Defining optimal Hyperspectral Narrowbands as proximal sensing in the early detection of Xylella fastidiosa in olive trees

POLITECNICO DI BARI

D.R.R.S

06

2017

Doctor of Philosophy in Environmental and Building Risk and Development

Coordinator: Prof. Michele Mossa

XXIX CYCLE
Curriculum: Environment and Natural Resources

DICATECh
Department of Civil, Environmental, Building Engineering and Chemistry

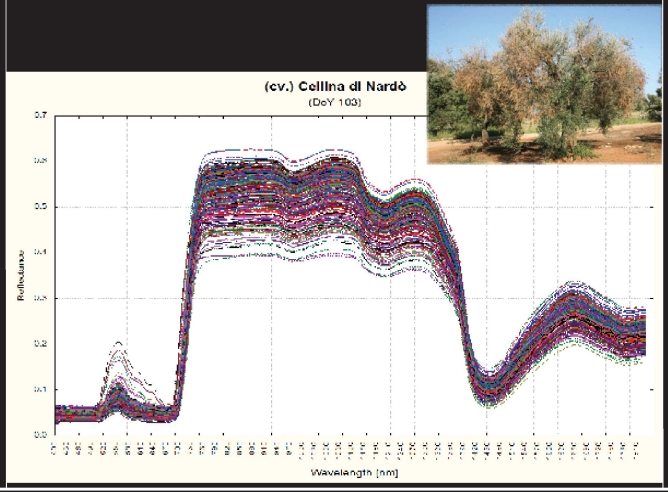
Stefania Gualano

Defining optimal Hyperspectral Narrowbands as proximal sensing in the early detection of *Xylella fastidiosa* in olive trees

Prof. Eufemia Tarantino
Department of Civil, Environmental, Building Engineering and Chemistry
Politecnico di Bari

Dr. Anna Maria D'Onghia
Division Integrated Pest Management
International Centre for Advanced Mediterranean Agronomic Studies,
Mediterranean Agronomic Institute of Bari (CIHEAM/MAIB)

Dr. Chariton Kalaitzidis
Dept of Geoinformation in Environmental Management
International Centre for Advanced Mediterranean Agronomic Studies,
Mediterranean Agronomic Institute of Chania (CIHEAM/MAICh)



Stefania Gualano
Defining optimal Hyperspectral Narrowbands as proximal sensing in the early detection of *Xylella fastidiosa* in olive trees
2017 DICATECh

POLITECNICO DI BARI

D.R.R.S

06

2017

Doctor of Philosophy in Environmental and Building Risk and Development

Coordinator: Prof. Michele Mossa

XXIX CYCLE
Curriculum: Environment and Natural Resources

DICATECh
Department of Civil, Environmental, Building Engineering and Chemistry

Defining optimal Hyperspectral Narrowbands as proximal sensing in the early detection of *Xylella fastidiosa* in olive trees

Prof. Eufemia Tarantino
Department of Civil, Environmental, Building Engineering and Chemistry
Politecnico di Bari

Dr. Anna Maria D'Onghia
Division Integrated Pest Management
International Centre for Advanced Mediterranean Agronomic Studies, Mediterranean Agronomic Institute of Bari (CIHEAM/MAIB)

Dr. Chariton Kalaitzidis
Dept of Geoinformation in Environmental Management
International Centre for Advanced Mediterranean Agronomic Studies, Mediterranean Agronomic Institute of Chania (CIHEAM/MAICh)

Stefania Gualano

POLITECNICO DI BARI

D.R.R.S

06

2017

Dottorato di Ricerca in Rischio e Sviluppo ambientale, territoriale ed edilizio

Coordinatore: Prof. Michele Mossa

XXIX CICLO
Curriculum: Ambiente e Risorse Naturali

DICATECh
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica

Definizione delle bande strette iperspettrali ottimali per il rilevamento prossimale finalizzato all'identificazione precoce di *Xylella fastidiosa* su olivo

Prof. Eufemia Tarantino
Dipartimento di Ingegneria Civile, Ambientale, del Territorio, Edile e di Chimica
Politecnico di Bari

Dr. Anna Maria D'Onghia
Divisione di Protezione Integrata delle Malattie delle Piante
Centro Internazionale degli Alti Studi Agronomici Mediterranei, Istituto Agronomico Mediterraneo di Bari (CIHEAM/IAMB)

Dr. Chariton Kalaitzidis
Dipartimento di Geoinformatica nella Gestione Ambientale
Centro Internazionale degli Alti Studi Agronomici Mediterranei, Istituto Agronomico Mediterraneo di Chania (CIHEAM/IAMCh)

Stefania Gualano

INDEX

ACRONYMS 5

LIST OF FIGURES.....9

LIST OF TABLES.....13

LIST OF EQUATIONS.....15

ABSTRACT English 17

ABSTRACT Italiano 21

CHAPTER 1: INTRODUCTION.....25

1.1 Contest and justification.....25

1.2. The Olive tree 26

1.3 *Xylella fastidiosa* and the Olive Quick Decline Syndrome in Apulia region 28

1.4 Remote sensing of plants 32

1.4.1 Aerial photography and photogrammetry.....33

1.4.2 Use GIS environment to create the plants disease database 36

1.5 Proximal and remote sensing of plants.....37

1.5.1 Hyperspectral RS of plant diseases 42
1.5.3 From univariati methods to multivariati statistics 49

1.6 Necessity and objective of the research 51

CHAPTER 2: Assisted analysis of aerial images at high geometric resolution for the identification of “Olive Quick Decline Syndrome” associated with the *Xylella fastidiosa* bacterium in Apulia 55

2.1 Introduction 55

2.2 Materials and Methods 58

2.3 Results 61

2.4 Conclusions 64

CHAPTER 3: Determining Optimal Hyperspectral Wavebands for detection of *Xylella fastidiosa* using Reflectance data: an Internal Clustering Criteria approach 67

3.1 Introduction 67

3.2 Materials and methods 75
3.2.1 Study Area 75
3.2.2 Plant sampling collection 76
3.2.3 Hyperspectral RS data collection 79
3.2.4 Spectral pre-processing: Savitzky-Golay filter 81
3.2.5 Dimensionality reduction of hyperspectral reflectance data 82
3.2.6 Statistical methods: selection of optimal spectral bands 83
3.2.6.1 Heuristic approach with combined models 83
3.2.6.2 Combined general purpose detection method: interval PCA Internal Clustering Validation (iPCA-ICV) 87
3.2.6.3 Heuristic Approach Model vs interval PCA Internal Clustering Validation 92

3.3 Results 93

3.3.1 Laboratory analysis....93
3.3.2 Comparison of the cultivars: infection rate....93
3.3.3 Spectral analysis....94
3.3.4 Selection of optimal spectral bands through Heuristic Approach model (HAM): LLR^2, PCA and Wilks' Lambda methods....98
3.3.5 Selection of optimal spectral bands through a combined general purpose detection method: iPCA-ICV....103
3.3.5 Best overall model: optimal band identification....110

3.4 Conclusions....113

CONCLUSIONS....115

REFERENCES....119

ACRONYMS

Absorbed Photosynthetically Active Radiation (APAR)
Airborne Visible Infrared Imaging Spectrometer (AVIRIS)
ANalysis Of VAriance (ANOVA)
Anthocyanin Reflectance Index (ARI)
Artificial Neural Networks (ANN)
Blue/Green Index (BIG2)
Calinski-Harabasz (CH)
Canonical Correlation Analysis (CCA)
Chlorophyll Normalized Difference Index (ChlNDI)
Citrus tristeza virus (CTV)
Cluster Analysis (CA)
Cluster Validity Analysis Platform (CVAP)
Colour Infrared (CIR)
Complesso del Disseccamento Rapido dell'Olivo (CoDiRO)
Day of Year (DoY)
Digital Number (DN)
Direct Tissue Blot Immuno Assay (DTBIA)
Double Difference Index (DD)
Electromagnetic (EM)
Enzyme-Linked ImmunoSorbent Assay (ELISA)
European and Mediterranean Plant Protection Organization (EPPO)
European Food Safety Authority (EFSA)
European Union (EU)

