, stone fruits and forest trees, landscape, and ornamental trees, with high impacts (EFSA, 2021). The economic impact of *Xf* in olives alone has been estimated for Italy, Greece and Spain from 2.38 to 7.49 billion euros per year, if *Xf* is not contained (Schneider et al., 2020). *Xf* is listed as a priority pest for the EU based on its probability of spreading, establishment, and consequences for the EU (Commission Delegated Regulation (EU) 2019/1702).

Xf is a xylem-invading pathogen that colonizes its host's vascular tissue and eventually blocks water flow through the xylem. *Xf* infections

* Corresponding author.
E-mail address: carlos.camino-gonzalez@ec.europa.eu (C. Camino).

https://doi.org/10.1016/j.rse.2022.113281
Received 14 February 2022; Received in revised form 18 August 2022; Accepted 17 September 2022
Available online 4 October 2022

gradually block the xylem, leading to a reduction of sap flow due to bacterial growth and plant physiological responses (Sicard et al., 2018). Symptoms in the canopies of infected plants include foliar discoloration, wilting of apical shoots, dieback of twigs and branches, and general decline (Carlucci et al., 2013) that may be confounded with a generic water stress responses due to the xylem-limiting nature of the infection processes (Hopkins, 1989). Several studies in different hosts have shown that disease progress and severity of disease caused by *Xf* increase with water stress (Purcell and Hopkins, 1996), e.g. liana vines (McElrone et al., 2001), and citrus (Machado et al., 2007). In almond trees, almond leaf scorch (ALS) symptoms mainly consist of marginal scorching of leaves, often with a golden yellow band between the brown or tan necrotic tissue and the inner green parts, with necrosis generally starting at the leaf margins and progressing to the midrib (Marco-Noales et al., 2021).

Xf bacterium grows in the xylem, a plant niche characterized by low concentrations of amino acids and other compounds. Sabella et al., 2018 showed that lignin deposition plays an important role in plant defense, suggesting that an increase in lignin amounts in the xylem may reduce *Xf* movement, slowing disease progression. It is known that the availability of minerals and trace elements affects host-pathogen interactions, especially pathogen survival, the expression of virulence traits, and host physiology (Cruz et al., 2012). Thus, there is evidence that *Xf*-infection causes changes in the mineral status of the host plant that may contribute to successful infection and symptom development. Indeed, it is reported that *Xf*-infection caused significant increases in concentrations of calcium (Ca) prior to the appearance of symptoms and decreases in concentrations of phosphorous (P) after symptoms appeared (De La Fuente et al., 2013).

The links between the physiological symptoms and defense mechanisms of *Xf*-infected plants, and the pigments and plant processes that can be assessed from remote sensing (RS), open avenues for Earth Observation to support the detection of *Xf*-infection in plants. In this regard, significant progress has been made in recent years towards discriminating *Xf*-infected from healthy plants based on hyperspectral imaging in the visible and near infrared regions (Poblete et al., 2020; Zarco-Tejada et al., 2018, 2021) and the short-wave infrared region (Camino et al., 2021) of the electromagnetic spectrum. Zarco-Tejada et al. (2018) showed that, in the *Xf* outbreak in Southern Italy, changes in specific plant physiological traits, estimated using hyperspectral and thermal RS data, can detect *Xf* asymptomatic infections before symptoms are visible in olive trees. Camino et al. (2021) showed, in almond orchards in the *Xf* outbreak area in Alicante province (Spain), that *Xf*-infected trees can also show symptoms resembling water stress and/or nutrition deficiency that compromises leaf-physiological processes in the plant. Other RS approaches have linked changes in chlorophyll (C_{ab}) and chlorophyll-fluorescence emissions to early phases of plant diseases (e.g., Calderón et al., 2013; Newman et al., 2003; Ye et al., 2020). Poblete et al. (2021) showed that plant traits and spectral indices derived from hyperspectral and thermal imagery can be used to discriminate visual symptoms caused by two vascular pathogens, *Xf* from *Verticillium dahliae* in olive trees. In this regard, Zarco-Tejada et al. (2021) demonstrated the existence of divergent pathogen- and host-specific spectral fingerprints for biotic and abiotic stresses that can disentangle pathogen-induced symptoms from those caused by water stress.

The most common RS approach used to estimate physiological traits are based on relationships between vegetation indices (VIs) or spectral bands and leaf components, such as C_{ab} (Haboudane et al., 2002). Other RS studies use regression models to capture relationships between spectral features and biophysical traits. Partial least squares (PLS) regression has been successfully implemented to retrieve leaf area index (LAI), C_{ab} (Kanning et al., 2018), potassium (K) and P contents (Pimstein et al., 2011) and to generate predictive models for maximum carboxylation rate (V_{cmax}) of Ribulose-1,5-Bisphosphate carboxylase (RuBisCO) enzyme, maximum electron transport rate and percentage leaf N (Meacham-Hensold et al., 2019). However, the finding from these empirical methods is difficult to transfer to other settings or scales. Radiative transfer (RT) models can overcome these limitations by providing a physics-based instruments to retrieve plant biophysical traits from leaf to canopy scales (Jacquemoud and Baret, 1990; Wang et al., 2015). Leaf and canopy parameters in crops can be estimated by coupling the PROSPECT-PRO leaf reflectance model (Féret et al., 2021) with the fourSAIL turbid medium canopy radiative transfer model (Verhoef and Bach, 2007). However, photosynthesis parameters are dependent on environmental conditions and therefore require more complex modeling strategies. Recent studies have successfully estimated V_{cmax} retrievals from satellite data to quantify the solar-induced chlorophyll fluorescence (SIF) (Guan et al., 2016; Koffi et al., 2015). Or even by employing hyperspectral imaging (Camino et al., 2019) using the Soil Canopy Observation, Photochemistry and Energy fluxes (SCOPE) model developed by van der Tol et al. (2009).

In recent years, machine learning algorithms have gained importance in RS, including detection of plant pests and diseases and to retrieve biophysical traits through inversion approaches coupled with so-called hybrid methods (see Verrelst et al. (2019) for a full review). These hybrid methods combine physical models with parametric or non-parametric models such as support vector machines (SVM), PLS regression, gaussian process regression or ensemble learning methods like random forest (RF). For example, Berger et al. (2020) used a hybrid ML method with Gaussian regression to estimate crop N based on the coupled fourSAIL model with the leaf PROSPECT-PRO model. Despite these demonstrations, hurdles remain to be tackled before it can be used for systematic and widespread, pest detection. These include the algorithm choice, the need to balance training and testing datasets, and feature selection for ML models (Verrelst et al., 2019). In this regard, we need to develop a ML framework that could identify the right ML algorithm based on a robust set of performance metrics such as accuracy, interpretability, complexity, scalability, and computational cost associated to multicollinearity issues. While there is no universal method that can be applied across ML problems, a modeling framework that helps tackle the aforementioned challenges can provide a roadmap for retrieving biophysical plant traits from hyperspectral data and RT models.

In this work, we develop and assess the robustness of an operational ML framework to detect *Xf*-infected almond trees under two well differentiated water regimes. For this, we explore the feasibility of using a parsimonious ML model by adding a novel set of plant traits, such as V_{cmax} and carbon-based constituent (CBC; including lignin deposits), in addition to leaf biochemical constituent, leaf N, structural traits and thermal-based indicator to distinguish *Xf*-infected trees from heathy almond trees across two different water regimes. Previous works (e.g., Poblete et al., 2021; Zarco-Tejada et al., 2021) used fluorescence emission and spectral indices as proxy of photosynthesis in the early *Xf* detection, but in this work, we track photosynthetic capacity with a single leaf parameter, namely the maximum carboxylation rate of the enzyme RuBisCO.

2. Material and methods

2.1. Study area and field data collection

The study site was located in Majorca Island (Balearic Islands, Spain) that is one of the most important almond-producing regions in Spain. Its climate is typically Mediterranean, with warm average temperatures and seasonal rains. We selected 14 almond orchards in different locations of the island within five municipalities covering a wide range of ALS incidence and severity (Table 3). The orchards included different water regimes (Supplementary Fig. S1), almond cultivars, plant density, and crop age or tree crown size. Orchard selection was based on information provided by the Agricultural Service, Conselleria de Medi Ambient, Agricultura I Pesca of the Regional Government of the Balearic

Islands. We carried out a field and airborne campaign during the 8th and 12th of July 2019 that included the visual, molecular diagnosis of *Xf*-infection, leaf physiological measurements and airborne hyperspectral and thermal imagery. The meteorological conditions (air temperature, vapor pressure deficit, air relative humidity, and daily rainfall) registered by the nearest meteorological stations during the days of the field campaign are shown in Supplementary Fig. S4. To account for non-disease-induced water stress in the detection of *Xf*, we used *Xf*-free orchards under irrigated conditions as reference (orchards 1, 7–1, 7–2 and 15 in Table 3).

The *Xf* disease severity (DS) assessment consisted of a visual inspection of each almond tree using a rating scale between 0 and 5, where zero refers to no visual symptoms (i.e., asymptomatic), one, two, three and four refer to trees with visual *Xf* symptoms in between 1 and 25%, 25–50%, 50–75% and 75–95% of the tree-crown, respectively, and five corresponds to a practically dead tree or a tree with mostly dead branches (Supplementary Fig. S2). To validate the visual evaluation and the RS methods we tested *Xf*-infection in the studied almond orchards through recombinase-polymerase-amplification (RPA) using the AmplifyRP XRT + test (Agdia®, Inc., Elkhart, IN) (Harper et al., 2010) using crude extracts from leaf petioles following manufacturer's instructions. Almond sampling from mature branches was performed according to the latest version of the standard protocol PM 7/24(4) for *Xf* diagnosis of the European and Mediterranean Plant Protection Organization (EPPO, 2019). About 10% of the trees in each orchard were sampled ($N = 360$) and included all disease severity levels found within each orchard. Additionally, *Xf* -infection from a subset of samples was determined by DNA extraction from leaf petioles using the CTAB protocol and by the qPCR assay of Harper et al. (2010, erratum 2013) according to EPPO (2019).

We conducted leaf physiological measurements using a hand-held Dualex 4 device (Force-A, Orsay, France; Supplementary Fig. S3) on 104 almond trees randomly selected across six rainfed orchards and nine irrigated ones (Supplementary Fig. S1). We carried out the field measurements in a two-day window around the day of the aerial flights. In those almond trees identified as *Xf*-symptomatic based on visual inspection, we split the leaf measurements into two groups (asymptomatic and *Xf*-symptomatic leaves). The C_{ab}, Nitrogen Balance index (NBI) and the anthocyanin (A_{nth}) content were measured on 1471 asymptomatic leaves and 846 *Xf*-symptomatic ones. On the same trees, leaf samples were collected immediately after recording the Dualex readings for chemical analysis of macro- and micronutrients. Total N concentration (N_{total}) was determined using a LECO TruSpec C/N (St. Joseph, MI, USA) elemental analyzer, and P, K Ca, Iron (Fe), Zinc (Zn) and Magnesium (Mg) content were determined by inductively coupled plasma with optical emission spectroscopy (ICP-OES) using a Thermo iCAP 6500 Series Duo instrument (Thermo Scientific, Leicestershire, UK). These analyses were carried out at the Ionomic Service at CEBAS-CSIC (Spain). We used two-way analysis of variance (ANOVA) to assess the effects of *Xf*-infection on Dualex readings and macro-micronutrients, the effect of the water regimes on the measurements and their interaction between *Xf*-infection and the water regime. The statistical analysis of the field data was conducted in R software (R Core Team, 2020).

2.2. Airborne campaign

The airborne campaign was carried out on 9-10th July 2019 flying with the heading of the aircraft on the solar plane and at 250–300 m above ground level (AGL) altitude and 140 km/h flight speed. High-resolution hyperspectral and thermal images were obtained using two micro-hyperspectral imagers and a thermal camera installed in tandem on a Cessna aircraft (Fig. 1), all operated by the Laboratory for Research Methods in Quantitative Remote Sensing (QuantaLab), IAS-CSIC.

The two hyperspectral sensors recorded in tandem, from the same aircraft, in the visible-near infrared (Headwall Photonics, Fitchburg, MA, USA) and in the near infrared and shortwave infrared region (Micro-Hyperspec NIR-100, Headwall Photonics). The micro-hyperspec VNIR sensor is a linear-array hyperspectral camera configured to record 320 spectral bands in 12-bit. Measurements covered the visible-near infrared between 400 and 800 nm, yielding a 6.4 nm FWHM with a 25-μm slit. The frame rate was set at 50-fps with an integration time of 18 ms. We used a lens with an 8 mm focal length, which combined with the IFOV of 0.93 mrad and an fFOV angle of 50° resulted in a swath of 540 m at 50 cm spatial resolution from the aircraft.

The micro-hyperspec NIR-100 camera was set up with a configuration of 165 spectral bands and 16-bit radiometric resolution in the spectral region of 950–1750 nm, yielding 6.05 nm FWHM with a 25-μm slit and an optical aperture of f/1.4. After the spectral calibration, a Cornerstone 260 1/4 m Monochromator (model 74,100; Oriel Instruments, USA) and the XE-1 Xenon Calibration Light Source (Oceanic

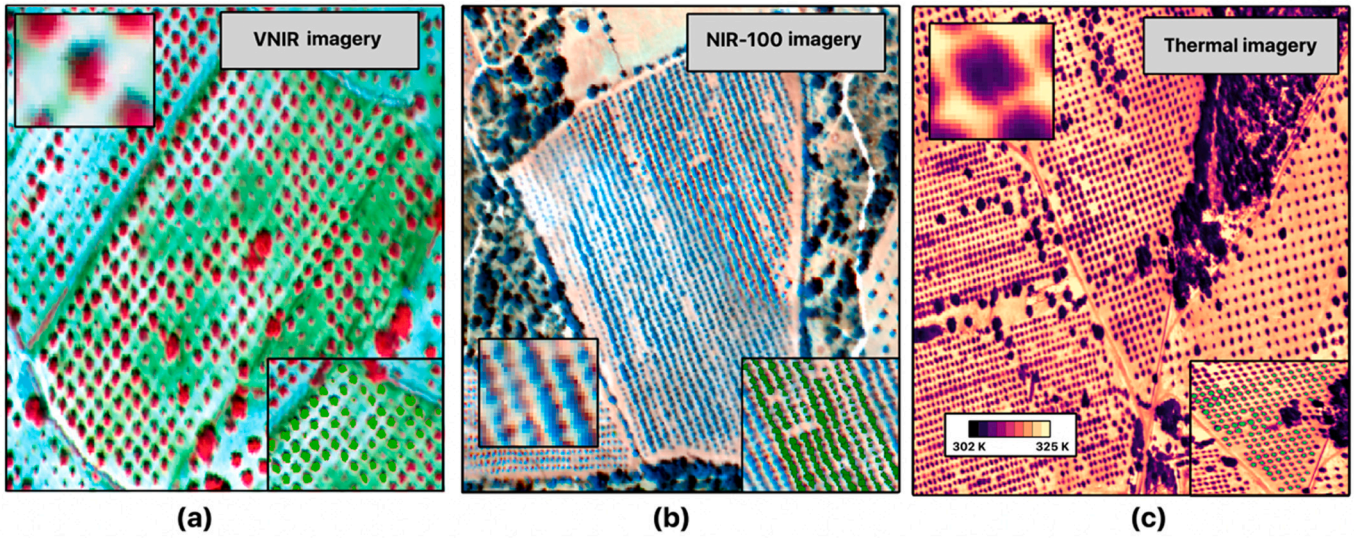

Fig. 1. Images of an almond orchard taken with a hyperspectral VNIR sensor (a, composite: 800 (R), 679 (G) and 540 (B) nm), hyperspectral NIR-100 sensor (b, composite: 985 (R), 1285 (G) and 1550 (B) nm) and thermal camera (c). The spatial resolution of the images is 50 (a), 60 (b), and 40 (c) cm per pixel. Inset panels show sections of the images in greater detail with the automatic segmentation of the tree crowns in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Optics, USA) were used to derive the FWHM and the center wavelength for each spectral band. The images were taken with a frame rate of 50 fps at an integration time of 40 ms and an angular FOV of 38.6°. The 12.5-mm focal length lens allowed recording images with a spatial resolution of 60 cm.

Thermal images were captured by a FLIRSC655 camera (FLIR Systems, Wilsonville, OR, USA) that was installed alongside the hyperspectral cameras. The thermal imagery was obtained in the 7.5–13 μm range, with a resolution of 640 × 480 pixels and a 13.1 mm optics focal length. The images were taken with an angular FOV of 45° and a ground resolution of 50 cm. The radiometric and atmospheric corrections were made as described in detail in Camino et al. (2021).

2.3. Operational machine learning framework for *Xf*-infected trees detection

In this study, we developed an operational machine learning framework to detect *Xf*-infected almond trees growing in orchards with two water regimes. The framework combines four ML stages: 1) individual tree-crown segmentation (Fig. 2A); 2) V_{cmax} and plant trait retrievals using biophysical models (Fig. 2B), 3) N_{total} estimation (Fig. 2C); and 4) generation of a parsimonious ML to detect *Xf*-infected trees (Fig. 2D). The operational ML framework was implemented in Python 3.6 and in R software (R Core Team, 2020). The following sections describe the individual stages.

2.3.1. Tree-crown segmentation

We developed a streamlined framework to detect and outline individual tree crowns in the hyperspectral and thermal imagery (Fig. 2A). As a first step, we co-registered all images using a single orthophoto as reference image. We then performed segmentations on each of the images to delineate tree crowns (Supplementary Fig. S5). This step avoids errors associated with the different spatial resolution of the imaging sensors (50, 60 and 40 cm per pixel for the VNIR, the NIR-100 and the thermal sensor, respectively). For each hyperspectral sensor, we used two spectral bands to generate a vegetation index (VegI) mapping based on the normalized difference vegetation index (NDVI; Rouse et al., 1973). We built the VegI-Mapping in the VNIR hyperspectral scenes taking advantage of the strong absorption of chlorophyll in the 675 nm band and the high reflectance at 800 nm produced by pure vegetation pixels. In contrast, for the NIR-SWIR hyperspectral scenes we calculated the VegI-Mapping based on the strong absorption of the proteins at 1510 nm and the strong absorption peak found at 1080 nm for the water content. First, we performed an unsupervised K-means clustering to separate almond tree-crowns from soil background for each VegI-Mapping and thermal scene. To avoid misclassified pixels at the edge of tree crowns, we then used a second classifier (an SVM algorithm) to remove those pixels classified by K-means algorithm as tree pixels that might be more related to spectral artifacts or soil effects. In the SVM algorithm, to reduce the tree-crowns we used the full reflectance bandsets for both hyperspectral sensors, while for the thermal sensor, we used the available individual thermal images. For the SVM algorithm, we carried out a random spatial sampling procedure where each pixel is classified from the value obtained in the K-means classification. We then split the random spatial points into two groups: the training points with 75% of the spatial points and the testing points with the remaining 25%. Then, we used the reflectance spectra of each hyperspectral scene as well as the thermal information (in kelvin degrees) in the thermal scenes to train a SVM model to separate individual tree crowns from other covers (i.e., soil).

The non-linear SVM classification method was applied using the radial basis function kernel and a cost function to penalize errors associated with the misclassification of tree crowns. We evaluated the performance of each SVM model based on a confusion matrix, calculated from the testing points, and derived statistical metrics (Supplementary Table S1). As a final step, a boundary refinement of the predicted tree-crown was performed to avoid isolated pixels and to be able to separate tree crowns that were close together. Therefore, based on the centroid and the area of each tree (Supplementary Fig. S5), we segmented and isolated each tree crown. We used the spatial centroid of the trees to minimise any error in the spectral assignment of each almond tree. This step enabled the union between the two hyperspectral sensors, field data, and thermal information. The tree-crown segmentations kept the vegetation pixels inside tree-crowns in each high-resolution image, removing the edges of the almond trees that are usually composed of soil-contaminated pixels, shadows, and other canopy background components.

After the automatic tree crown extraction conducted in the hyperspectral and thermal scenes, we extracted the average radiance and reflectance spectra and the tree crown temperature from the individual segmented tree crowns using each high-resolution sensor.

2.3.2. Multi-tier ML approach for retrieving V_{cmax} and plant traits

We designed a Multi-tier ML approach (Fig. 2B) to retrieve plant traits at tree-crown scales through different ML models and simulations using two RT models: (i) PROSAIL-PRO model which couples the PROSPECT-PRO leaf reflectance model with the fourSAIL turbid medium canopy radiative transfer model; and (ii) SCOPE v2.0 which integrates various radiative transfer models and one leaf biochemical photosynthesis model (Farquhar et al., 1980) with an energy balance model. We made 50,000 simulations with the RT models and the plant trait ranges described in Tables 1 and 2 (for PROSAIL-PRO and SCOPE, respectively), to create a look-up table (LUTs) linking trait combinations and reflectance. To avoid potential ill-posed solutions (Combal et al., 2003) in the simulations, we constrained all inputs based on field measurements, information from literature, and preliminary simulations to ensure that the resulting LUTs were within the range of the observations made by the hyperspectral sensors over the tree crowns. In addition, for the SCOPE LUT, we applied a uniform distribution transformation for V_{cmax} and varied each plant trait within the specified range shown in Table 2. For the PROSAIL-PRO LUT, pigments and LAI followed a gaussian distribution, except for LIDFa and viewing angles where we instead used a uniform distribution. Furthermore, for carotenoids (C_{ar}) and A_{nth} we distributed the data with a multivariate normal distribution function that maintained a correlation with C_{ab}. In addition, we fixed Cbrown to 0 for both LUTs because Cbrown having a relatively low impact on canopy reflectance and does not greatly contribute to the spectral differentiation of variations in plant functioning traits (Kattenborn et al., 2018).

To avoid noise in the data affecting the model inversion, we applied smoothing of the simulated spectra at 1 nm resolution using a Savitzky-Golay derived calculation method (Savitzky and Golay, 1964). We then resampled each simulated spectrum to adjust its resolution to the bandwidth of the sensors using Gaussian spectral response functions defined by the FWHM values of each sensor (FWHM = 6.5 nm for the VNIR hyperspectral sensor and FWHM = 6.05 nm for the NIR-100 hyperspectral sensor).

To make the estimates, the resampled spectra simulated via PROSAIL-PRO and SCOPE models were first randomly partitioned into two groups, the training sample with 80% of the simulations and the testing sample with the 20%. We then trained various ML models using three regression algorithms namely, Random Forest (Leo, 2001), SVM (Boser et al., 1992), and a 3-layer Neural Network (NNe). We used these three ML algorithms to retrieve each plant trait using the LUTs configuration described in Tables 1 and 2. We estimated V_{cmax}, standardized to a reference temperature for 25 °C by exclusively using SCOPE inversions based on simulated reflectance and radiance spectra. In particular, we cropped the radiance spectra to 740–790 nm, while two spectral ranges were used for reflectance spectra: (i) 500–800 nm and (ii) 640–800 nm. As shown in Camino et al. (2019), the spectral region located at 740–790 nm is suitable for retrieving V_{cmax} based on radiance spectra due to its close link with the chlorophyll fluorescence emission. The

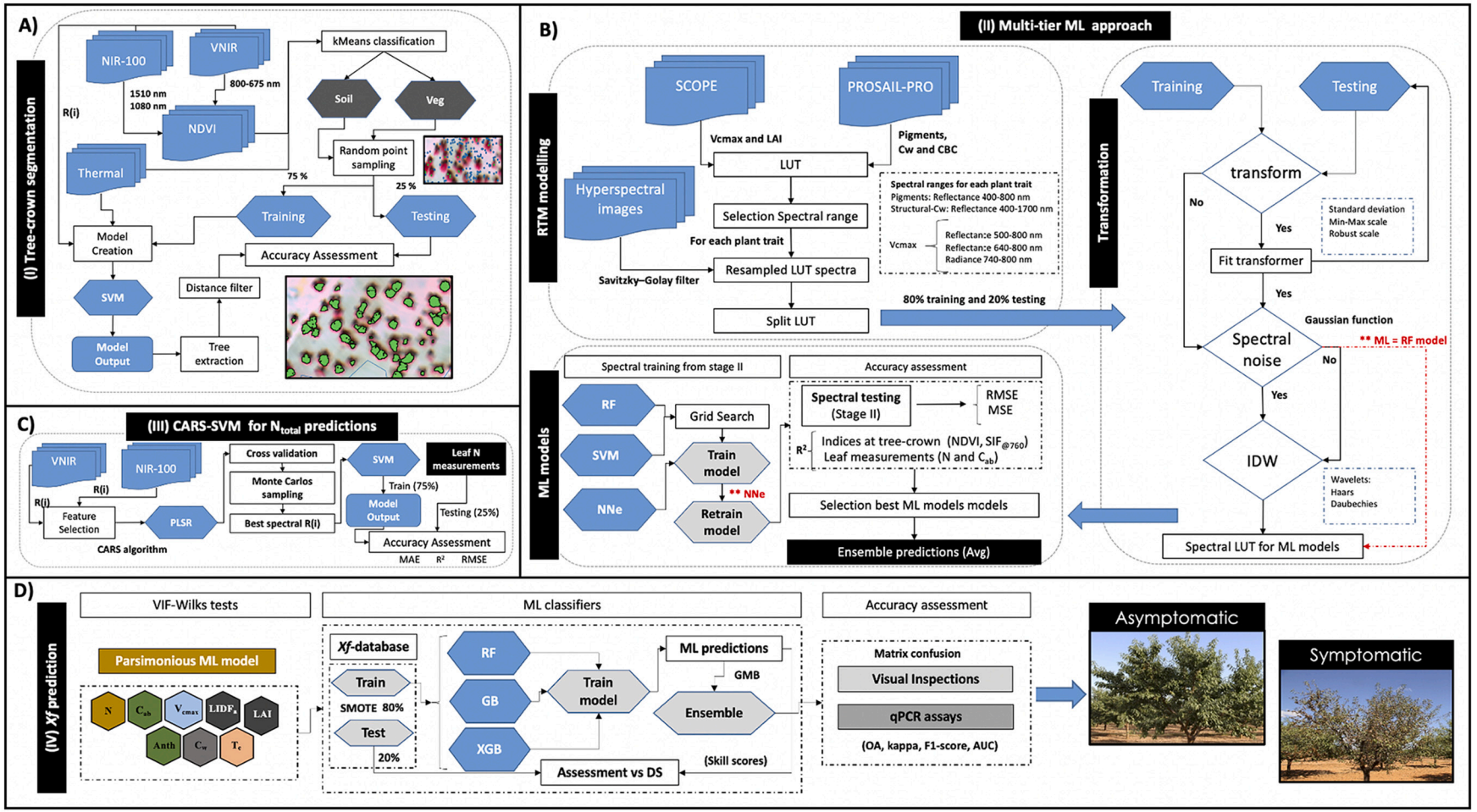

Fig. 2. Overview of the automated machine learning framework for *Xylella fastidiosa* detection using SCOPE and PROSAIL-PRO models, leaf measurement and hyperspectral imagery. For details on the individual panels in the figure, please see section 2.3 in Material and methods.

Table 1
Input variables used in PROSAIL-PRO and the statistical distributions used to generate the Look-up tables for model inversions.

Parameters	Description	Range	Unit	Distribution
PROSPECT-PRO				
N	Leaf structure parameter	1.3–3.2	[−]	Gaussian $_{Avg.=2.2;\ std=0.2}$
C_{ab}	Chlorophyll a + b content	7–52	µg/cm^2	Gaussian $_{Avg.=30;\ std=5}$
C_{ar}	Carotenoid content	0–11	µg/cm^2	Multivariate normal distr.
A_{nth}	Anthocyanin content	0–7	µg/cm^2	Multivariate normal distr.
C_{brown}	Brown pigment	0	[−]	…
C_w	Equivalent water thickness	0.009–0.017	cm	Gaussian $_{Avg.=0.009;\ std=0.002}$
Proteins	Protein content	0.001–1.85	mg/cm^2	Gaussian $_{Avg.=0.001;\ std=0.00025}$
CBC	Carbon-based constituents	0.001–2.55	mg/cm^2	Gaussian $_{Avg.=0.001;\ std=0.002}$
SAILH-5B				
LAI	Leaf area index	0.3–3.8	[−]	Gaussian $_{Avg.=3\ std=0.5}$
$LIDF_a$	Leaf angle distribution (a)	30–90	deg	Uniform
$LIDF_b$	Leaf angle distribution (b)	0	[−]	…
Hspot	Hotspot parameter	0.01–0.50	deg	Uniform
tts	Solar zenith angle	0–45	deg	Uniform
tto	Observer zenith angle	0	deg	…
psi	Relative azimuth angle	0	deg	…

Table 2
Input variables used in SCOPE model and statistical distributions used for generating the LUT.