Extra Virgin Olive Oil (EVOO)
Feature Selection Methods (FSM)
Full Width Half Maximum (FWHM)
Full-Spectral-Analysis (FSA)
Generalised Discriminant Analysis (GDA)
Geographic Information System (GIS)
Heuristic Approach model (HAM)
Hyperspectral Reflectance (HR)
Index-SPAD
International Olive Council (IOC)
Interval PCA Internal Clustering Validation (iPCA- ICV)
Istituto Nazionale di Statistica (ISTAT)
Istituto Servizi per il Mercato Agricolo Alimentare (ISMEA)
Lambda-Lambda R^2 model (LLR^2)
Leave One Out Cross Validation (LOOCV)
Levels of Severity (LS)
Lichtenthaler Indices (Lic)
Matrix Laboratory (Matlab)
Modified Chlorophyll Absorption in Reflectance Index (MCARI)
Modified Chlorophyll Absorption Integral (mCAI)
Modified Simple Ratio (mSR)
MultiLayer Perceptrons (MLP)National Plant Protection Organization (NPPO)
Multiple Linear Regression (MLR)
Near Infrared (NIR)
Normalized Differential Vegetation Infex (NDVI)
Normalized Phaeophytinization Index (NPQI)
Olive Quick Decline Syndrome (OQDS)
Optimal Vegetation Index (Viopt)
Partial Least Squares Regression (PLSR)
Pest Disease Detection (PDD)
Photochemical Reflectance Index (PRI)
Pigment Specific Normalized Difference (PSNDa)

Pigment Specific Normalized Difference (PSNDb)
Pigment Specific Normalized Difference (PSNDcar)
Plant Senescence Reflectance Index (PSRI)
Polymerase Chain Reaction (PCR)
Principal Component Analisys (PCA)
Principal Component regression (PCr)
Proximal Sensing (PS)
Quadratic Discriminant Analysis (QDA)
Red Edge Position (REP)
Red Edge Vegetation Stress Index (RVSI)
Red Green Blue (RGB)
Reflectance (R)
Remote Sensing (RS)
Self-Organizing Maps (SOM)
Short Wave Infrared (SWIR)
Silhouette (S)
Spectral Angle Mapper (SAM)
Spectral Feature Fitting (SFF)
Spectral Vegetation Index (SVI)
Spectral Vegetation Indices (SVIs)
Stepwise Discriminant Analysis (SDA)
Stepwise Multiple Linear Regression (SMLR)
Structure Insensitive Pigment Index (SIPI)
Support Vector Machine (SVM)
Thermal Infrared (TIR)
Unione Nazionale Associazioni PRoduttori di OLive (UNAPROL)
Unmanned Aerial Vehicle (UAV)
Virgin Olive Oil (VOO)
Visible (VIS)
Water Index (WI)
Xylella fastidiosa (Xf)
Yellow Index (YI)

LIST OF FIGURES

Fig. 1/1 - Olive-producing countries in the Mediterranean region 28

Fig. 1/2 - World distribution of *Xylella fastidiosa* (EPPO, 2016).....29

Fig.1/3 - The Quick Decline of Olive trees in the outbreak area of *Xylella fastidiosa*.....30

Fig. 1/4 - The official demarcated area for *Xylella fastidiosa* in Apulia Region (EU Decision 2015/789). 31

Fig. 1/5 - A: The interaction of leaf tissue with light depends on structural and leaf chemical properties. B: During pathogenesis, plant pathogens influence leaf structural and chemical properties, and by this process the leaf optics are altered (Mahlein, 2015)38

Fig. 1/6 - Typical spectral signature of vegetation by Pu and Gong (2011) 40

Fig. 1/7 - Chlorophyll a fluorescence to photosynthesis from the photosystem to the canopy level and to leaf level (Porcar-Castell et al., 2014).....41

Fig. 2/1 - Clear cut symptoms of OQDS in olive tree.....56

Fig. 2/2 - Subject of the study – the infected area circled in red. Olive groves are in green. 58

Fig. 2/3 - Images in false color, used in the photointerpretation phase manual, within the GIS framework 59

Fig. 2/4 - Characteristic symptoms of OQDS on olive trees infected by Xf: a) “leaf scorch”, b) desiccation of branches and c) Quick decline of the tree 61

Fig. 2/5 - Key photo types selected for the three levels of gravity (white frame: mild; yellow frame: medium; red frame: high) 62

Fig. 2/6 - OQDS classified olive trees..... 63

Fig. 2/7 - Photointerpretation of OQDS trees inspected in fields and which tested positive (circled in red) to serological and molecular testing in June 2014. In the enlarged area, trees that resulted positive from the official monitoring have been overlapped for contrasting effect (circled in cyan) 64

Fig. 3/1 - The study area: small olive orchard located in Gallipoli, Province of Lecce 75

Fig. 3/2 - Localization of the plant subjects according to the cultivar..... 76

Fig. 3/3 - Leaf scorching and twig and branch die-back observed in the orchard..... 77

Fig. 3/4 - Sampling scheme..... 78

Fig. 3/5 - Collected twigs are divided into 3 samples (on the right) with specific codes... 79

Fig. 3/6 - Acquisition scheme 80

Fig. 3/7 - *Xylella fastidiosa* Infection rate according to the cultivar studied..... 94

Fig. 3/8 - Field hyperspectral measurement (n=1380): (cv.) Cellina di Nardò (A) and (cv.) Leccino (B) 96

Fig. 3/9 - Correlation matrix LLR2, relative to both varieties, in function with the 143 narrowbands: (a) Cellina di Nardò and (b) Leccino 99

Fig. 3/10 - UNIQUE correlation matrix (LLR2), relative to both varieties, in funzione with the 143 narrowbands: (a) Cellina di Nardò and (b) Leccino 100

Fig. 3/11 - Spectrum of separable potentials between healthy and infected leaves of both varieties: (a) Cellina di Nardò and (b) Leccino 101

Fig. 3/12 - Excellent narrowbands selected from HAM model to select leaves infected or not infected by *Xf*, superimposed on the relative spectral signature average: (a) Cellina di Nardò and (b) Leccino.....102

Fig. 3/13 - Selection of the observations (spectral signatures of the leaves) deemed "extreme cases", relative to the two varieties, through the dissimilarity function. The observations that exceed 5 are discarded from the model: (a) Cellina di Nardò and (b) Leccino.....104

Fig. 3/14 - Global and local analyses of PCA relative to the Leccino variety.....105

Fig. 3/15 - Internal indices overlap the average spectral signature used to identify the excellent bands related to the Cellina di Nardò variety: (a) Calinski-Harabasz index, (b) Silhouette index.....107

Fig. 3/16 - Spectrum of Calinski-Harabasz index related to both varieties superimposed on the spectral signatures: (a) Cellina di Nardò and (b) Leccino.....108

Fig. 3/17 - Spectrum of average Silhouette index related to both varities superimposed on the spectral signatures: (a) Cellina di Nardò and (b) Leccino.....109

Fig. 3/18 - Canonical Discriminant Analysis applied to the bands selected by the two models: (1a) HAM on Cellina di Nardò, (2a) HAM on Leccino, (1b) iPCA-ICV on Cellina di Nardò, (2b) iPCA-ICV on Leccino.....111

LIST OF TABLES

Table 1/1 - Major Vegetation indices employees in Plant Disease Detection.....47

Table 3/1 - Statistics related to samples taken from the field and the related tests for the assessment of *Xf*.....93

Table 3/2 - Spectral signatures and related leaves tested for *Xf* in the sampling period ..95

Table 3/3 - Confusion matrix derived from GDA-LOOCV applied to the bands selected through the (i) HAM (LLR2, PCA, Wilks' Lambda) and (ii) iPCA-ICV.....110

Table 3/4 - Cross-validation error rates (%) calculated through a confusion matrix derived from GDA-LOOCV applied to the bands selected through the (i) HAM (LLR2, PCA, Wilks' Lambda) and (ii) iPCA-ICV. The number of bands used 110

Table 3/5 - Wilks' Lambda calculated through Canonical Discriminant Analysis applied to the bands selected through the (i) HAM (LLR2, PCA, Wilks' Lambda) and (ii) iPCA-ICV. The number of bands used 112

LIST OF EQUATIONS

Equation 1/1 - Reflectance. 39

Equation 1/2 - Transmittance. 39

Equation 1/3 - Absorbance. 39

Equation 3/1 - Savitzky Golay polynom. 81

Equation 3/2 - Wilks' Lambda parameter. 86

ABSTRACT English

The occurrence of the Olive Quick Decline Syndrome (OQDS) caused by *Xylella fastidiosa* (*Xf*) in Apulia region (Italy), with the strain Co.Di.RO (**Co**mplesso del **Di**ssecamento **R**apido dell'**O**livo) affecting mainly the olive trees, poses a serious threat for olive production in all Mediterranean countries.

Xf is a regulated pathogen in Europe (EPPO A1 list) because it affects more than 350 plant species worldwide. Infected olive trees may die as a consequence of the multiplication of the bacterium inside the vascular system which restricts the water flow from the roots to the canopy of the tree. Around 95% of olive cultivation is concentrated in the Mediterranean region and Italy ranks second worldwide.

Accordingly*, Xf* represents the main threat of olive trees worldwide due to the severe symptoms induced (mainly leaf scorch, dieback and quick decline of the tree), the long list of sap-feeding insects which may efficiently spread the pathogen, as the *Philaenous spumarius* in Apulia, and the large number of secondary hosts.

Xf restricts the cultivation of olive trees and the preservation of the historical heritage of olive trees in the Mediterranean region. Currently, no control measures are fully effective in the control of the bacterium and in the management of the olive quick decline; therefore the early detection of infected trees, their immediate eradication and vector control strategies are the only means of avoiding or containing the risk of contamination. These measures could be more effective if the infection is identified at early stages of disease development, in order to mitigate the spread of the pathogen and infections to neighbouring trees. However, visual inspections in the field are time-consuming and expensive.

To this aim, remote sensing could be a useful tool to detect water stress induced by *Xf* infection in olive trees at early stages.

Recently, an increase in research occurred in the application of Geomatic techniques, due to a greater availability of Remote and Proximal Sensing (RS, PS) instruments which has led to significant progress in the monitoring of complex biological phenomena and relative data management for running in, stand-alone, or web-based Geographic Information System (GIS) platforms. In this way it is possible to integrate heterogeneous spatial data in a single operative environment. Such data can be obtained by means of direct methods or indirect methods. The resulting data can be used for the implementation of provisional models to identify a plant adversity in order to rationalize the intervention strategy.

The first research of this work, the suitability of photointerpretation techniques to recognize and classify the plants damaged by OQDS in GIS environment was evaluated, for this purpose very high geometrical resolution aerial images were used by processing visible (VIS) and near infrared (NIR) data on a study area in South of Apulia region, which represents the first outbreak area of *Xf*.

The remotely acquired radiometric measurements were aimed at identifying appropriate photo-types, morphologically suitable in detecting the alteration of olive trees associated to different levels of OQDS-like symptoms. The use of spatially defined images strengthened by the presence of the near infrared band has greatly facilitated the identification of signs of OQDS starting with key photo types which are well correlated to the expression of the disease.

The technique made it possible to identify 20% of the photo interpreted OQDS-trees and infected by *Xf*. This achievement is the prerequisite to thoroughly examine and improve the methodology through the use of stereoscopic restitution in the GIS environment.

However, a second research was aimed at assessing the potential of hyperspectral reflectance data (HR) to identify the infection of *Xf* in olive at early stages of development.

Sampling was carried out on infected plants belonging to the two main olive varieties varieties (cvs. “Cellina di Nardò” and “Leccino”) grown in a commercial grove located in the outbreak area of *Xf* in south Apulia. Each sample was made of leaves collected

from 10 twigs/tree with different levels of infection. The study focused on the: (i) the discrimination between infected asymptomatic and non infected leaves; (ii) the selection of the best wavelengths for highlighting this discrimination and (iii) the identification of biophysiological indicators (vegetation indices) correlated to the OQDS induced by *Xf*.

The discrimination of infected leaves has been made using pre-elaborated data acquired with a field spectroradiometer, in the spectral wavelengths range between 400 and 1830 nm. A heuristic approach to variable selection, used in literature (*Lambda-Lambda* R^2 model - LLR^2, *Principal Component Analisys* model - PCA and *Wilks' Lambda*) and a combined general purpose detection method, proposed in this research, named interval PCA Internal Clustering Validation, iPCA-ICV have been compared.

The unsupervised method proposed, divides the spectrum of reflectance data into a determined number of intervals, calculates the PCA within them (iPCA) and validates the goodness of the groupings obtained (classes) through *Cluster Validity index* measurement. The discriminative ability of selected wavelengths by the two methods was assessed by generalized discriminant analysis based on canonical correlation and measurement of error type such as leave-one-out cross-validation, through confusion matrices.

From both methods it was possible to discriminate leaves infected by *Xf* and to select specific narrowbands. However, the best discriminative power was obtained from iPCA-ICV for both varieties (error rates of 23.7% and of 22.02% respectively for cv. Cellina di Nardò and cv. Leccino), compared to the reference method (error rates equivalent to 42.47% and 22.02% respectively for cv. Cellina di Nardò and cv. Leccino).

The two methods have shown differences in number and in the position in the narrowbands selected (each of 10 nm) between the two varieties. In particular, both agree with the VIS regions (close to the blue and the red) and that of Short Wave Infrared (SWIR) as portions of the spectrum increase the discrimination of Leccino, the variety less affected by the infection (23.1%), while, for Cellina, the species more affected (85.7% of positive findings). The iPCA-ICV identifies the absorption bands of water around 1180 and 1400 nm (and many bands of SWIR). The heuristic method identifies two bands of 705 and 805 nm, as determinants in the identification of Xylella.

The identification of critical regions of the spectrum, therefore, is the first logical step towards the development of indicators of robust stress based on hyperspectral images. The band selection techniques, also, are extremely useful not only to improve the

power of predictive models, but also for the interpretation of the data or design of specific sensors for Pest Disease Detection (PDD).

key words: *Xylella fastidiosa*, Hyperspectral proximal sensing, Stress detection, Feature election, Principal Component Analysis, Cluster Validity index.

ABSTRACT Italiano

La presenza in Puglia (Italia) del **Co**mplesso del **Di**sseccamento **R**apido dell'**O**livo (CoDiRO) causato da *Xylella fastidiosa (Xf)*, il cui ceppo Co.Di.RO colpisce prevalentemente gli alberi di olivo, rappresenta una seria minaccia per la produzione olivicola in tutti i Paesi mediterranei.

Xf è un patogeno regolamentato in Europa come organismo di quarantena (lista EPPO A1) perché colpisce più di 350 specie vegetali in tutto il mondo. La maggior parte degli olivi infetti muore a seguito della moltiplicazione del batterio all'interno del sistema vascolare che limita il flusso dell'acqua dalle radici alla chioma dell'albero. Circa il 95% della coltivazione olivicola è concentrata nella regione mediterranea e l'Italia è il secondo Paese produttore a livello mondiale. Quindi, *Xf* rappresenta una seria minaccia per l'olivo nel mondo a causa della gravità dei sintomi indotti (soprattutto bruscatura delle foglie, disseccamento dei rami e deperimento rapido dell'albero), della lunga lista di vettori che possono diffondere efficientemente il patogeno, come il *Philaenous spumarius* in Puglia, e l'elevato numero di ospiti secondari del patogeno.

Xf rappresenta un limite per la coltivazione dell'olivo e per la tutela del patrimonio storico olivicolo nella regione mediterranea. Ad oggi, non esistono misure efficaci di controllo e di lotta diretta al batterio e al CoDiRO; quindi, l'identificazione precoce degli alberi infetti, la loro immediata eradicazione e le strategie di controllo dei vettori sono gli unici mezzi per impedire o limitare il rischio di contaminazione. Tali misure potrebbero essere più efficaci se l'identificazione dell'infezione avvenisse nei primi stadi di sviluppo della malattia, in modo da poter contenere la diffusione del patogeno e la sua trasmissione agli alberi circostanti. Comunque, i rilievi visivi in campo richiedono tempo e sono costosi.

A questo scopo, il telerilevamento potrebbe essere uno strumento utile all'identificazione di stress idrici causati dai primi stadi dell'infezione di *Xf* negli alberi di olivo.