Parameters	Description	Range	Unit	Distribution
Leaf biophysical and biochemical parameters				
N	Leaf structure parameter	1.5–2.5	[−]	Uniform
C_{ab}	Chlorophyll a + b content	10–60	µg/cm^2	Uniform
C_{ar}	Carotenoid content	0–15	µg/cm^2	Uniform
A_{nth}	Anthocyanin content	0–7.5	µg/cm^2	Uniform
C_s	Senescence factor	0	[−]	….
C_w	Equivalent water thickness	0.004–0.015	cm	Uniform
C_m	Dry matter content	5–25	mg/cm^2	Uniform
Lead biochemistry				
V_{cmax}	Maximum carboxylation rate	0–150	µmol/m^2·s	Uniform
m	Ball-berry conductance	8	[−]	….
Rd_{param}	Dar respiration	0.015	[−]	….
K_v	Extinction coef. For v-profile V_{cmax}	0.64	[−]	….
K_c	Cowan's water use efficiency	500	[−]	….
Ps	Stress multiplier for V_{cmax}	0.05	[−]	….
Stressfact	Stress impact on V_{cmax}	1	[−]	….
kNPQs	Rate thermal dissipation	0	[−]	….
qLs	Fraction active photosystems	1	[−]	….
Fqe	Fraction of photons partitioned to PSII	0.02	[−]	….
Canopy parameters				
H_c	Canopy height	30.05	m	….
L_w	Leaf width	0.07		….
LAI	Leaf area index	0.5–4	m^2/m^2	Uniform
$LIDF_a$	Leaf angle distribution (a)	−0.5 − 0.5	[−]	Uniform
$LIDF_b$	Leaf angle distribution (b)	−0.5–0.5	[−]	Uniform
Micrometeorological				
p	Air pressure	999	hPa	….
U	Wind speed	3	m^{-1}	….
Oa	O_2 concentration in the air	209	per mille	….
Ea	Atmospheric vapor pressure	15	hPa	….
Ca	CO_2 concentration air	380	Ppm	….
Ta	Air temperature	31	°C	….
R_{in}	Incoming shortwave radiation	800	W·m^{-2}	….
R_{il}	Incoming longwave radiation	300	W·m^{-2}	….

selected reflectance windows cover the spectral region most sensitive to changes in Vcmax (Suarez et al., 2021), which are the green region (505–560 nm) and the chlorophyll fluorescence emission peaks generated by the photosystems I-II (650–800 nm). To retrieve the remaining plant traits, we used the simulated reflectance from both biophysical models using exclusively spectral regions where they have the greatest contribution: In the case of pigments (C_{ab}, C_{ar} and A_{nth}), we used the VNIR placed between 400 and 800 nm. However, for structural parameters (LAI and $LIDF_a$), and CBC, we selected the coupled VNIR and SWIR (400–1700 nn). For the C_w we use the spectral region >900 nm.

As a pre-processing step prior to training, we built, for each ML regressor (SVM, RF, and NNe), different parametrizations using four criteria: i) spectral scaling; ii) feature scaling on the dependent variable; iii) wavelet transform; and iv) gaussian noise. Each combination produced a ML model that we then used to predict each plant trait. As first step, the feature scaling was carried out using different transformations: (i) standardization scaling, (ii) min-max scaling, and (iii) a robust scaling based on the Interquartile Range (IQR). Such data transformations were applied for the non-rule-based algorithms (i.e., SVM and NNe) to enable a more stable and faster convergence during the learning process. Second, a feature scaling was applied on the dependent variable by normalizing the values to the 0–1 range, which ensures that the predictions have a range similar to that of the measurement's values. Third, spectral gaussian noise was added to ensure that the model was sufficiently robust and would generalize well to new data. Finally, we applied discrete wavelet transforms (i.e., the Daubechies wavelets (db38) and the Haar wavelet transformation), which is a signal-processing technique that extracts both spectral and temporal information and is suitable for dynamic signals that are inherent to the spectral reflectance signatures. Several studies based on hyperspectral data demonstrated the benefits of wavelet analysis for spectral smoothing and noise removal (Banskota et al., 2013; Cheng et al., 2011; Kattenborn et al., 2017).

In the RF and SVM, to select the optimum combination of parameters, we applied a randomized search in the parameters space of the training data by sampling 10 random combinations and evaluating them by means of a 3-fold cross-validation (CV). Due to the high computational effort required by the larger parameters space of the NNe algorithm, this randomized search was not performed.

For each ML model (RF, SVM, and NNe), we retrained the ML models and evaluated their performance by calculating the root-mean square error (RMSE), and the mean square error (MSE) against the test set. For each plant trait, the best ML models were selected based on these performance metrics in the validation step (Supplementary Tables S3-S7). We also eliminated those ML models with predictions outside of the training conditions used in the LUTs. A third ML selection was then applied based on relationships found between the predictions for each plant trait and the leaf measurements (e.g., C_{ab}, N_{total}) and the most relevant spectral indices derived from hyperspectral and thermal sensors (Supplementary Fig. S12-S18). This step was used to discard those pre-

selected ML models with unrealistic predictions. Finally, the best 5–10 ML algorithms with their most suitable parametrizations for each plant trait were selected, after which their corresponding predictions were averaged to obtain an ensemble prediction. The latter was then used in the succeeding steps as described in Section 2.3.4. In the third ML selection, given the relationship between V_{cmax} and SIF (Camino et al., 2019), we used SIF signal retrieved by the Fraunhofer Line Discrimination (FLD) method (Plascyk and Gabriel, 1975), to select the best ML model for V_{cmax} and the remaining plant traits. However, for LAI we relied on the NDVI, which has been widely used in the literature to estimate LAI indirectly. We decided to use NDVI, despite its saturation problems in tree crowns with high vegetation density, for two reasons: i) we ruled out measuring LAI during the field campaign, as it is a time-consuming measurement; and ii) we found that the NDVI retrievals at tree-crown had a range between 0.18 and 0.70 (Fig. 5C); which is minimally affected by saturation phenomena typically of dense vegetation canopies.

The ensemble predictions were assessed to evaluate whether each plant trait showed significant differences between asymptomatic tree crowns (DS = 0) and tree crowns showing *Xf* symptoms (DS ≥ 1). For this assessment, Kruskal-Wallis test or the Wilcoxon post-hoc test with Bonferroni correction were performed for each plant trait to evaluate the differences among the severity classes under two water regimes. The Multi-tier ML approach was implemented in Python and the statistical analysis in R software.

2.3.3. Total nitrogen concentration

We estimated the N_{total} of individual tree-crowns using a competitive adaptive reweighted sampling (CARS) method with a non-parametric SVM regression algorithm (Fig. 2C). It ingested the reflectance between 400 and 1750 nm. Prior to the feature selection, we performed a first derivative transformation of the reflectance spectra in the spectral signal. We then applied the CARS method to reduce the number of spectral bands and to identify those that contributed most significantly to the prediction of N_{total} in almond trees in the study. The CARS approach is a robust algorithm for screening vital variables by coupling PLS regression with CV techniques (Li et al., 2009). It comprises three steps: first, an optimal combination of reflectance bands from full spectra was selected using PLS with CV. Depending on the importance of each wavelength, *n* subsets of wavelengths were sequentially chosen from N-Monte Carlo sampling runs in an iterative and competitive manner. Finally, the wavelength subset with the lowest root-mean-square error of CV (RMSE-CV) was selected for the ML model. This band set was used in a non-linear SVM regression to estimate leaf total nitrogen concentration. The best radial function and *c* parameter were found using a grid search method for CV. The r^2, the RMSE, and the mean absolute error (MAE) between the measured leaf nitrogen concentration and predicted N_{total} values were used as skill scores to validate the performance of each model (Supplementary Table S8). The statistical analysis was conducted in R software. To avoid overfitting on the training sample, 10-fold CV with 5 repeats was used in the SVM models predicting each plant-trait.

2.3.4. Parsimonious machine learning for the early detection of Xylella fastidiosa

We assessed the performance of the different ML classifiers used for *Xf* detection and determined the most parsimonious ML (pML) model in three water scenarios in which the presence of *Xf* is confirmed by a molecular diagnostic test: (i) almond trees growing under irrigated conditions; (ii) almond trees growing under rainfed conditions; and (iii) almond trees growing in both watering conditions. We assessed the multicollinearity based on variance inflation factor (VIF) analysis to find the optimal set of plant traits related to physiological processes and plant traits that can be embedded in a ML model to detect *Xf*-infected trees in irrigated and rainfed almond orchards (Fig. 2d). We then assessed the ability of the selected remote sensing-based physiological parameters to detect the presence of *Xf*-infected trees either asymptomatic or symptomatic (DS = 0, vs DS > 0) using an RF classifier, a gradient boosting (GB) machine (Friedman, 2001), and the advanced X gradient boosting (X-GB) model (Chen and Guestrin, 2016). We conducted the *Xf* detection based on three decision tree-based classifiers as previous works have largely demonstrated the effectiveness of such algorithms in plant pest detection (e.g., Wójtowicz et al., 2021; Zarco-Tejada et al., 2021; Xu et al., 2022).

First, the set of plant trait predictor variables were reduced through multicollinearity analysis using VIF with a threshold ≤10 and Wilks' lambda method with an *F*-test decision ≤0.05. VIF-Wilks' lambda allowed us to obtain the most pML model with a minimal number of plant traits. After VIF analysis, pML was built exclusively on pigment-related plant functional traits (C_{ab} and A_{nth}), structure-based functional traits (LAI, $LIDF_a$), C_w, CBC, N_{total}, V_{cmax} and tree crown temperature (T_c) extracted from thermal imagery. From the set of independent variables comprised by the aforementioned plant traits, we generated a training and a validation set containing 80% and 20% of the data, respectively, by means of a stratified random split based on the number of asymptomatic/symptomatic trees to ensure equal proportion of both disease categories in both sets. To deal with the fact that there were fewer symptomatic than asymptomatic trees in our data set (i.e., class imbalance), we first removed from the dataset those orchards with no *Xf*-infection and then, we leveraged the class weight parameter to strongly penalize the misclassification of the minority class. In addition, we also applied the synthetic minority oversampling technique (SMOTE, Chawla et al. (2002)) on the training set, which augments the minority class samples by synthesizing new instances from the existing ones.

The RF algorithm is an ensemble technique that fits decision tree classifiers on sub-samples of the dataset and is one of the most widely used ML methods across domains. Like RF, the GB classifier is also an ensemble of decision trees, but it fits the trees differently. As opposed to RF which builds the decision trees independently and combines the outcome at the end of the learning process, GB builds trees sequentially and aggregates the results along the sequence. X-GB, which is another ensemble tree-based classifier that we used in our empirical experiments, is an improved implementation of GB which implements regularization parameters (L1 and L2) to reduce overfitting. The best parameters for each of these classification models were found using a randomized search method like the parameter tuning process described in Section 2.3.3, but instead using a 10-fold CV.

We conducted a feature evaluation using a permutation importance method (Leo, 2001) based on 100 iterations for each ML algorithm (RF, GB, and X-GB) to analyze the contribution of each plant trait in the pML model. Finally, we assembled the ML predictions with a stacking GB machine model to benchmark the classification accuracy of the ML disease detection models using the visual assessment and the subset of trees analyzed by the RPA molecular diagnostic test to determine their *Xf*-infection status. The classification accuracies of the ML algorithm for the pML model were evaluated by calculating the overall accuracy (OA, in %), the kappa value **κ** (Cohen, 1960), the F1-scores and the Area-Under-the-Curve (AUC) scores.

3. Results

3.1. Disease incidence and severity

Four of the orchards (i.e., 1, 7–1, 7–2 and 15) under irrigation showed no visual *Xf* symptoms. In the other three irrigated orchard, 90.3 ± 4.4%, 8.9 ± 4.0% and 0.8 ± 0.4% of the trees were asymptomatic (DS = 0), showing early *Xf*-symptoms (DS = 1) or moderate to severe *Xf*-symptoms (DS > 1). In the nine rainfed orchards, 39.8 ± 15.7%, 29.8 ± 9.1% and 30.4 ± 11.3% of the trees were asymptomatic (DS = 0), showing early *Xf*-symptoms (DS = 1) or moderate to severe *Xf*-symptoms (DS > 1), respectively (Table 3). Results of molecular diagnosis using the RPA test are presented in Table 3. In total, 71% (*n* = 148) and 31% (*n* =

Table 3
Percentage of almond trees with different *Xylella fastidiosa* (*Xf*) disease status identified during the field campaign in 14 almond orchards under irrigated ($n = 2815$ trees) and rainfed conditions ($n = 1593$ trees). The disease severity (DS) classification ranged between 0 and 4, where DS = 0 refers to asymptomatic trees and DS = 1, DS = 2, DS = 3 and DS = 4 refer to trees with visual *Xf* symptoms between 1 and 25%, 25–50%, 50–75% and > 75% of the tree-crown. The number of almond trees tested by Recombinase-polymerase amplification (RPA) molecular diagnostic test is also shown.

Irrigation regimen	DS_0	DS_1	DS_2	DS_3	DS_4	DS_{1-4}	RPA	
Rainfed							negative	positive
Orchard 4	92.8%	6.7%	0% %	0.5%	0%	7.2%	7	8
Orchard 8	24.1%	46.6%	23.5%	5.6%	0.3%	75.9%	9	20
Orchard 9	3.3%	20.7%	41.5%	34.3%	0.2%	96.7%	2	23
Orchard 10	41%	27%	20.2%	11.2%	0.6%	59%	11	22
Orchard 16	9.2%	61.4%	24.5%	3.8%	1.1%	90.8%	8	18
Orchard 17	68.3%	16.4%	3.3%	7.1%	4.9%	31.7%	9	11
Irrigated								
Orchard 1	100%	0%	0%	0%	0%	0%	10	0
Orchard 7–1	100%	0%	0%	0%	0%	0%	52	2
Orchard 7–2	100%	0%	0%	0%	0%	0%	25	0
Orchard 11	96%	4%	0%	0%	0%	4%	26	17
Orchard 12	68.8%	28.3%	2.9%	0%	0%	31.2%	12	23
Orchard 13	84.2%	14.9%	0.4%	0.2%	0.2%	22.2%	10	9
Orchard 14	77.8%	19.8%	2%	0.4	0%	22.2%	8	12
Orchard 15	100%	0%	0%	0%	0%	0%	8	0

214) of trees analyzed in the rainfed orchards and irrigated orchards, respectively, tested positive for *Xf*-infection.

3.2. Leaf field measurements

Overall, *Xf*-asymptomatic leaves displayed a better physiological and nutritional status than that of *Xf*-symptomatic leaves in both irrigated and rainfed trees (Table 4). For both water regimes, *Xf*-symptoms were associated with lower ($P < 0.05$) C_{ab}, NBI, N_{total}, and C_{total} values but higher ($P < 0.05$) A_{nth}, Fe and Zn. However, Ca, P, K and Mg content depended on the water regime and occurrence of *Xf*-symptoms. On rainfed conditions with high occurrence of *Xf*-symptoms Ca was higher on *Xf*-asymptomatic trees, while P, K, and Mg content was lower on *Xf*-symptomatic trees than asymptomatic trees. No significant differences ($P \geq 0.05$) existed for the remaining macro-nutrients either between the two symptomatic classes or water regimes (Table 4).

To further investigate the effects of *Xf*-infection in C_{ab}, A_{nth} and N_{total}, leaf measurements were performed in almond trees (n = 104) where the presence or absence of *Xf*-infection was confirmed by an RPA *Xf*-specific molecular diagnostic test (Fig. 3). Results showed that *Xf*-infection reduced C_{ab} and N_{total} significantly ($P < 0.05$) in both irrigated and rainfed trees. Conversely, A_{nth} increased in *Xf*-leaves. We highlight that the relationship found between leaf N_{total}, and C_{ab} (Fig. 3C) showed a significant and positive trend ($r^2 = 0.28$ and $P < 0.05$), with a drop in values for both leaf measurements as the severity of *Xf* symptoms increased.

Table 4
Average and standard deviation of leaf measurements performed with a Dualex, and the macro-micronutrients quantified on leaves from almond trees showing no visual symptoms of *Xylella fastidosa* (*Xf*) infection and *Xf*- symptomatic leaves ($n = 104$ and 100 trees Dualex measurements and macro-micronutrients, respectively) grown in orchards under rainfed and irrigated conditions. The number of measurements (n) for each leaf group and method is also shown. Symbols indicate a statistically significant difference ($P < 0.05$) of the studied grouped (asymptomatic vs symptomatic leaves (*), water regimes (+) and their interaction between *Xf*-infection and the water regime (x)).

	Irrigated conditions		Rainfed conditions		*p*-value
	Xf-Asympt.	*Xf*-Sympt.	*Xf*-Asympt.	*Xf*-Sympt.	
Dualex measurements					
C_{ab}	31.50 ± 6.68	23.25 ± 9.23	30.56 ± 6.91	21.84 ± 7.79	*/+/…
NBI	16.21 ± 4.26	12.46 ± 5.90	15.71 ± 4.41	11.27 ± 4.65	*/+/…
A_{nth}	0.03 ± 0.04	0.078 ± 0.069	0.03 ± 0.03	0.073 ± 0.06	*/…/…
$n_{-Dualex}$	653	175	818	671	
macro- micronutrients					
N_{total}	2.09 ± 0.23	1.94 ± 0.32	1.80 ± 0.27	1.67 ± 0.32	*/+/…
C_{total}	41.10 ± 0.87	40.08 ± 1.06	40.94 ± 1.01	40.3 ± 0.89	*/…/…
Ca	2.97 ± 0.60	2.88 ± 0.51	3.22 ± 0.68	3.42 ± 0.79	…/+/…
P	0.098 ± 0.017	0.092 ± 0.027	0.082 ± 0.014	0.079 ± 0.02	…/…/…
K	1.35 ± 0.58	1.12 ± 0.41	1.34 ± 0.55	1.25 ± 0.57	…/…/…
Mg	0.50 ± 0.18	0.63 ± 0.12	0.53 ± 0.13	0.49 ± 0.13	…/…/x
Fe	130.5 ± 44.6	135.7 ± 34.6	109.4 ± 27.45	147.4 ± 40.65	*/+/…
Zn	20.1 ± 10.5	21.95 ± 14.2	17.12 ± 5.4	20.8 ± 7.97	*/…/x
$n_{-samples}$	51	13	54	44	

3.3. Capacity of plant trait retrievals to distinguish Xf-infected trees in two water regimes

C_{ab}, CBC (including cellulose, lignin, hemicellulose, starch, and sugars), A_{nth}, and C_{ar} contents were significantly higher ($P < 0.05$) in asymptomatic almond trees, compared to those showing initial *Xf*-symptoms. Among *Xf*-symptomatic trees (DS > 0), for rainfed trees, the content on these plant traits tended to decrease (C_{ab} and C_{ar}) or increase (CBC and A_{nth}) with the increase in symptom severity, being higher (Cab and C_{ar}) or lower (CBC and A_{nth}) in trees showing initial *Xf*-symptoms compared to those with moderate or severe *Xf*-symptoms, respectively (Fig. 4). We should note that the C_{ab} C_{ar} and A_{nth} predictions were also able to distinguish between trees showing moderate from those showing severe *Xf*-symptoms when growing under rainfed conditions. In contrast, under irrigated conditions, only C_{ar} and CBC could differentiate ($P \geq 0.05$) between asymptomatic trees and trees with *Xf* symptoms. On the other hand, the C_w, the $LIDF_a$ and the tree-crown temperature were able to distinguish ($P < 0.05$) asymptomatic trees from *Xf*-symptomatic trees grown when growing in both water conditions (Supplementary Fig. S7).

Leaf area index and NDVI were higher ($P < 0.05$) in asymptomatic trees and decreased significantly ($P < 0.05$) as *Xf*-symptom got more severe (Fig. 5A, B). Moreover, the Multi-tier ML approach predict LAI reliably. In fact, the relationships found between NDVI (Fig. 5C) for each DS class, which is a spectral indicator widely used to validate LAI

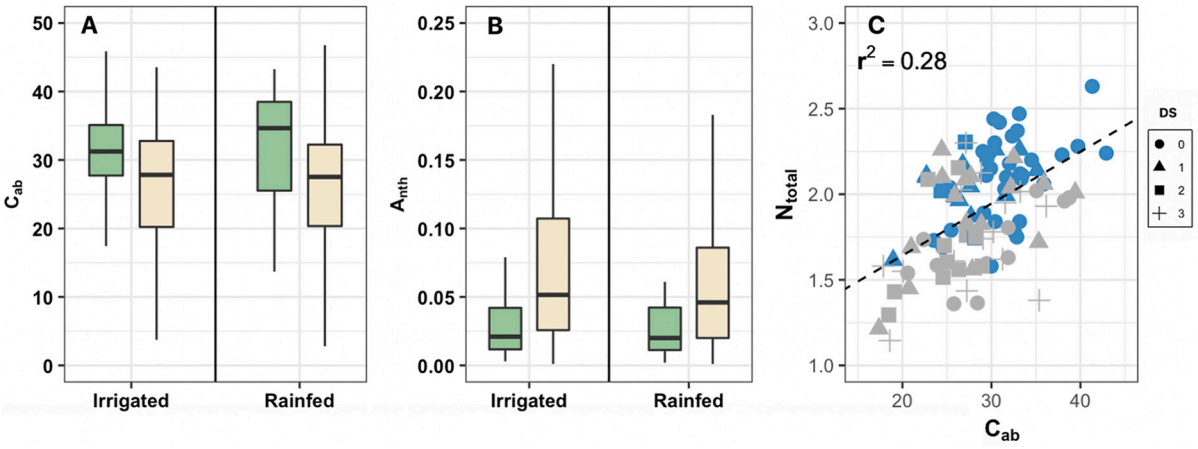

Fig. 3. Chlorophyll content (C_{ab} in µgcm^{-2}; a) and anthocyanin index (b; A_{nth}), in almond trees (n = 104 trees) under two water regimes for asymptomatic (DS = 0 and RPA negative) or sowing almond leaf symptoms (DS > 0) caused by *Xylella fastidiosa* (*Xf*) on which infection by *Xf* was confirmed by an RPA molecular diagnostic test. C) Scatter plot between total nitrogen concentration (N_{total}) and chlorophyll content measured in the leaves of almond trees located in orchards managed under rainfed (in grey) and irrigated (in blue) regimes (where n = 104 trees; DS_0 = 46, DS_1 = 28, DS_2 = 14, DS_3 = 16). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

retrievals, obtained a remarkably strong linear agreements ($r^2 > 0.79$), showing the highest LAI values in those trees growing under irrigated conditions and without initial *Xf*-symptoms.

In two selected orchards with high and low incidence of *Xf*-infection under rainfed conditions., the spatial mapping estimation of pigments based on the proposed ML framework coupled with inverse distance weighting (IDW) techniques tracked well the spatial variability of pigments (e.g., A_{nth}, see Supplementary Fig. S8) or even CBC (Supplementary Fig. S9).

3.4. Total nitrogen, V_{cmax}, and carbon-based constituent

The N_{total} quantified based on the CARS-SVM model showed robust predictive performance when compared to the N_{total} measured at leaf level during the testing phase (RMSE = 0.11% and MAE = 0.09%). Overall, the fitted model (Fig. 6A; r^2 = 0.83) also showed that almond trees growing under irrigated conditions displayed better nutritional N status compared to that under rainfed conditions. We validated V_{cmax} retrievals based on their relationships against N_{total} and chlorophyll content, as leaf assimilation rate measurements could not be performed. This lack of assimilation measurements in the field is due to the complexity of net assimilation measurements on almond leaves, as well as the fact that these measurements on leaves are time consuming and therefore very restricted for operational purposes. A linear relationship was also found between predicted N_{total} and V_{cmax} retrieved by the Multi-tier ML approach using the SCOPE model at tree-crown level (r^2 = 0.61; $P < 0.05$; Fig. 6B) as well as for the relationship between N_{total} and the CBC (Fig. 6C; r^2 = 0.62), indicating that the carbon-based constituent derived using the PROSAIL-PRO model yielded good match with the total nitrogen concentration. Similar agreement was found between the predictions of V_{cmax} and chlorophyll content (r^2 = 0.65; $P < 0.05$; Fig. 6D). In relation to *Xf* incidence, N_{total} predictions and V_{cmax} showed negative trends as *Xf* severity increased in both irrigated and rainfed plants (Fig. 7). In fact, V_{cmax} was significantly higher ($P < 0.05$) in

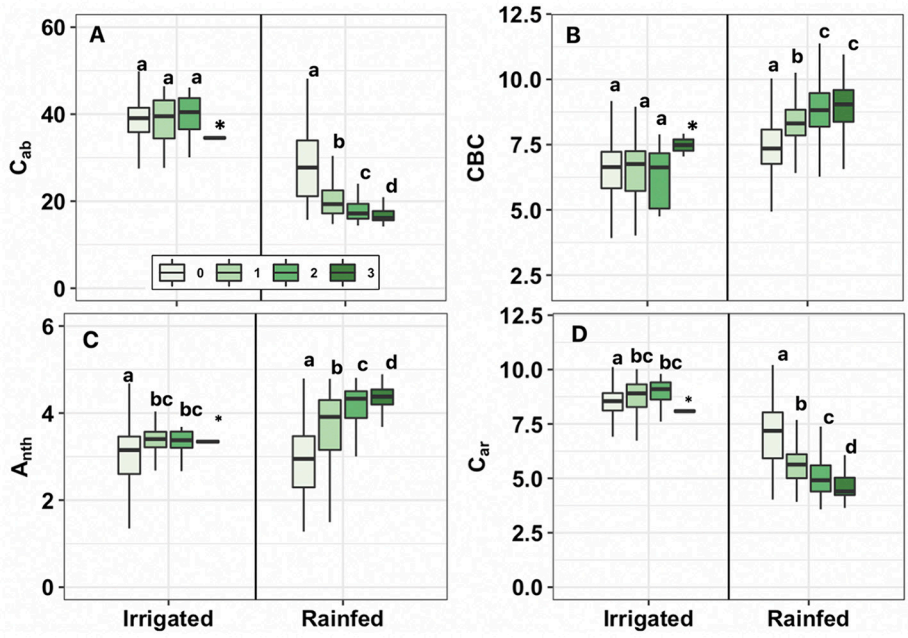

Fig. 4. A) Chlorophyll content (C_{ab} in µg/cm^2), B) carbon-based constituent (CBC in mg/cm^2), C) anthocyanins content (A_{nth} in µg/cm^2), and D) carotenoids content (C_{ar} in µg/cm^2) predicted for almond tree crowns assessed as asymptomatic (DS = 0), to trees with different *Xylella fastidiosa* symptoms severity (DS ≥ 1) under irrigated (DS_0 = 2524, DS_1 = 268, DS_2 = 20 and DS_3 = 2) and rainfed (DS_0 = 506, DS_1 = 483, DS_2 = 370 and DS_3 = 220) conditions. Disease severity levels with the same letter are not significantly different (Wilcoxon post-hoc tests, $P <$ 0.05). Asterisks denote that this class is not used in the Wilcoxon post-hoc due to the low number of cases).