In tempi recenti si è assistito ad un aumento della ricerca nelle applicazioni delle tecniche Geomatiche, favorito dalla maggiore disponibilità di strumenti di rilevazione da remoto e da vicino, che ha condotto ad un significativo avanzamento della possibilità di monitorare fenomeni biologici complessi e di gestire, in ambiente Geographic Information System (GIS), i relativi dati sia in modalità stand-alone che in rete. In tal modo è possibile integrare, in un unico ambiente operativo, dati spaziali eterogenei derivanti dall'impiego di metodi diretti, come le azioni di monitoraggio, o dall'utilizzo di metodi indiretti, come l'elaborazione dei dati telerilevati. I dati così prodotti possono essere utilizzati per l'implementazione di modelli previsionali nella difesa delle avversità sul territorio e per potere così adattare la strategia di intervento e razionalizzare la difesa delle colture.

La prima ricerca in questo lavoro ha come obiettivo la valutazione dell'idoneità delle tecniche di fotointerpretazione per riconoscere e classificare piante colpite dal CoDiRO in ambiente GIS. A tal fine sono state utilizzate immagini da aereo ad alta risoluzione geometrica nel visibile e nell'infrarosso vicino relative ad un'area di studio nel sud della Regione Puglia, che rappresenta la prima area focolaio di *Xf*.

Le misure radiometriche rilevate da remoto sono state orientate all'individuazione di appropriati fototipi, morfologicamente in grado di rilevare l'alterazione associata a diversi livelli di sintomi ascrivibili al CoDiRO. L'uso di immagini spaziali definite, rafforzato dalla presenza della banda nel vicino infrarosso, ha facilitato notevolmente l'identificazione dei segnali di CoDiRO a partire dai fototipi chiave che sono ben correlati all'espressione della malattia.

La tecnica ha reso possibile l'identificazione del 20% di alberi fotointerpretati con CoDiRO ed infetti da *Xf.* Questo risultato rappresenta un buon presupposto per poter esaminare in modo approfondito e migliorare la metodologia attraverso la restituzione stereoscopica in ambiente GIS.

La seconda ricerca è stata invece finalizzata all'accertamento del potenziale dei dati di riflettanza iperspettrale (HR) per poter identificare l'infezione di *Xf* nei primi stadi di sviluppo su olivo.

I campionamenti hanno riguardato piante infette delle due principali varietà di olivo (cvs “Cellina di Nardò” e “Leccino”) coltivate in un campo commerciale localizzato nell’area focolaio di *Xf* nel Sud della Puglia. Ogni campione era composto da foglie raccolte da 10 rametti/albero con diversi livelli di infezione. Lo studio ha avuto come obiettivo la: (i) discriminazione tra foglie infette asintomatiche e foglie non infette; (ii) la selezione delle migliori bande per evidenziare tale discriminazione e il (iii) confrontato tra due metodi di selezione delle variabili a sostegno dell'analisi delle riflettanze iperspettrali.

La discriminazione delle foglie infette asintomatiche da quelle non infette, utilizzando dati pre-elaborati acquisiti con uno spettroradiometro da campo, è stata definita nell’intervallo di lunghezze d’onda 400 - 1830 nm dello spettro. Un approccio euristico di selezione delle variabili, utilizzato in letteratura (*Lambda-Lambda* R^2 model - LLR^2, *Principal Component Analisys* model - PCA e *Wilks' Lambda*) e un combined general purpose detection method, proposto in questa ricerca, denominato interval PCA Internal Clustering Validation (iPCA-ICV) sono stati messi a confronto.

Il metodo non supervisionato proposto, divide lo spettro dei dati di riflettanza in un numero determinato di intervalli, calcola la PCA all'interno di essi (iPCA) e convalida la bontà dei raggruppamenti ottenuti (classi) attraverso misure di *Cluster Validity index*. La capacità discriminante delle lunghezze d'onda selezionate dai due metodi è stata valutata mediante analisi discriminante generalizzata basata sulla correlazione canonica e sulla misura dell'errore di tipo leave-one-out cross-validation, attraverso matrici di confusione.

Da entrambi i metodi è stato possibile discriminare foglie infette da *Xylella fastidiosa* e selezionare bande strette specifiche. Tuttavia, il miglior potere discriminante è stato ottenuto da iPCA-ICV per entrambe le varietà (percentuale di errore del 23.7% e del 22.02% rispettivamente per cv. Cellina di Nardò e cv. Leccino), rispetto al metodo di riferimento (percentuale di errore del 42.47% e del 22.02% rispettivamente per cv. Cellina di Nardò e cv. Leccino).

I due metodi hanno evidenziato differenze nel numero e nella posizione delle bande strette selezionate (ciascuna di 10 nm) tra le due varietà. In particolare, entrambi concordano con le regioni del VIS (vicini al blu e al rosso) e dello Short Wave Infrared (SWIR) come porzioni dello spettro a maggior peso sulla discriminazione della Leccino, varietà meno colpita dall'infezione (23.1%), mentre, per la Cellina, varietà più colpita (85.7% di positività riscontrata), i due metodi risultano discordanti. Il iPCA-ICV individua le

bande di assorbimento dell'acqua intorno a 1180, 1400 nm e in molte bande dello SWIR, il metodo euristico individua due bande a 705 e 805 nm, come determinanti nell'individuazione di Xylella.

L'identificazione di regioni critiche dello spettro, dunque, costituisce il primo passo logico verso lo sviluppo di indicatori di stress robusti basati su immagini iperspettrali. Le tecniche di selezione delle bande, inoltre, risultano estremamente utili non solo per migliorare il potere dei modelli predittivi, ma anche per l'interpretazione dei dati o il design di sensori specifici Pest Desease Detection (PDD).

key words: *Xylella fastidiosa*, telerilevamento prossimale iperspettrale, rilevazione di forme di stress, Feature selection, Analisi delle Componenti Principali, indici per la validità dei raggruppamenti.

CHAPTER 1: INTRODUCTION

1.1 Contest and justification

Agriculture currently faces the great challenge of guaranteeing future food security to the increasing worldwide population (Fereres et al., 2011). A sustainable rural economy is highly dependant on healthy plants.

Unfortunately, plant pests (pathogens and insects), including invasive alien species, are having unprecedented impacts in yield and quality losses, which could be estimated between 20 and 40% of global agricultural productivity (Oerke, 2006). However, impacts can be wide-ranging and include not only direct economic impacts (crop yields, employment, tourism, etc.), but also environmental impacts (loss of habitats and biodiversity) and social/anthropological impacts (religious and spiritual values, educational values and cultural diversity and heritage).

The phytosanitary threat is significantly enhanced due to the worldwide increase in global trade and the movement of people, which are responsible for human-mediated introductions of invasive pests and diseases over long distance, and to climate changes (increase of temperatures, decrease of rainfall, thus increasing droughts), which favour the establishment of infectious agents in regions where they have not been previously reported, or emerge from indigenous vegetation to invade plant species.

Control means against most of invasive pests are difficult or may not even exist. Preventive measures based on the early pest surveillance and detection are economically and environmentally more efficient than eliminating phytosanitary outbreaks from which the infestation/infection can fast spread, even by vectors as for some severe pests.

Therefore, the early detection of plant diseases in the field (before the onset of disease symptoms) could be a valuable source of information for the application of specific measures to prevent the development and the spread of plant pathogens (Mahlein et al., 2012). Current techniques are not sufficiently advanced to manage large numbers of samples and to early detect all relevant pests; moreover, National Plant Protection Organizations (NPPO) and phytosanitary stakeholders need an efficient support for the decision they have to take at territorial and farm scales (Steiner et al., 2008).

Furthermore, diseases as well as abiotic stress conditions are commonly heterogeneous in time and space in a production field. Thus, site-specific disease management has to be assessed by detailed recording of spatial distribution and disease development, requiring large-scale and geo-referenced monitoring of diseases in the crop for precise timing and application of control measures (Nutter et al., 2010). Consequently, precise and time saving methods are essential for early disease surveillance and detection.