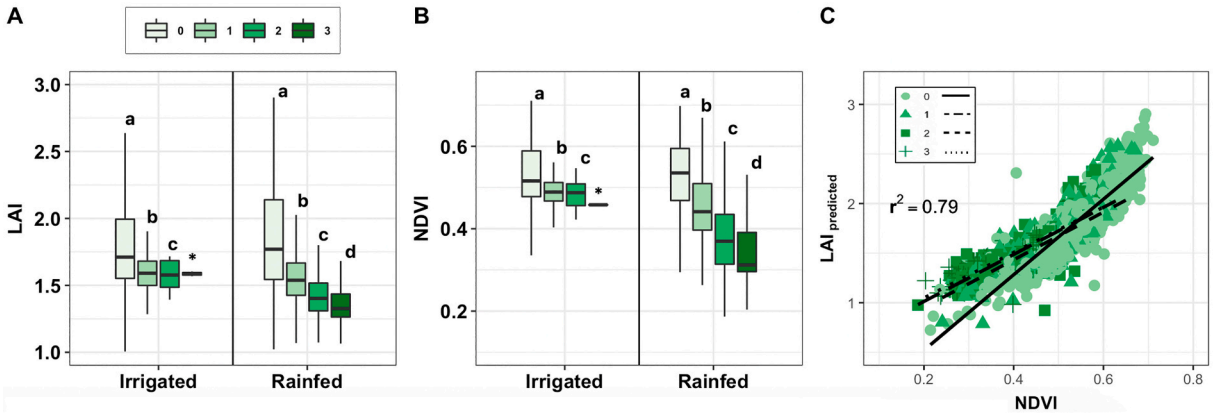

Fig. 5. A) Predicted leaf area index (LAI in m^2/m^2) and B) NDVI at tree-crown assessed as asymptomatic (DS = 0), to trees with different *Xylella fastidiosa* symptoms severity (DS ≥ 1) under irrigated ($DS_0 = 2524$, $DS_1 = 268$; $DS_2 = 20$ and $DS_3 = 2$) and rainfed ($DS_0 = 506$, $DS_1 = 483$, $DS_2 = 370$ and $DS_3 = 220$) conditions. Disease severity levels with the same letter are not significantly different (Wilcoxon post-hoc tests, $P < 0.05$). C) Predicted LAI vs. airborne NDVI retrieved for tree-crowns (b; where $n = 4408$ trees; $DS_0 = 3030$, $DS_1 = 751$, $DS_2 = 390$, $DS_3 = 222$). Asterisks denote that this class is not used in the Wilcoxon post-hoc due to the low number of cases).

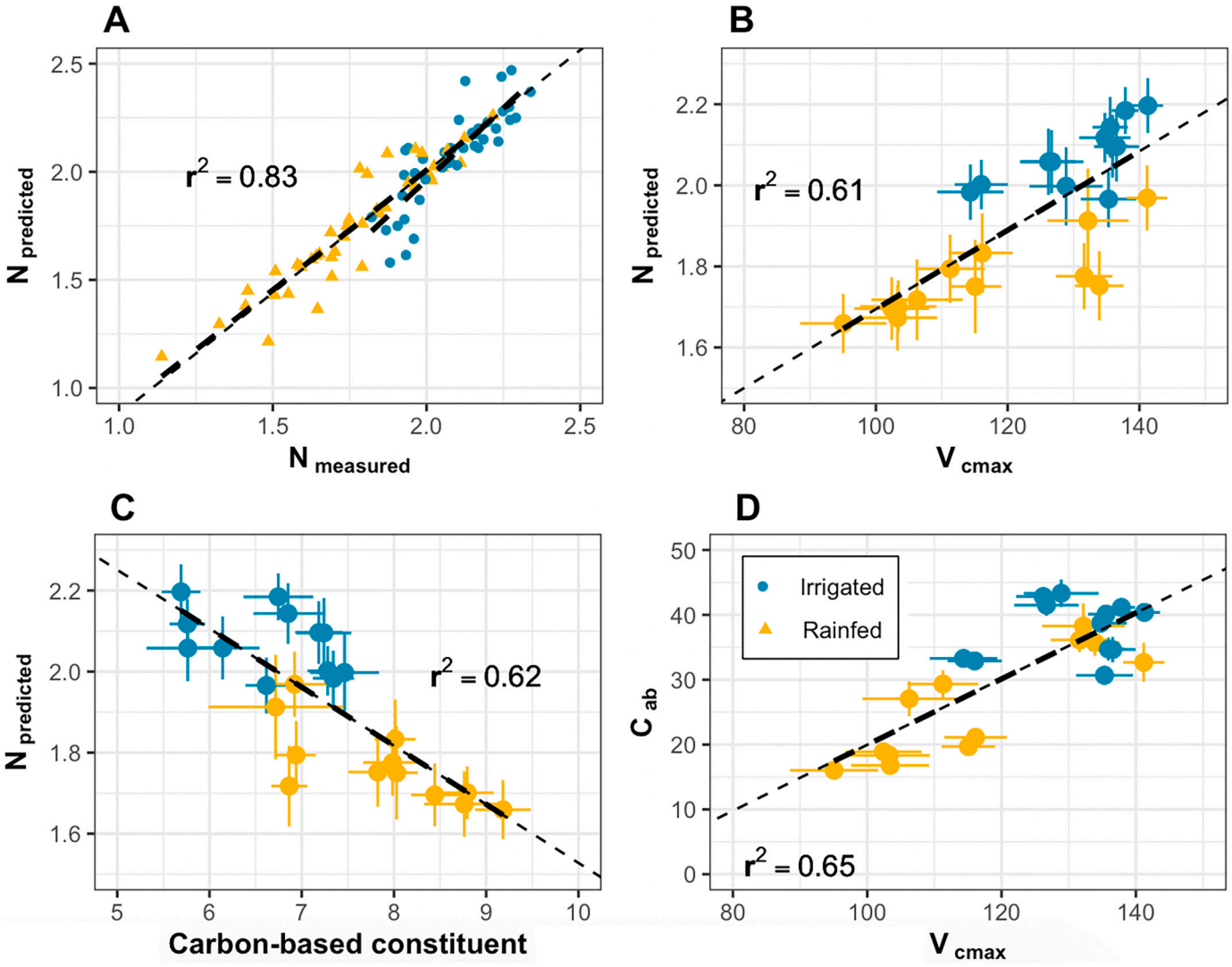

Fig. 6. Relationship between predicted total nitrogen concentration (A, N_{total} in %) and measured N_{total} at tree-crown level (A; in %, $n = 98$), maximum carboxylation rate (B; V_{cmax} in μmol· m^{-2}·s^{-1}) and carbon-based constituent (C; CBC in mg/cm^2) predicted at tree-crown level. For the same trees, the relationship between chlorophyll content (C_{ab} in $\mu g cm^{-2}$) and Vcmax is also shown (D). In (B), (C) and (D), tree values were averaged by orchard, water regimen and incidence of *Xylella fastidiosa* on almond trees.

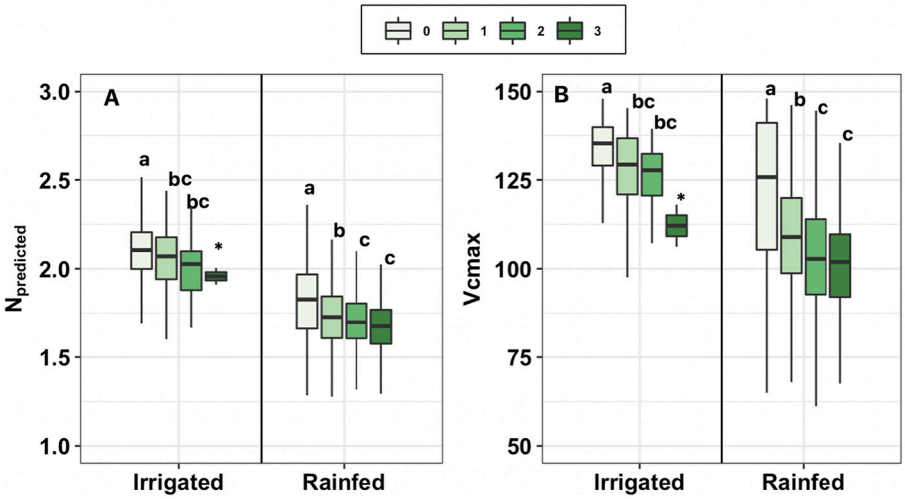

Fig. 7. Predicted total nitrogen (N) concentration (A; %) and maximum carboxylation rate (B; V_{cmax}; in $\mu mol \cdot m^{-2} \cdot s^{-1}$) predicted for almond tree crowns assessed as asymptomatic (DS = 0), to trees with increasing *Xylella fastidiosa* symptoms severity (DS ≥ 1) under irrigated ($DS_0 = 2524$, $DS_1 = 268$, $DS_2 = 20$ and $DS_3 = 2$) and rainfed ($DS_0 = 506$, $DS_1 = 483$, $DS_2 = 370$ and $DS_3 = 220$) conditions. Severity levels with the same letter are not significantly different (Wilcoxon post-hoc tests, $P < 0.05$). Asterisks denote that this class is not used in the Wilcoxon post-hoc due to the low number of cases).

asymptomatic trees than that on *Xf*-symptomatic with the only exception of N_{total} predicted for trees under rainfed conditions. On *Xf*-symptomatic trees, both variables tended to decrease with the increase in *Xf* severity with lower values ($P < 0.05$) estimated in trees showing moderate or severe *Xf*-symptoms that reached similar values ($P \geq 0.05$).

The spatial mapping of V_{cmax} (Fig. 8) and N_{total} patterns (Supplementary Fig. S6) showed that overall almond trees in rainfed orchards with a higher *Xf* incidence and severity had also lower photosynthetic capacity and N_{total} than those with lower *Xf* incidence and severity levels.

3.5. Parsimonious model to detect Xf-infections under different water conditions

The most pML model showed a good performance for the three ML classifiers when fitted to almond orchards affected by *Xf* under either irrigated or rainfed conditions (Table 5). In fact, for pooled data, the OA and AUC reached 83%, and the Cohen's Kappa coefficient exceeded 0.66 for the three ML algorithms (RF, GB, and X-GB). However, the performance of the pML model decreased when fitted to data from each water regime separately, in particular for orchards under irrigated conditions (Table 5). Similar accuracies in OA, F1-score, and AUC were also found when the pML model was fitted for almond trees growing under rainfed conditions, but a decrease in the performance of the model occurred when fitted to trees growing only under irrigated conditions that reduced the Cohen's kappa coefficient up 0.24 and 0.26 values. Differences in the relative proportion of *Xf*-symptomatic trees between both irrigation regimes could account for these differences in model performance. In fact, on pooled data the proportion of asymptomatic and *Xf*-symptomatic trees has a good balance (58% for asymptomatic trees vs. 42% for *Xf*-symptomatic trees), but, not on rainfed orchards in which *Xf*-incidence is much higher (32% for asymptomatic vs 68% of *Xf*-symptomatic trees).

Upon our analysis of the ROC curves for the three ML classifiers and the *Xf* detection parsimonious ML model, the ROC showed that the sensitivity/specificity ratio was highest when data for both water regimes were used for ML classifiers (Fig. 9A). Under rainfed conditions we observed an increase in specificity (85%), but sensitivity decreased to 80% (Fig. 9C). On the contrary, under irrigated conditions, the specificity dropped to values lower than 40%, while the specificity yielded similar values (Fig. 9C), compared to the pooled data. These changes in sensitivity and specificity are probably also associated with differences in the relative proportion of asymptomatic and *Xf*-

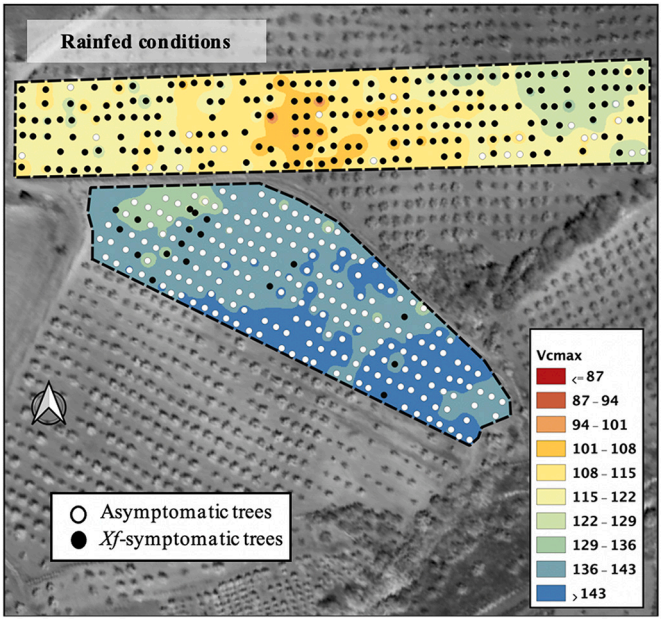

Fig. 8. Map showing the spatial distribution of predicted maximum carboxylation rate (V_{cmax} in $\mu mol \cdot m^{-2} \cdot s^{-1}$) in almond orchards under rainfed conditions. The selected almond orchards were two orchards with high and low incidence of *Xylella fastidiosa*. Inverse distance weighted (IDW) was applied from tree-crowns to generate the spatial map.

Table 5
Overall accuracy, F1-score, Cohen's kappa coefficient and AUC yielded by ML algorithms (RF, GB, and X-GB) for the most parsimonious model (pML) on almond orchards affected by leaf scorch caused by *Xylella fastidiosa* under irrigated and rainfed conditions.

	OA	F1-score	kappa	AUC
Irrigated and rainfed conditions				
Random forest (RF)	84%	0.84	0.67	0.83
Gradient boost (GB)	84%	0.84	0.67	0.84
X Gradient boost (X-GB)	83%	0.83	0.66	0.83
Irrigated conditions				
Random forest (RF)	80%	0.79	0.24	0.61
Gradient boost (GB)	80%	0.79	0.26	0.62
X Gradient boost (X-GB)	79%	0.79	0.25	0.62
Rainfed conditions				
Random forest (RF)	83%	0.83	0.61	0.81
Gradient boost (GB)	82%	0.82	0.59	0.79
X Gradient boost (X-GB)	81%	0.81	0.57	0.79

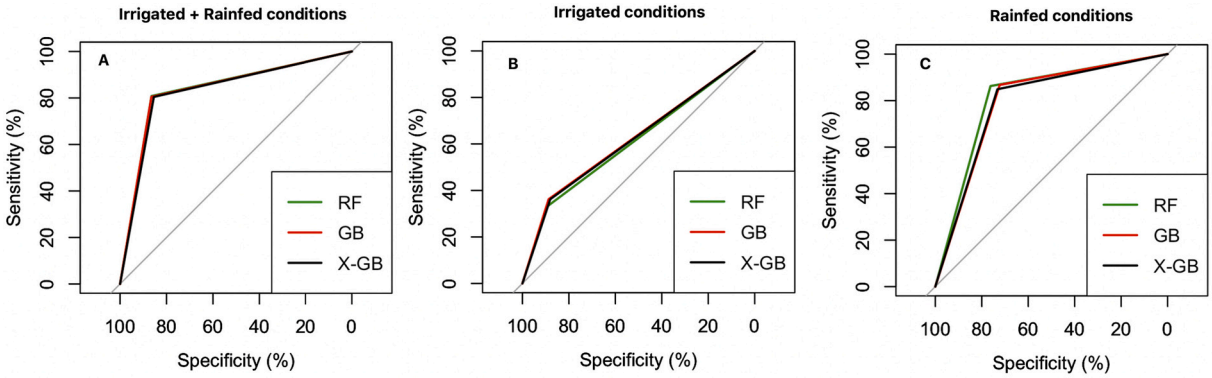

Fig. 9. Receiver operating characteristic (ROC) analysis for detecting *Xylella fastidiosa* symptoms in almond trees for the three studied ML classifiers (RF, GB, and X-GB) and the most parsimonious ML model using the testing samples under irrigated and rainfed conditions (A), irrigated (B) and rainfed conditions (C).

symptomatic tress. Thus, although ML performed well overall to detect *Xf*-symptomatic trees for different disease incidence levels, a more accurate *Xf* detection is reached when *Xf*-incidence is higher than 50%.

3.6. Plant trait contributions in the parsimonious model for Xylella fastidiosa detection

To evaluate the significance of each of the plant traits to detect *Xf*-symptomatic trees, for each ML classifier, we estimated the predictor importance in the best-fitting pML model (Fig. 10). Overall, both *Xf*-incidence and water regime determined the set of plant traits and their contribution in the pML model. Thus, in orchards under irrigated conditions (Fig. 10A) the V_{cmax}, C_{ab}, CBC, $LIDF_a$, and A_{nth} were the plant traits that contributed most to the ML classifiers, with an average contribution of 20%, 17%, 14%, 12%, and 11% respectively (Supplementary Table S2). However, in orchards under rainfed conditions (Fig. 10B; Supplementary Table S2), A_{nth}, and T_c increased their average contribution to 28% and 15% respectively, CBC and $LIDF_a$ kept the same averaged contribution, while the averaged contribution of the V_{cmax} and C_{ab}, decreased to 9% and 8%, respectively. This fact suggested that under water-limited conditions where the proportions of *Xf*-symptomatic trees (75%; where 45% are trees rated with a DS $\geq$ 2 and the

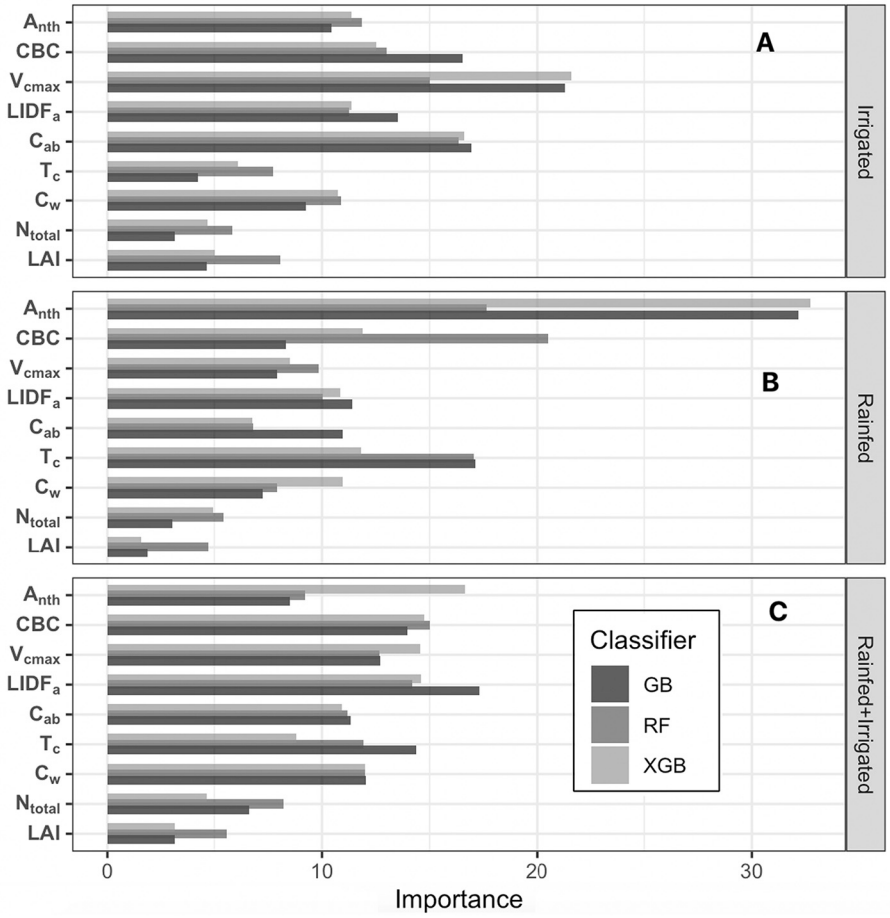

Fig. 10. Plant trait contribution for detecting *Xylella fastidiosa* (*Xf*) symptoms for the three ML classifiers (RF, GB, and XGB) for the most parsimonious ML model under irrigated (A) rainfed conditions (B) and both water regimes (C).

remaining as DS = 1) is higher than asymptomatic trees (26%), the plants traits related to water stress (T_c), $LIDF_a$, and plant defense, such as anthocyanins or lignin deposition (including in CBC) played a significant role when high *Xf* pressure is already present. Finally, when the analysis was performed using almonds under both water conditions (Fig. 10C; Supplementary Table S2), V_{cmax}, C_{ab} and C_w yielded similar averaged importance than A_{nth}, T_c, and CBC, being lower for N_{total} and LAI.

3.7. Assessment of parsimonious model through the RPA molecular diagnostic tests

We evaluated the accuracy of the ML model for *Xf* detection compared to that reached by the visual inspection in the field for the subset of almond trees (n = 360) where the presence/absence of the bacterium was confirmed by molecular diagnosis (Table 6). Additionally, an ensemble GB model was built based on the predictions of the three ML classifiers for *Xf* detection. Overall, individually, the three ML classifier (RF, GB, and X-GB) yielded an overall accuracy >88% with F1-score and UAC > 0.88 and Cohen's kappa coefficient > 0.76 (Table 6). Interestingly, the use of the ensemble GB classifier, increased the performance of the model reaching an overall accuracy of 90% and Cohen's kappa values of 0.80, that are similar to the performance of visual inspections in the field that reached an overall accuracy of 91% and a Cohen's kappa of 0.82 (Table 6).

4. Discussion

In this study, we showed that disease incidence and disease severity were higher in rainfed orchards, compared to orhcards under irrigation. Disease severity in rainfed orchards ranged from 0.08 to 2.07 (on a 0–4 scale), while that in irrigated orchards was 0 in four of the eight orchards and ranged 0.04 to 0.34 in those that showed *Xf* symptoms. As previously stated, *Xf* infection caused a block water flow through the vascular system reduces transpiration and induces water stress, thus causing analogous symptoms that can be confounded with abiotic stress. Irrigated orchards are characterized by water availability but also by current best management that respond better to *Xf* infection.

4.1. Leaf physiological measurements and plant traits retrieval at tree-crown scales

This study showed that pigments (C_{ab} and A_{nth}), plant traits related to photosynthetic capacity (V_{cmax}) and carbon-based constituents are valuable plant traits for the detection of almond trees affected by ALS caused by *Xf* in orchards under two irrigation regimes located in different municipalities in Majorca Island (Balearic Islands).

Our results indicated that *Xf*-infection causes a reduction in C_{ab} and an increase in A_{nth} at both leaf (Table 4) and tree-crown scales (Fig. 4A and C), particularly noticeable in trees managed under rainfed conditions. This is consistent with the results of Zarco-Tejada et al. (2018, 2021) and Poblete et al. (2021) in olive trees affected by the Quick decline syndrome in southern Italy and that of Camino et al. (2021) in almond trees affected in the *Xf*-outbreak in Alicante province of Spain.

Table 6
Overall accuracy, F1-score, Cohen's kappa coefficient and AUC for visual inspections for the most parsimonious ML model (pML) and the ensemble GB model on almond trees where the presence/absence of *Xylella fastidiosa* is confirmed by the RPA molecular diagnostic test (n = 360 trees).

	OA	F1-score	kappa	AUC
pML model				
Random forest (RF)	88%	88.66	0.76	0.88
Gradient boost (GB)	88%	88.89	0.76	0.88
X Gradient boost (X-GB)	89%	90.10	0.79	0.90
Ensemble (stacking RF-GB-XGB)	90%	90.95	0.80	0.90
Visual inspections	91%	91.79	0.82	0.91

In this context, Pereira et al. (2019) also reported an accumulation of anthocyanin in leaves of *Arabidopsis* ecotype Col-0 inoculated with *Xf* subsp. *Pauca*. Interestingly, it is known that *A. thaliana* grown in low P conditions increases anthocyanin content (Sanchez-Calderón et al., 2006) and how *Xf*-infection causes P deficiency in many crops (e.g., De La Fuente et al., 2013) as was the case in our study, in which *Xf*-symptomatic leaves had lower P content (Table 4). In our study, C_{ab} and C_{ar} content showed a similar trend, with a reduction as *Xf* symptoms severity increased (Fig. 4D), likely due to the strong correlation between both plant traits (Garrity et al., 2011). The reduction in N_{total} found in *Xf*-infected leaves (Table 4) and *Xf*-infected trees (Fig. 3A and Fig. 6B) agrees with the findings of Purcino et al. (2007), who suggested that *Xf*-infection disrupts nitrogen metabolism and causes alterations in nitrogen uptake, assimilation and distribution in the plant. This leaf nitrogen deficiency found mainly in trees growing under rainfed conditions (Fig. 6A), could be related to the accelerated leaf senescence due to the coupled effects of *Xf*-infection with water stress (Hsiao, 1973), and could potentially hamper the plant's ability to protect photosynthetic machinery under high light conditions during partial stomatal closure. In this line, (McElrone and Forseth, 2004) also reported how water stress induced an increase in severity of symptoms and progression of *Xf* along the stem of *Parthenocissus quinquefolia*.

We showed that RuBisCO-related V_{cmax} (Fig. 6B) decreased in *Xf*-infected plants for both water conditions. This decrease in V_{cmax} is linked to a decline in photosynthetic activity, which is known to be associated to light-induced inhibition of photosystem II activity (Guidi et al., 2019). Furthermore, V_{cmax} reduction (Fig. 6B and Fig. 7B), is also attributed to an alteration in the metabolic pathways of photosynthesis, such as a reduction in the biochemical capacity to regenerate RuBP (Farquhar and Sharkey, 1982; Tezara et al., 1999). Furthermore, this decrease in V_{cmax} may also be linked to other abiotic factors, such as drought stress or nutrient deficiencies. Under drought conditions and high irradiance levels, trees induce stomatal closure when temperature increases. However, nutrient deficiencies cause chlorosis and symptoms that can be confounded with *Xf* infection. In trees growing under limited irrigation conditions, the reduction of V_{cmax} is more pronounced (Fig. 7B), where *Xf* infection coexists with nutrient deficit (Table 4 and Fig. 7A) and water stress. This latter abiotic stress factor also reduced the photosynthesis capacity of trees by a decline in stomatal conductance in response to increasing canopy temperature (Supplementary Fig. S7C). On the contrary, nitrogen deficiency increases chlorophyll degradation in leaves, limiting the plant photosynthetic capacity during leaf senescence (Fig. 6D). These results confirm that the plant reduced its photosynthetic capacity as ALS symptoms progressed in the almond trees, being more severe in limited water conditions (Fig. 7B). These findings agree with other RS works using high-resolution hyperspectral sensors and SCOPE model, which showed that V_{cmax} drop when water nutrient and water stress are limiting factors (e.g., Camino et al., 2019; Suarez et al., 2021). In general, as *Xf* symptoms increase, we found a decrease in chlorophyll and photosynthetic activity (e.g., Fig. 6D), which is magnified when the *Xf* infection coexists with other abiotic factor, such as water stress or nutritional deficiencies.

We compared the variability of main plant traits obtained at the tree-crown level in *Xf*-affected orchards with reference orchards without *Xf* (Supplementary Fig. S10 and S11). We observed that reference orchards have similar variation in pigment, photosynthetic activity, nitrogen, and tree-crown temperature as non-diseased trees under irrigation, compared to symptomatic trees under rainfed conditions. The degradation of plant traits is more intense under rainfed conditions and in the presence of *Xf*, while under irrigation and good management practices *Xf* induced a smaller change for each plant trait. In this regard, Boyer (2003) stated that a predisposition to disease is often observed in host plants during water deficiencies. Although the biochemical and biophysical causes are not known with certainty; this author proposed two main mechanisms to explain how water stress increases the susceptibility of plants to attack by pathogens: (i) gradients in water potential

can be altered by water deficits and prevent growth in the host without altering growth of the pathogen; and (ii) there is a decrease in photosynthetic activity and protein synthetic activity when water deficits develop that could decrease the synthesis of metabolites and enzymes important for disease resistance (Boyer, 2003). Moreover, McElrone et al. (2003) reported on *Parthenocissus quinquefolia* plants subjected to drought stress and infected with *Xf*, that the major effect of infection by *Xf* occurs due to reduced hydraulic conductance caused by clogging of the vessels, and not increased cavitation and embolism of xylem elements.

We also highlight our findings for carbon-based constituents, which include lignin among other compounds (e.g., cellulose, lignin deposits), that showed an uptick in their levels as *Xf* severity increased (Fig. 4B and Fig. 6C) irrespective of the water regime. This increased lignin synthesis in *Xf*-infected trees is reported as a defense response to pathogen infection, and is a well-known defense mechanism that has been widely demonstrated for different pathosystems (Gayoso et al., 2010; Wallis and Chen, 2012; Xu et al., 2011) including *Xf* in olive (Sabella et al., 2018) and grapevine (Wallis and Chen, 2012).

We remark that the proposed ML framework yielded good performance for all the estimated plant traits and in the two water conditions. In particular, we also highlight the robust linear relationship found between V_{cmax}/CBC and leaf N_{total} (Fig. 6) and between NDVI and LAI (Fig. 5C). Overall, the ML framework showed that the use of ML models could successfully track the changes of plant traits in almond trees growing under irrigation or rainfed conditions, nutritional and *Xf* disease status. Therefore, the robustness of the ML framework allows the transferability of this methodology to other plant species and plant pests. The operational ML frameworks developed in this study with the combination of geostatistical methods (e.g., IDW) would be a powerful tool for the characterization of healthy/disease status on *Xf* outbreak areas in near real time. This spatial mapping of plant traits (e.g., V_{cmax}, leaf N_{total}) enabled us to obtain a spatial scheme of affected orchards from discrete tree-crown data based on the assumption of spatial autocorrelation. Furthermore, the spatial mapping show the potential for monitoring disease progress based on a set of specific plant traits that are sensitive to leaf biochemical changes (e.g., A_{nth}, lignin) and photosynthesis process (V_{cmax}) related to the *Xf*-infection. However, additional field measurements should be also assessed in future studies to be able to quantify the uncertainties identified in the ML models.