Therefore, in the last decade remote sensing methods and modelling supported by high statistic approaches have made progress in order to provide useful tools to detect disease symptoms at early stages of development and spatial heterogeneities due to pathogens at canopy scale (Bock et al., 2010, Nutter et al., 2010, Mahlein et al., 2012).

The early detection of within-field differences in crop status or growth conditions caused by diseases would enable the farmer to streamline input factors thereby optimizing his profit margin, while simultaneously improving the overall stability of the agrosystem.

1.2. The Olive tree

Olive (*Oleaeuropea* L.) is the most cultivated non-tropical fruit tree in the world, with 97% of world production located in the Mediterranean Basin (Figure 1/1) and the remaining 3% spread among other geographical areas (López-Escudero and Mercado-Blanco, 2011). Olive cultivation occupies 10 million hectares, which include an estimated 865 million trees (IOC, 2013). Olive is therefore the traditional and characteristic tree in the Mediterranean region, but it represents the symbol of peace worldwide, too.

Olives have characteristics beneficial for human health related to their unique composition of fatty acids and to numerous minor components. Olive trees have also been a powerful tool in the protection of the environment and are still important in sustainability of agriculture in extreme or near extreme environments. Especially for the European Union, the production of extra virgin and virgin olive oil (EVOO and VOO) is one of the most economically significant, as it generates and exports more of these products than any other single country or trading bloc.

Italy ranks second in Europe after Spain, with a total producing area of 1.110.701 hectares (ISTAT, 2016). Conversely, Italian olive oil production is only one third of the Spanishproduction due to the high costs for olive cultivation and harvesting. Therefore, the olive oil sector in Italy depends onolive oil importations, primarily from Tunisia (ISMEA, 2016).

Most of the olive farms are located in the South of Italy and the olive production in Apulia is 30% of the national production. In this region, the olive tree is not only a crop of economical importance but characterises the landscape and the heritage of the region with great impact from asocial and environmental point of view. Indeed, Apulia ranks first in Italy for the number of centennial olive trees, about 40% of the national heritage,which are protected by regional law for their monumental value (Law 144 of 14/02/1951).

The olive tree is therefore grown in the whole Apulian territory which has different pedoclimatic characteristics; therefore a number of varieties and different agricultural practices are adopted, such as traditional cropping in marginal areas to intensive plantations in irrigated areas (De Gennaro and Roselli, 2013). Main cultivar is “Coratina” with 90.000 hectares of cultivation, followed by “Ogliarola Salentina”, “Cellina di Nardò” and “Ogliarola Barese”.

The year 2015 has been considered the worst year for olive oil, due to a considerable decrease in production and quality of olives not only in Apulia, but in the whole of Europe. According to UNAPROL (Unione Nazionale tra le Associazioni di PRoduttori di OLive), this situation was mainly caused by strong attacks of the olive fly (*Bactrocera oleae* Rossi) and damages mainly caused by the diseases of anthracnose (*Colletotrichum* spp.) and Vertifcillium wilt (*Verticillium dahliae*). Unfortunately, the Apulian olive trees are now threatenedby a newly introduced bacterium, *Xylella fastidiosa*, which has already

caused the death of thousands of trees,since its first finding in 2013, and for which no control means are available (Saponari et al., 2013).

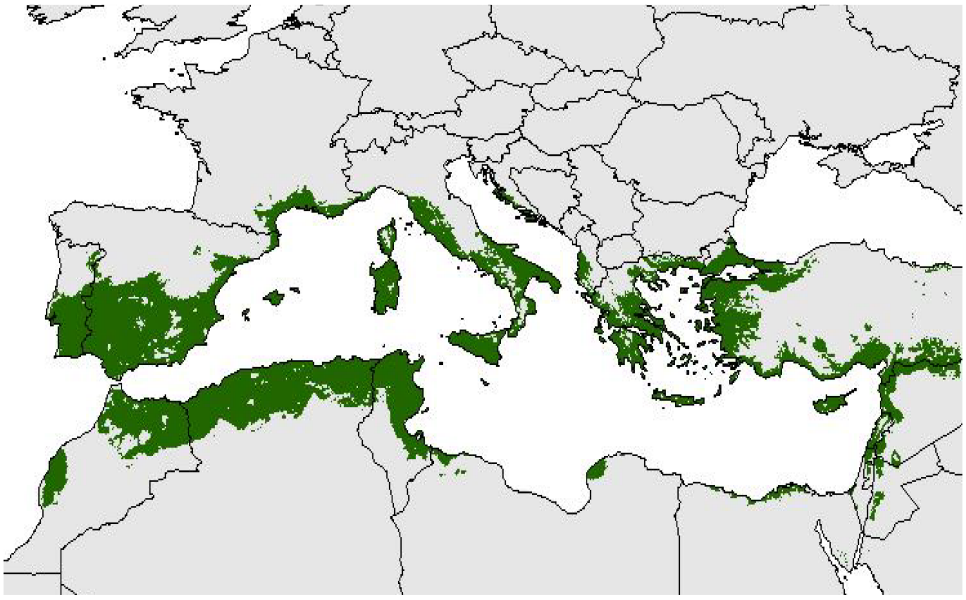


Fig. 1/1 - Olive-producing countries in the Mediterranean region.

1.3 Xylella fastidiosa and the Olive Quick Decline Syndrome in Apulia region

Xylella fastidiosa (*Xf*) (Wells et al., 1987), a negative gram and xylem-limited bacterium of the family *Xanthomonadaceae*, is a destructive pathogen belonging to the EPPO A1 list of quarantine organisms. *Xf* is subdivided in four subspecies (*fastidiosa*, *multiplex*, *pauca* and *sandyi*) which have a different geographic distribution in the American continent, although some *Xf* subspecies have been recently identifiedin Italy, France, Iran, Germany and Spain (Figure1/2). This pathogen represents a serious threat worldwide due to the large number of hosts (more than 359 according to EFSA, 2016), the severe symptoms induced (mainly leaf scorch, wilting, dieback and quick decline) and the long list of sap-feeding insects vectors (*Homoptera*, *Cicadellidae* and *Cercopidae*) (Almeida et al., 2005).

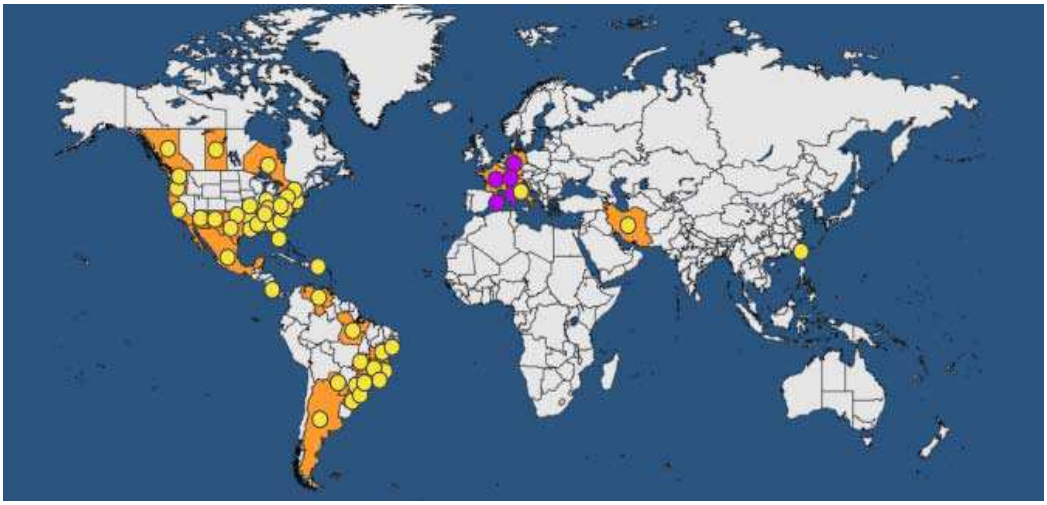


Fig. 1/2 - World distribution of *Xylella fastidiosa* (EPPO, 2016).

The bacterium is persistently transmitted by adults in which they multiply at the foregut level. Both nymph and adult vectors soak up the bacteria by feeding on an infected plant xylem, but only winged adults, because of their high mobility, can inoculate the bacteria in healthy plants immediately after acquisition.

Indeed, *Xf* is associated with economically important diseases, such as Pierce's disease of grapevine, Plum leaf scald, Phony disease, Almond leaf scorch, Citrus variegated chlorosis and Coffee leaf scorch (Janse and Obradovic, 2010). However, some host plants remain symptomless and can serve as a source of inoculum for vectors (Hopkins and Purcell, 2002).

Colonization of the xylem vessels results in their clogging by the bacterial biofilm which impairs water uptake (Newman et al., 2003). The discovery of *Xf* in Italy (Saponari et al., 2013) has made more evident the risk that this pathogen represents for European and Mediterranean commercial crops, landscape trees and ornamentals.

The Italian *Xf* subspecies pauca strain CoDiRO detected in olive trees and other host species was the same Strain Typing to theone infecting oleander and coffee in Costa Rica (Nunney et al., 2014). In Europe, the green leafhopper *Cicadellaviridis* L. (Cicadellidae) and the meadow spittlebug *Philaenus spumarius* L. (Aphrophoridae) were reported

as potential vectors for *Xf* (Janse and Obradovic, 2010). However only *P. spumarius* was proven to be an effective vector in Apulia region of Italy (Saponari et al., 2014).

The Olive Quick Decline Syndrome induced by *Xf* is initially characterized by the heterogeneous desiccation of twigs and small branches. The majority of local cultivars is highly susceptible and dies within a few years. Experimental evidence shows that grapevines and citrus are not infected by the CoDiRO strain. The most severe symptoms are displayed on centenarian olive trees infected by *Xf*, in which tracheomicotic fungus species are more considered as aggravators of the OQDS.

Observations of preliminary investigations suggest that, under the same natural conditions, old and new Italian olive cultivars expressed different levels of symptom severity to the infection: the old cultivars (e.g. Ogliarola Salentina, Cellina di Nardò), largely grown in Apulia, showed more susceptibility, regardless the age of orchard, compared to new varieties (e.g. Leccino, Coratina, Nociara) that show less symptoms under same growing conditions (Boscia et al., 2014).

Fig.1/3 - The Quick Decline of Olive trees in the outbreak area of *Xylella fastidiosa*.

After this finding, several reports on *Xylella* in olive occurred in Argentina (Haelterman et al., 2015) and in Brazil (Della Coletta-Filho et al., 2016).

Being a quarantine pathogen, *Xf* is regulated in the European Union by Directives 2000/29/CE, followed by the Decision 2015/789/CE. Drastic control methods were soon adopted in the outbreak areas of Apulia for pathogen containment (e.g. mandatory uprooting of the infected trees and the nearby hosts; mandatory control of the vector *P. spumarius*) and in the *Xf*-free areas for preventing pathogen introduction. As shown in Figure 1/4 a demarcated area (infected zone and buffer zone) was officially established in Apulia region for the application of different phytosanitary measures. The emergency posed by *Xf* prompted the Italian Ministry of Agriculture to appoint a special Commissioner for its control and to establish a national advisory scientific committee. More than 80.000 trees were monitored and new foci have been found outside the initial outbreak area of Lecce province (Brindisi and Taranto provinces), showing the rapid spreading of the infection. Nurseries are submitted to regulatory restrictions concerning the production and movement of *Xf*-host plants. Within this contest the early pathogen surveillance and detection is of utmost importance for supporting the National Plant Protection Organization (NPPO) in the rapid application of containment measures.

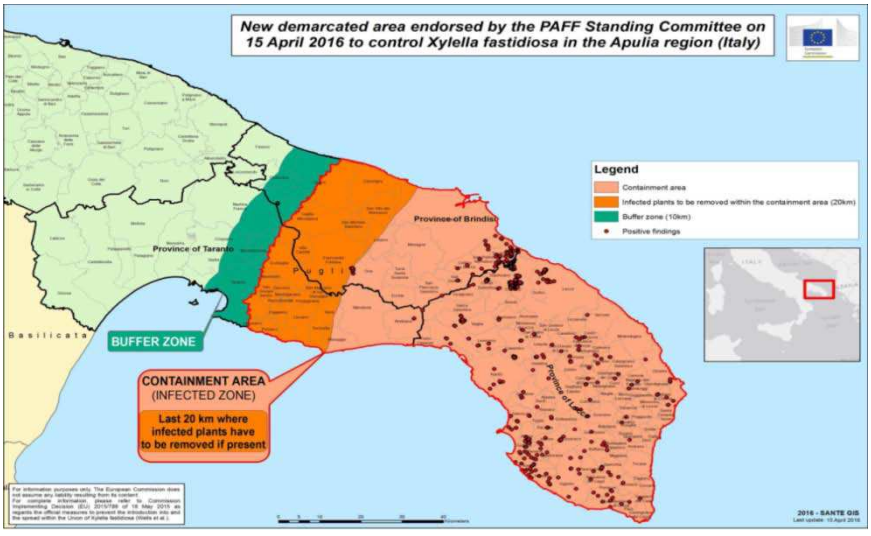


Fig. 1/4 - The official demarcated area for *Xylella fastidiosa* in Apulia Region (EU Decision 2015/789).

1.4 Remote sensing of plants

With the rise of space technology, Remote Sensing, a technique that allows the collection of information from a remote acquisition system without coming into contact with the object "to be observed", has, in the last fifty years permitted the obtaining of various and important information regarding the characteristics of the land and ocean surfaces over large areas across the planet (Jones and Vaughan, 2010).

These techniques have been made possible according to the physical principle that different materials reflect, absorb and transmit light radiation in a different way, according to the different wavelength of electromagnetic energy that intercepts them. In essence, different materials may be characterized by a specific spectral signature that uniquely characterizes them.

Regarding the different types of sensors, those that dominate major ground-based observations, operate in the "optical region" of the electromagnetic spectrum (VIS, NIR and SWIR), especially for applications that involve studies on vegetation, where the main absorption diagnostic characteristics basically fall in this region of the spectrum (Ustin et al., 2009, Kokaly et al., 2009).

The first sensors used for remote sensing, acquired the reflected energy of the light coming from the objects hit (target) in a few wavelength ranges (*broadband or multispectral sensor*).

Since the early 1980s, instead, it was possible to develop hyperspectral sensors, that increase the availability of the number of information channels (reflectance) (passing from 3 - 10 bands to 100 - 1000 bands, approximately), increasing the spectral resolution through the reduction of the bandwidth (*Full Width Half Maximum* - FWHM), from over the 100 nm in multispectral sensors to approximately 1-10 nm in new technology hyperspectrals (Hunt, 1980).

The hyperspectral data have led to improved estimates of biochemical indicators and their morphological and structural characteristics correlated with the vegetation compared to those obtained from traditional broadband multispectral data (Lee et al., 2004, Zhao et al., 2007).

In particular, if starting from the traditional multispectral data, containing information limited to a few broad bands of the ElectroMagnetic(EM) spectrum (typically the red

and the NIR such as Normalized Differential Vegetation Infex - NDVI - for example), a single index was made to correlate with almost all the features of the vegetation.

Through the use of hyperspectral data (hundreds of narrowbands) connections unique to the characteristics of the vegetation can be established (Hyperspectral Narrow Band Vegetation Index) as an example, hyperspectral indices associated with water content of the plant, soil moisture, structure of the biomass, plant pigments (chlorophyll, anthocyanins etc.).

Hyperspectral vegetation indices have therefore improved the accuracy in modelling and in vegetation mapping up to thirty percent more compared to the use of broadband indices (Thenkabail et al., 2013, Bolton and Friedl, 2013).

The use of hyperspectral data finds its applicationin the recovery of biochemical parameters related to the nutritional status of the plant, such as nitrogen (Ramoelo et al., 2013), phosphorus (Mutanga and Skidmore, 2004), lignin/cellulose (Daughtry et al., 2004), water content (Mirzaie et al., 2014), to biochemical pigments such as carotenoids (Blackburn, 2007), anthocyanins and in particular cholorophyll (Huang and Blackburn, 2011, Navarro-Cerrillo et al., 2014).

Biophysical/structural indicators starting from hyperspectral data provide instead, information regarding the vegetation fractional coverage of the plant (Guerschman et al., 2009, Somers et al., 2009), biomass content per unit area (Casas et al., 2014), the leaf area index (Darvishzadeh, 2008), specific leaf area (Van Wittenberghe et al., 2014), the diameter of the trunk and the average height of trees (Cho et al., 2009), etc..