4.2. Mineral composition on Xylella fastidiosa-infected leaves

We conducted a mineral analysis at leaf-level to check nutritional differences between leaves sampled from healthy and *Xf*-infected trees. We found that Zn showed higher levels in *Xf*-symptomatic almond leaves than in those showing no disease symptoms and grown under rainfed conditions. Previous studies indicated that Zn is an important virulence factor for *Xf* and is required for efficient colonization of the host with Navarrete et al. (2015) suggesting that the host leaf ionome is correlated with bacterial virulence. In this regard, in our study, Ca concentration significantly increased in leaves with *Xf* symptoms under water-limiting conditions, whereas P significantly decreased in *Xf*-infected leaves (Table 4). These findings are in agreement with De La Fuente et al. (2013), who found that *Xf*-infection causes significant increases in concentrations of Ca prior to the appearance of symptoms and decreases in P concentrations after symptoms appeared. In this same context the higher content in Fe in *Xf*-symptomatic leaves also agree with the results of previous studies (i.e., Silva-Stenico et al., 2009), that found high concentrations of Fe and Zn in citrus leaves affected by *Xf*.

Our findings indicate that by studying mineral constituents may reveal changes associated with *Xf*-infection and symptom development. To successfully capture mineral changes at canopy scale for *Xf* detection with hyperspectral sensors, we need further studies based on upscaling techniques for estimating leaf mineral composition. In this regard, PLS regression techniques shown in other RS works (e.g., Pimstein et al., 2011) or the reflectance-based ML model proposed in this study to quantify N_{total} at tree scale (Fig. 6), could be promising modeling strategies to characterize mineral constituents at canopy scale. Therefore, the feasibility of including mineral composition in new RS studies could reduce the uncertainties of the *Xf* detection ML models.

4.3. Plant trait contributions in ML models for Xylella fastidiosa detection

The most reliable functional traits to discriminate between asymptomatic and *Xf*-symptomatic almond trees were those related to pigment, V_{cmax}, CBC and thermal-based indicator. In particular, the contribution of pigments was confirmed in earlier studies for this same pathogen in olive but focused on VNIR hyperspectral sensors (Zarco-Tejada et al., 2018). Additionally, Camino et al. (2021) demonstrated the relevance of spectral N indicators when developing models for *Xf* detection in almonds at tree crown scales by combining RT models and hyperspectral sensors. In contrast, to the best of our knowledge, there are no RS studies to date that include V_{cmax}, as proxy of photosynthetic capacity, and lignin deposition (including in CBC) in the remote detection of *Xf*-infection. In this study, we have quantified the relative contribution of plant physiological traits when adding V_{cmax} and CBC in the pML model, and their effects when fitted to field data with different disease prevalence and water regimes. In this regard, V_{cmax}, C_{ab}, and CBC, accounted for at least 51% in irrigated orchards, but dropped up to 31% in rainfed conditions. Under these later conditions, A_{nth}, and T_c accounting together up to 43% of the total, followed by CBC, LIDFa and V_{cmax}. This study confirms the findings by Zarco-Tejada et al. (2021), that showed the importance of using a thermal indicator, as proxy of transpiration rates, especially when *Xf*-infection coexists with abiotic factors, such as water stress and nitrogen deficiency. When we analyzed the contribution of plant traits in the pML in almond trees under both water regimes, we found that pigments (C_{ab} and A_{nth}), T_c, C_w, CBC, and V_{cmax} yielded similar contribution in the ML models, while LAI and N_{total} showed the lowest contribution.

4.4. Parsimonious ML model for Xylella fastidiosa detection

In the most pML model for *Xf*-detection, which combined V_{cmax} with thermal-based T_c, CBC, and a limited set of plant traits, the three ML classifiers yielded a higher performance (Table 4) when almond orchards are managed under rainfed conditions (Table 4). However, the ability of the models for *Xf* detection decreased when orchards were under irrigated conditions with a relative low proportion of trees affected by *Xf*-infection. In particular, the three ML classifiers reduced their OA by about 2–3% compared to the performance reached by the pML under rainfed conditions. These results confirm the ability of a parsimonious model to detect *Xf*-infection under heterogeneous disease prevalence and with different types of orchard management (e.g., water, nutritional, or pest management). These results achieved similar accuracies with previous work (e.g., Zarco-Tejada et al., 2018, 2021), where they implemented ML approaches by combining a high number of spectral indices and plant traits derived from RT models. Here, we highlighted that the decision tree-based classifiers could achieve an overall accuracy of >90% when comparing with the molecular diagnosis dataset, when only nine plant traits are included in the pML model. The decision tree-based classifiers we used for *Xf* detection identify the most important input features or variables used in each model; this is important as it permits a feature analysis and identifies the contribution of each plant trait in the model for detecting *Xf* symptoms. Although neural network-based algorithms have been shown to outperform other ML models in many different use-cases, their learning process typically produces features that are too abstract to reveal underlying processes.

5. Conclusions

This study constitutes a significant progress in *Xf* detection through

an operational ML framework that allows the monitoring of photosynthetic activity (e.g., V_{cmax}), pigments, carbon-based constituents and thermal-based T_c changes potentially associated to *Xf*-infection in trees under two water regimes and different orchard management practices. We show that the development of a simple ML model where adding photosynthetic capacity besides plant traits related to transpiration rates, pigments-nitrogen, lignin deposits, and structural decay was able to successfully detect *Xf*-infected trees, obtaining an overall accuracy of 90% when the ML model was compared to trees tested for *Xf*-infection by the RPA molecular diagnostic test. The proposed ML framework reduces the uncertainties of plant traits using a biophysical model and opens up a new approach for large-scale application of ML to different crops and plant pathogens.

CRediT authorship contribution statement

C. Camino: Conceptualization, Methodology, Software, Validation, Data curation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **K. Araño:** Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **J.A. Berni:** Conceptualization, Methodology, Software, Validation, Investigation, Formal analysis, Writing – original draft, Writing – review & editing. **K.H. Dierkes:** Software, Validation, Writing – original draft, Writing – review & editing. **J.L. Trapero-Casas:** Data curation, Writing – review & editing. **G. León-Ropero:** Data curation, Writing – review & editing. **M. Montes-Borrego:** Data curation, Writing – review & editing. **M. Roman-Écija:** Data curation, Writing – review & editing. **M.P. Velasco-Amo:** Data curation, Writing – review & editing. **B.B. Landa:** Conceptualization, Methodology, Data curation, Supervision, Project administration, Funding acquisition, Resources, Writing – original draft, Writing – review & editing. **J.A. Navas-Cortes:** Conceptualization, Methodology, Data curation, Supervision, Project administration, Funding acquisition, Resources, Writing – original draft, Writing – review & editing. **P.S.A. Beck:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition, Resources, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by grant: ITS2017-095: Design and Implementation of control strategies for *Xylella fastidiosa*, Project 5. Government of the Balearic Islands, Spain. Data collection was partially supported by the European Union's Horizon 2020 research and innovation program through gran agreement XF-ACTORS (727987). The authors would like to thank the staff of the QuantaLab Laboratory at the IAS-CSIC (Cordoba, Spain) for their assistance and support provided during the airborne campaigns and the image processing. Special thanks to Professor P.J. Zarco-Tejada for his support and scientific advice during this work. D. Sacristan is also acknowledged for his support during the field campaigns, and M. Gomila for confirmation of *Xf* diagnosis in a selection of almond trees by qPCR analysis. We are also grateful for the information provided by Juan De Dios García, Agricultural Service, Conselleria de Medi Ambient, Agricultura I Pesca of the Regional Government of the Balearic Islands on the phytosanitary status and management system of almond orchards in the affected area. The views expressed are purely those of the authors and may not in any circumstance be regarded as stating an official position of the European Commission.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2022.113281.

References

Banskota, A., Wynne, R.H., Thomas, V.A., Serbin, S.P., Kayastha, N., Gastellu-Etchegorry, J.P., Townsend, P.A., 2013. Investigating the utility of wavelet transforms for inverting a 3-D radiative transfer model using hyperspectral data to retrieve forest LAI. Remote Sens. 5, 2639–2659. https://doi.org/10.3390/rs5062639.

Berger, K., Verrelst, J., Féret, J.-B., Hank, T., Wocher, M., Mauser, W., Camps-Valls, G., 2020. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. Int. J. Appl. Earth Obs. Geoinf. 92, 102174 https://doi.org/10.1016/j.jag.2020.102174.

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory, COLT '92. Association for Computing Machinery, New York, NY, USA, pp. 144–152. https://doi.org/10.1145/130385.130401.

Boyer, J.S., 2003. Biochemical and biophysical aspects of water deficits and the predisposition to disease. Annu. Rev. Phytopathol. 33, 251–274.

Calderón, R., Navas-Cortés, J.A., Lucena, C., Zarco-Tejada, P.J., 2013. High-resolution airborne hyperspectral and thermal imagery for early detection of verticillium wilt of olive using fluorescence, temperature and narrow-band spectral indices. Remote Sens. Environ. 139, 231–245. https://doi.org/10.1016/j.rse.2013.07.031.

Camino, C., Calderón, R., Parnell, S., Dierkes, H., Chemin, Y., Román-Écija, M., Montes-Borrego, M., Landa, B.B., Navas-Cortes, J.A., Zarco-Tejada, P.J., Beck, P.S.A.A., 2021. Detection of xylella fastidiosa in almond orchards by synergic use of an epidemic spread model and remotely sensed plant traits. Remote Sens. Environ. 260, 112420 https://doi.org/10.1016/j.rse.2021.112420.

Camino, C., Gonzalez-Dugo, V., Hernandez, P., Zarco-Tejada, P.J.P.J., 2019. Radiative transfer vcmax estimation from hyperspectral imagery and SIF retrievals to assess photosynthetic performance in rainfed and irrigated plant phenotyping trials. Remote Sens. Environ. 231, 111186 https://doi.org/10.1016/j.rse.2019.05.005.

Carlucci, A., Lops, F., Marchi, G., Mugnai, L., Surico, G., 2013. Has xylella fastidiosa "chosen" olive trees to establish in the Mediterranean basin? Phytopathol. Mediterr. 541–544.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Snopes.Com: two-striped telamonia spider. J. Artif. Intell. Res. 16, 321–357.

Chen, T., Guestrin, C., 2016. XGBoost. Proc. 22nd ACM SIGKDD int. Conf. Knowl. Discov. Data Min. https://doi.org/10.1145/2939672.2939785.

Cheng, T., Rivard, B., Sanchez-Azofeifa, A., 2011. Spectroscopic determination of leaf water content using continuous wavelet analysis. Remote Sens. Environ. 115, 659–670. https://doi.org/10.1016/j.rse.2010.11.001.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Meas. 20, 37–46. https://doi.org/10.1177/001316446002000104.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. 84, 1–15. https://doi.org/10.1016/S0034-4257(02)00035-4.

Cruz, L.F., Cobine, P.A., De La Fuente, L., 2012. Calcium increases xylella fastidiosa surface attachment, biofilm formation, and twitching motility. Appl. Environ. Microbiol. 78, 1321–1331. https://doi.org/10.1128/AEM.06501-11.

De La Fuente, L., Parker, J.K., Oliver, J.E., Granger, S., Brannen, P.M., van Santen, E., Cobine, P.A., 2013. The bacterial pathogen xylella fastidiosa affects the leaf ionome of plant hosts during infection. PLoS One 8, e62945. https://doi.org/10.1371/journal.pone.0062945.

Eppo, 2019. PM 7/24 (4) xylella fastidiosa. EPPO Bull. 49, 175–227. https://doi.org/10.1111/epp.12575.

EFSA (European Food Safety Authority), 2019. Update of the scientific opinion on the risks to plant health posed by Xylella fastidiosain the EU territory. EFSA J. 17 (5), 5665. https://doi.org/10.2903/j.efsa.2019.5665, 200 pp.

EFSA (European Food Safety Authority), 2019. Scientific opinion on the effectiveness of in plantacontrol measures for Xylella fastidiosa. EFSA J. 17 (5), 5666. https://doi.org/10.2903/j.efsa.2019.5666, 17 pp.

EFSA (European Food Safety Authority), 2021. Scientific report on the update of the Xylella spp. host plant database–systematic literature search up to 31 D December 2020. EFSA J. 19 (6), 6674. https://doi.org/10.2903/j.efsa.2021.6674, 70 pp.

Farquhar, G.D., Sharkey, T.D., 1982. Stomatal conductance and photosynthesis. Annu. Rev. Plant Physiol. 33, 317–345. https://doi.org/10.1146/annurev. pp.33.060182.001533.

Farquhar, G.D., von Caemmerer, S., Berry, J.A., 1980. A biochemical model of photosynthetic CO2 assimilation in leaves of C3 species. Planta 90, 78–90. https://doi.org/10.1007/BF00386231.