The increasing availability of hyperspectral data therefore, acquired from aerial and satellite platforms more frequently, has stimulated and interested worldwide research towards the development of new methodologies for the recovery of the vegetation parameters (Lee et al., 2004).

1.4.1 Aerial photography and photogrammetry

Photogrammetry applies the principles of optics and knowledge of the interior geometry of the camera and its orientation to reconstruct dimensions and positions of objects represented within photographs (Campbell and Wynne, 2011).

The photogrammetric images are created by a sensor that transforms the electromagnetic radiation flux into electrical impulses subsequently converted in numeric format. The resulting image (Red, Green, Blue, NIR) is formed by basic elements, generally of square shape, called pixels (picture elements). Each pixel has orthogonal Cartesian coordinates, which define the spatial location within the image, and a numerical value called Digital Number (DN) that defines the brightness.

One of the most valuable regions of the spectrum is the NIR region, characterized by wavelengths just longer than the longest region of the visible spectrum.

This region carries important information about vegetation and is not subject to atmospheric scattering. The colour infrared model (CIR) creates a three-band colour image by discarding the blue band from the visible spectrum and adding a channel in the NIR.

It shows living vegetation and water bodies very clearly and greatly reduces atmospheric effects compared with the natural-colour model, so it is very useful for high-altitude aerial photography, which otherwise is subject to atmospheric effects that degrade the image.

This band combination is important for studies in agriculture, forestry, and water resources, to list only a few of many (Campbell and Wynne, 2011).

An early link between crop stress and a remote measurement was made in 1956 when Colwell showed that infrared photography could be used to record the changes in internal leaf reflectance that were caused by disease (Colwell, 1956).

His work, as well as that on potato late blight by Manzer and Cooper (1967) in which they displayed the possibility to detect the disease on photographs 2-3 days before the symptoms became evident on ground surveys, has inspired many researchers in agriculture and forestry

Since then a number of researchers have demonstrated the utility of colour infrared photography in detecting diseases in the field.

Photointerpretation is based on three steps: image interpretation process, elements of image interpretation, image interpretation keys.

Image Interpretation process

For purposes of clarification, it is important to distinguish between these separate functions:

Classification is the assignment of objects, features, or areas to classes based on their appearance on the imagery.

Enumeration is the task of listing or counting discrete items visible on an image.

Measurement, or mensuration, is an important function in many image interpretation problems. Two kinds of measurement are important.

First is the measurement of distance and height and, by extension, of volumes and areas as well. The practice of making such measurements forms the subject of photogrammetry, which applies knowledge of image geometry to the derivation of accurate distances.

A second form of measurement is quantitative assessment of image brightness. The science of photometry is devoted to measurement of the intensity of light and includes estimation of scene brightness by examination of image tone, using special instruments known as densitometers. If the measured radiation extends outside the visible spectrum, the term radiometry applies.

Delineation is the identification of edges or boundaries between different areas, in order to separate distinct areal units that are characterized by specific tones and textures (Campbell and Wynne, 2011).

Elements of Image Interpretation

The combination of the eight elements of image interpretation describes characteristics of objects and features as they appear on remotely sensed images:

- *image tone* denotes the lightness or darkness of a region within an image;
- *image texture* refers to the apparent roughness or smoothness of an image region;
- *shadow* is an especially important clue in the interpretation of objects;

- *pattern* refers to the arrangement of individual objects into distinctive recurring forms that facilitate their recognition on aerial imagery;
- *association* specifies the occurrence of certain objects or features, usually without the strict spatial arrangement implied by pattern;
- *shapes of features* are obvious clues to their identities;
- *size* can be the relative size of an object in relation to other objects on the image, or can be the absolute measurements (Campbell and Wynne, 2011).

Image Interpretation Keys

Image interpretation keys are valuable aids for summarizing complex information portrayed as images. They have been widely used for image interpretation. An image interpretation key is simply reference material designed to permit rapid and accurate identification of objects or features represented on aerial images.

A key usually consists of two parts: (1) a collection of annotated or captioned images or stereograms and (2) a graphic or word description, possibly including sketches or diagrams (Campbell and Wynne, 2011).

1.4.2 Use GIS environment to create the plants disease database

Advances in computing technology has led to form an automated cartography that played the main role in the development of Geographical Information System (GIS).

The GIS tool can help in mapping strategies and in the application of plant protection measures with a high precision manner. GIS can overcome the heterogeneity issue related to pests and abiotic factors, in time and space. This tool has a strong influence on spatial analysis. The increasing ability to capture and handle geographic data means that the spatial analysis is occurring within data-rich environment, where GIS provides platform for managing geographic data.

The main advantage of the photointerpretation in GIS environment, from aerial/satellite images with high spatial resolution, is to have a continuous view of the study area (large territorial scale), the integration with other geographic data and especially the ability to create easily updated geo-referenced databases.

1.5 Proximal and remote sensing of plants

As a technique without physical contact, RS, by definition, includes the use of spectral measurements derived from handheld instruments at small distances (less than a few meters), such as PS. These measurements are calculated and analyzed to retrieve information from the object observed (in this instance, the diseased trees) (De Jong and Van der Meer, 2007).

RS and PS are indirect evaluation techniques, capable of monitoring the characteristics of vegetation and its state of health at various distances and over vast territorial areas.

The sensors used in applications for the study of the vegetation can be distinguished as “active” or “passive” depending on the artificial radiation emitted, by directly measuring reflected or scattered energy of the object that is hit (active sensors), or by directly measuring reflected solar radiation or the thermal radiation emitted (passive sensors). Radar and Lidar are examples of active RS and PS instruments.

Passive or optic instruments are capable of measuring solar radiation in the visible (VIS, wavelength 400-700nm), near infrared (NIR, wavelength 700-1.100nm), in shortwave infrared (SWIR, 1.100- 2.500nm) and the energy emitted by thermal infrared (TIR, 3 to 15 micron) of the electromagnetic spectrum for applications related to Plant Disease Detection.

From the point of view of the optical properties of a plant system, the path that the light accomplishes when it enters into contact with a leaf is characterized by the following processes:

- transmission through the surface;
- absorption by internal chemical substances (example, pigments, water, sugars, lignin and amino acids);
- partial reflection by internal structures or total reflection from the contact surface (Figure 1/5A).

It is therefore clear that the process of reflection of light in plants is a complex phenomenon and is associated with numerous biophysical and biochemical interactions.

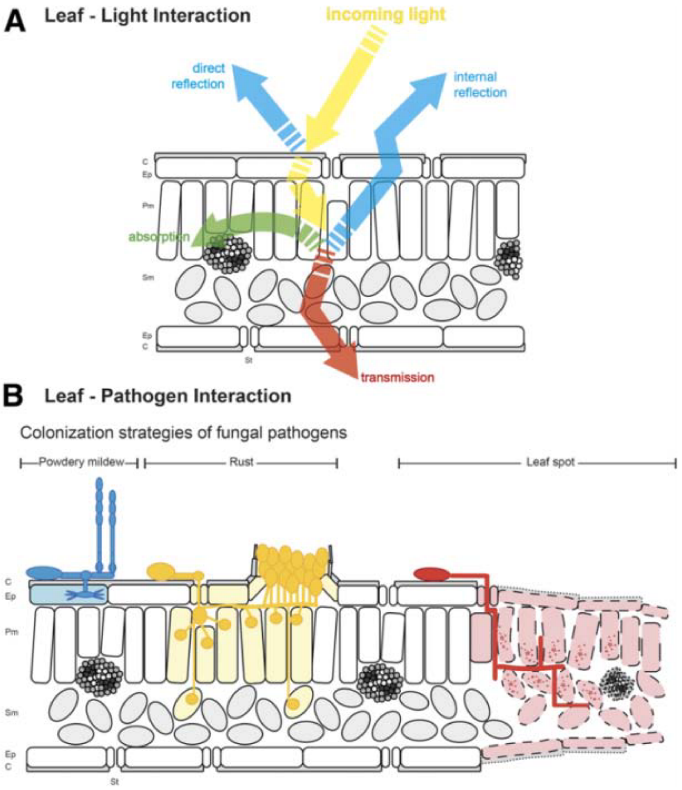


Fig. 1/5 - A: The interaction of leaf tissue with light depends on structural and leaf chemical properties. B: During pathogenesis, plant pathogens influence leaf structural and chemical properties, and by this process the leaf optics are altered (Mahlein, 2015).