Féret, J.B., Berger, K., de Boissieu, F., Malenovský, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based

constituents. Remote Sens. Environ. 252 https://doi.org/10.1016/j.rse.2020.112173.
Friedman, J., 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29 (5), 1189–1232. https://doi.org/10.1214/aos/1013203451.
Garrity, S.R., Eitel, J.U.H., Vierling, L.A., 2011. Disentangling the relationships between plant pigments and the photochemical reflectance index reveals a new approach for remote estimation of carotenoid content. Remote Sens. Environ. 115, 628–635. https://doi.org/10.1016/j.rse.2010.10.007.
Gayoso, C., Pomar, F., Novo-Uzal, E., Merino, F., de Ilárduya, O.M., 2010. The ve-mediated resistance response of the tomato to verticillium dahliae involves H2O2, peroxidase and lignins and drives PAL gene expression. BMC Plant Biol. 10, 232. https://doi.org/10.1186/1471-2229-10-232.
Guan, K., Berry, J.A., Zhang, Y., Joiner, J., Guanter, L., Badgley, G., Lobell, D.B., 2016. Improving the monitoring of crop productivity using spaceborne solar-induced fluorescence. Glob. Chang. Biol. 22, 716–726. https://doi.org/10.1111/gcb.13136.
Guidi, L., Piccolo, L., Landi, M., Lo Piccolo, E., Landi, M., 2019. Chlorophyll fluorescence, photoinhibition and abiotic stress: does it make any difference the fact to be a C3 or C4 Species? Front. Plant Sci. 10, 174. https://doi.org/10.3389/fpls.2019.00174.
Haboudane, D., Miller, J.R., Tremblay, N., Zarco-Tejada, P.J., Dextraze, L., 2002. Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. Remote Sens. Environ. 81, 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4.
Harper, S.J., Ward, L.I., Clover, G.R.G., 2010. Development of LAMP and real-time PCR methods for the rapid detection of xylella fastidiosa for quarantine and field applications. Phytopathology 100, 1282–1288. https://doi.org/10.1094/PHYTO-06-10-0168.
Hopkins, D.L., 1989. Xylella fastidiosa: xylem-limited bacterial pathogen of plants. Annu. Rev. Phytopathol. 27, 271–290. https://doi.org/10.1146/annurev.py.27.090189.001415.
Hsiao, T.C., 1973. Plant responses to water stress 0755. Ann. Rev. Plant Physiol. 24, 519–570. https://doi.org/10.1146/annurev. pp. 24.060173.002511.
Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. Remote Sens. Environ. 34, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z.
Kanning, M., Kühling, I., Trautz, D., Jarmer, T., 2018. High-resolution UAV-based hyperspectral imagery for LAI and chlorophyll estimations from wheat for yield prediction. Remote Sens. 10, 1–17. https://doi.org/10.3390/rs10122000.
Kattenborn, T., Fassnacht, F.E., Pierce, S., Lopatin, J., Grime, J.P., Schmidtlein, S., 2017. Linking plant strategies and plant traits derived by radiative transfer modelling. J. Veg. Sci. 28, 717–727. https://doi.org/10.1111/jvs.12525.
Kattenborn, T., Fassnacht, F.E., Schmidtlein, S., 2018. Differentiating plant functional types using reflectance: which traits make the difference? Remote Sens. Ecol. Conserv. 5 (1), 5–19. https://doi.org/10.1002/rse2.86.
Koffi, E.N., Rayner, P.J., Norton, A.J., Frankenberg, C., Scholze, M., 2015. Investigating the usefulness of satellite-derived fluorescence data in inferring gross primary productivity within the carbon cycle data assimilation system. Biogeosciences 12, 4067–4084. https://doi.org/10.5194/bg-12-4067-2015.
Leo, B., 2001. Random forests. Random For. 1–122 https://doi.org/10.1201/9780429469275-8.
Li, H., Liang, Y., Xu, Q., Cao, D., 2009. Key wavelengths screening using competitive adaptive reweighted sampling method for multivariate calibration. Anal. Chim. Acta 648, 77–84. https://doi.org/10.1016/j.aca.2009.06.046.
McElrone, A.J., Forseth, I.N., 2004. Photosynthetic responses of a temperate Liana to xylella fastidiosa infection and water stress. J. Phytopathol. 152, 9–20. https://doi.org/10.1046/j.1439-0434.2003.00794.x.
McElrone, A.J., Sherald, J.L., Forseth, I.N., 2001. Effects of water stress on symptomatology and growth of Parthenocissus quinquefolia infected by xylella fastidio. Plant Dis. 85, 1160–1164.
Mcelrone, A.J., Sherald, J.L., Forseth, I.N., 2003. Interactive effects of water stress and xylem-limited bacterial infection.
Machado, E.C., Oliveira, R.F., Ribeiro, R.V., Medina, C.L., Stuchi, E.S., Pavani, L.C., 2007. Water deficiency intensifies physiological symptoms of citrus variegated clorosis in 'Natal' sweet orange plants. Bragantia 3, 373–379.
Marco-Noales, E., Barbé, S., Monterde, A., Navarro, I., Ferrer, A., Dalmau, V., Aure, C.M., Domingo-Calap, M.L., Landa, B.B., Roselló, M., 2021. Evidence that xylella fastidiosa is the causal agent of almond leaf scorch disease in Alicante, mainland Spain (Iberian Peninsula). Plant Dis. https://doi.org/10.1094/PDIS-03-21-0625-SC.
Meacham-Hensold, K., Montes, C.M., Wu, J., Guan, K., Fu, P., Ainsworth, E.A., Pederson, T., Moore, C.E., Brown, K.L., Raines, C., Bernacchi, C.J., 2019. High-throughput field phenotyping using hyperspectral reflectance and partial least squares regression (PLSR) reveals genetic modifications to photosynthetic capacity. Remote Sens. Environ. 231, 111176 https://doi.org/10.1016/j.rse.2019.04.029.
Navarrete, F., De, L., Fuente, L., 2015. In: Zinc Detoxification Is Required for Full Virulence and Modification of the Host Leaf Ionome by Xylella fastidiosa, 28, pp. 497–507. https://doi.org/10.1094/MPMI-07-14-0221-R.
Newman, K.L., Almeida, R.P.P., Purcell, A.H., Lindow, S.E., 2003. Use of a green fluorescent strain for analysis of xylella fastidiosa colonization of Vitis vinifera. Appl. Environ. Microbiol. 69, 7319–7327. https://doi.org/10.1128/AEM.69.12.7319-7327.2003.
Pereira, W.E.L., Ferreira, C.B., Caserta, R., Melotto, M., de Souza, A.A., 2019. Xylella fastidiosa subsp. Pauca and fastidiosa colonize arabidopsis systemically and induce anthocyanin accumulation in infected leaves. Phytopathology 109, 225–232. https://doi.org/10.1094/PHYTO-05-18-0155-FI.
Pimstein, A., Karnieli, A., Bansal, S.K., Bonfil, D.J., 2011. Exploring remotely sensed technologies for monitoring wheat potassium and phosphorus using field spectroscopy. F. Crop. Res. 121, 125–135. https://doi.org/10.1016/j.fcr.2010.12.001.
Plascyk, J.A., Gabriel, F.C., 1975. The Fraunhofer line discriminator MKII an airborne instrument for precise and standardized ecological luminescence measurement. IEEE Trans. Instrum. Meas. 24, 306–313. https://doi.org/10.1109/TIM.1975.4314448.
Poblete, T., Camino, C., Beck, P.S.A., Hornero, A., Kattenborn, T., Saponari, M., Boscia, D., Navas-Cortes, J.A., Zarco-Tejada, P.J., 2020. Detection of xylella fastidiosa infection symptoms with airborne multispectral and thermal imagery: assessing bandset reduction performance from hyperspectral analysis. ISPRS J. Photogramm. Remote Sens. 162, 27–40. https://doi.org/10.1016/j.isprsjprs.2020.02.010.
Poblete, T., Navas-Cortes, J.A., Camino, C., Calderon, R., Hornero, A., Gonzalez-Dugo, V., Landa, B.B., Zarco-Tejada, P.J., 2021. Discriminating xylella fastidiosa from verticillium dahliae infections in olive trees using thermal- and hyperspectral-based plant traits. ISPRS J. Photogramm. Remote Sens. 179, 133–144. https://doi.org/10.1016/j.isprsjprs.2021.07.014.
Purcell, A.H., Hopkins, D.L., 1996. Fastidious xylem-limited bacterial plant pathogens. Annu. Rev. Phytopathol. 34, 131–151.
Purcino, R.P., Medina, C.L., Martins, D., Winck, F.V., Machado, E.C., Novello, J.C., Machado, M.A., Mazzafera, P., 2007. Xylella fastidiosa disturbs nitrogen metabolism and causes a stress response in sweet orange Citrus sinensis cv. Pêra. J. Exp. Bot. 58, 2733–2744. https://doi.org/10.1093/jxb/erm138.
R Core Team, 2020. R: a language and environment for statistical computing.
Rouse, J.W., Hass, R.H., Schell, J.A., Deering, D.W., 1973. Monitoring vegetation systems in the great plains with ERTS. Third Earth Resour. Technol. Satell. Symp. 1, 309–317 citeulike-article-id:12009708.
Sabella, E., Luvisi, A., Aprile, A., Negro, C., Vergine, M., Nicolì, F., Miceli, A., De Bellis, L., 2018. Xylella fastidiosa induces differential expression of lignification related-genes and lignin accumulation in tolerant olive trees cv. Leccino. J. Plant Physiol. 220, 60–68. https://doi.org/10.1016/j.jplph.2017.10.007.
Sanchez-Calderón, L.S., López-Bucio, J., Chacón-López, A., Gutiérrez-Ortega, A., Hernández-Abreu, E., Herrera-Estrella, L., 2006. Characterization of low phosphorus insensitive mutants reveals a crosstalk between low phosphorus-induced determinate root development and the activation of genes involved in the adaptation of arabidopsis to phosphorus deficiency. Plant Physiol. 140, 879–889.
Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least-squares procedures. Anal. Chem. 36, 1627–1639. https://doi.org/10.1021/ac60214a047.
Schneider, K., van der Werf, W., Cendoya, M., Mourits, M., Navas-Cortés, J.A., Vicent, A., Lansink, A.O., 2020. Impact of xylella fastidiosa subspecies pauca in european olives. Proc. Natl. Acad. Sci. U. S. A. 117, 9250–9259. https://doi.org/10.1073/pnas.1912206117.
Sicard, A., Zeilinger, A.R., Vanhove, M., Scharte, T.E., Beal, D.J., Daugherty, M.P., Almeida, R.P.P., 2018. Xylella fastidiosa: insights into an emerging plant pathogen. Annu. Rev. Phytopathol. 56 (1), 181–202.
Silva-Stenico, M.E., Pacheco, F.T.H., Pereira-Filho, E.R., Rodrigues, J.L.M., Souza, A.N., Etchegaray, A., Gomes, J.E., Tsai, S.M., 2009. Nutritional deficiency in citrus with symptoms of citrus variegated chlorosis disease. Braz. J. Biol. 69, 859–864. https://doi.org/10.1590/s1519-69842009000400013.
Suarez, L., González-Dugo, V., Camino, C., Hornero, A., Zarco-Tejada, P.J., 2021. Physical model inversion of the green spectral region to track assimilation rate in almond trees with an airborne nano-hyperspectral imager. Remote Sens. Environ. 252 https://doi.org/10.1016/j.rse.2020.112147.
Tezara, W., Mitchell, V.J., Driscoll, S.D., Lawlor, D.W., 1999. Water stress inhibits plant photosynthesis by decreasing coupling factor and ATP. Nature 401, 914–917. https://doi.org/10.1038/44842.
van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z., 2009. An integrated model of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy balance. Biogeosciences 6, 3109–3129. https://doi.org/10.5194/bg-6-3109-2009.
Verhoef, W., Bach, H., 2007. Coupled soil-leaf-canopy and atmosphere radiative transfer modeling to simulate hyperspectral multi-angular surface reflectance and TOA radiance data. Remote Sens. Environ. 109, 166–182. https://doi.org/10.1016/j.rse.2006.12.013.
Verrelst, J., Malenovský, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.P., Lewis, P., North, P., Moreno, J., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. Surv. Geophys. 40, 589–629. https://doi.org/10.1007/s10712-018-9478-y.
Wallis, C.M., Chen, J., 2012. Grapevine phenolic compounds in xylem sap and tissues are significantly altered during infection by xylella fastidiosa. Phytopathology 102, 816–826. https://doi.org/10.1094/PHYTO-04-12-0074-R.
Wang, Z., Skidmore, A.K., Darvishzadeh, R., Heiden, U., Heurich, M., Wang, T., 2015. Leaf nitrogen content indirectly estimated by leaf traits derived from the PROSPECT model. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8, 3172–3182. https://doi.org/10.1109/JSTARS.2015.2422734.
Wójtowicz, A.Piekarczyk J., Czernecki, B., Ratajkiewicz, H., 2021. A random forest model for the classification of wheat and rye leaf rust symptoms based on pure spectra at leaf scale. J. Photochem. Photobiol. B Biology 223 (2021), 112278. https://doi.org/10.1016/j.jphotobiol.2021.112278.
Xu, Z., Zhang, Q., Xiang, S., Li, Y., Huang, X., Zhang, Y., Zhou, X., Li, Z., Yao, X., Li, Q., Guo, X., 2022. Monitoring the severity of pantana phyllostachysae chao infestation in moso bamboo forests based on UAV multi-spectral remote sensing feature selection. Forests 13, 418. https://doi.org/10.3390/f13030418.
Xu, L., Zhu, L., Tu, L., Liu, L., Yuan, D., Jin, L., Long, L., Zhang, X., 2011. Lignin metabolism has a central role in the resistance of cotton to the wilt fungus verticillium dahliae as revealed by RNA-seq-dependent transcriptional analysis and histochemistry. J. Exp. Bot. 62, 5607–5621. https://doi.org/10.1093/jxb/err245.

Ye, H., Huang, W., Huang, S., Cui, B., Dong, Y., Guo, A., Ren, Y., Jin, Y., 2020. Recognition of banana fusarium wilt based on UAV remote sensing. Remote Sens. 12 (6), 938. https://doi.org/10.3390/RS12060938.

Zarco-Tejada, P.J., Camino, C., Beck, P.S.A.A., Calderon, R., Hornero, A., Hernández-Clemente, R., Kattenborn, T., Montes-Borrego, M., Susca, L., Morelli, M., Gonzalez-Dugo, V., North, P.R.J.J., Landa, B.B., Boscia, D., Saponari, M., Navas-Cortes, J.A., 2018. Previsual symptoms of xylella fastidiosa infection revealed in spectral plant-trait alterations. Nat. Plants 4, 432–439. https://doi.org/10.1038/s41477-018-0189-7.

Zarco-Tejada, P.J., Poblete, T., Camino, C., Gonzalez-Dugo, V., Calderon, R., Hornero, A., Hernandez-Clemente, R., Román-Écija, M., Velasco-Amo, M.P., Landa, B.B., Beck, P. S.A., Saponari, M., Boscia, D., Navas-Cortes, J.A., 2021. Divergent abiotic spectral pathways unravel pathogen stress signals across species. Nat. Commun. 12, 6088. https://doi.org/10.1038/s41467-021-26335-3.