In terms of wavelengths, the VIS range is principally dominated by the contents of the absorption of pigments present in the leaves; the NIR range depends on the structure, processes of internal scattering and the absorption of water and the SWIR range is influenced by the composition of chemical constituents and water (Jacquemoud and Ustin, 2001).

The nature and quantity of reflection, absorption and transmission, depend on the wavelength (λ) of electromagnetic radiation, angle of incidence (causing both specular or diffuse scattering), surface roughness (leaf surface cuticle) and especially, the differences

in the structures of leaves, biochemical constituents of optical characteristics, dielectric, or thermal, of vegetation elements (Kumar et al., 2001, Baret et al., 2007).

These properties are quantified by the dimensionless physical magnitudes of reflectance (R), transmittance (T) and absorbance (A) and defined by the following mathematical equations:

$$R(\lambda) = E(\lambda)_R / E(\lambda)_I \times 100$$

Equation 1/1 - Reflectance.

$$T(\lambda) = E(\lambda)_T / E(\lambda)_I \times 100$$

Equation 1/2 - Transmittance.

$$A(\lambda) = E(\lambda)_A / E(\lambda)_I \times 100$$

Equation 1/3 - Absorbance.

where $E(\lambda)_R$, $E(\lambda)_T$ and $E(\lambda)_A$, respectively, express reflected energy, transmitted and absorbed by the plant system (e.g. leaves), conforming to wavelength and $E(\lambda)_I$ the incident energy from the rising sun.

The magnitude of spectral response, reflectance R, is generally obtained from multispectral or hyperspectral sensors displayed in the functions of the spectral resolution. The resulting curve which shows the values of reflectance in the varied wavelengths is defined as the spectral signature and is typical of the kind of vegetation which was measured.

Figure 1/6 shows the typical spectral signature of vegetation in the range of wavelengths between 350 and 2500 nm (Pu and Gong, 2011).

A plant in biotic stress conditions (e.g. affected by bacteriosis) reacts with a protection mechanism which manifests itself in impairments in the leaf structure, and specific changes in the chemical composition of the tissue (Figure 1/4B) which generally lead to suboptimal growth. All of this is expressed by a spectral signature different from that of a healthy plant.

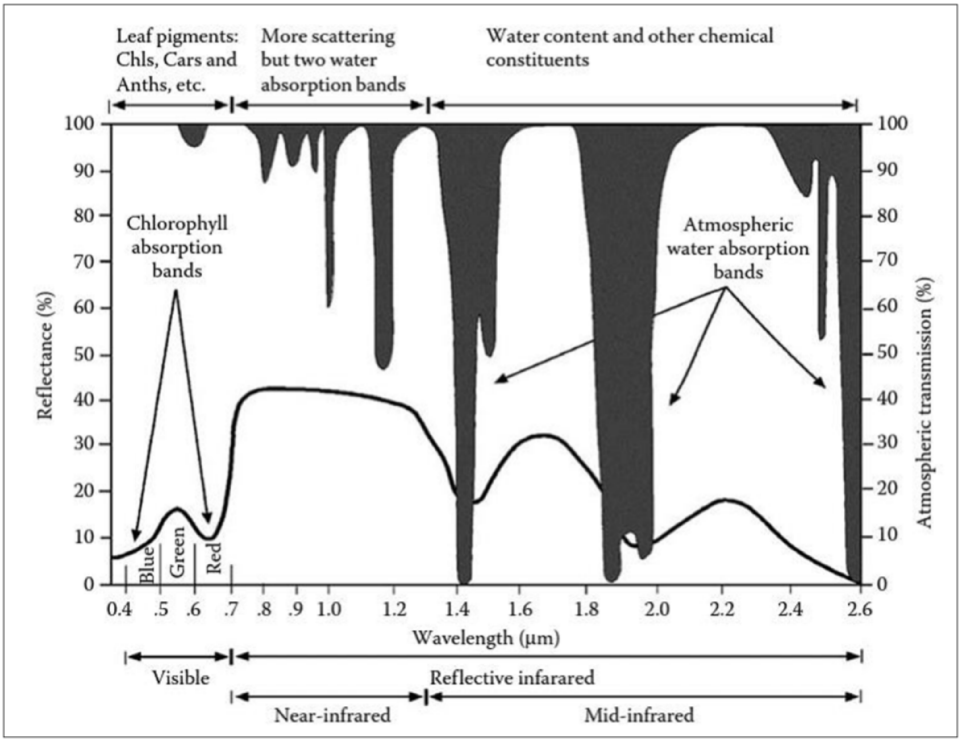


Fig. 1/6 - Typical spectral signature of vegetation by Pu and Gong (2011).

Meroni et al. (2010) have provided a clear and exhaustive description of the effects of stressors on a plant's physiology and of how RS can detect early or chronic changes induced by them.

The total energy absorbed by a plant (Absorbed Photosynthetically Active Radiation - APAR) is related to the total leaf surface and the pigment concentration (chlorophyll) of the foliage. A healthy plant utilizes APAR primarily for vital photochemical reactions (0-20%) and dissipates the rest in the form of heat (75-90%) and fluorescence (2-5%) (Meroni et al., 2010).

Fluorescence is a radioactive process which can verify the relaxation of an excited molecule (dissipation of energy) (Figure 1/7).

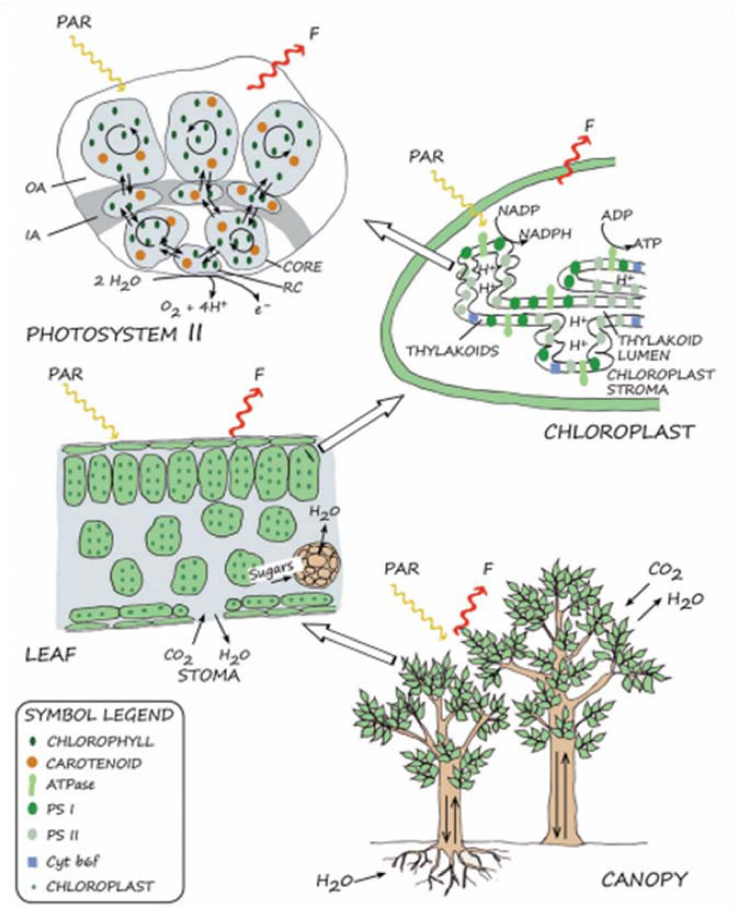


Fig. 1/7 - Chlorophyll a fluorescence to photosynthesis from the photosystem to the canopy level and to leaf level (Porcar-Castell et al., 2014).

When the plants are exposed to pathogens they activate defense responses, which in the initial stage (the absence of visual symptoms), reacts with physiological mechanisms such as the reduction in the rate of photosynthesis, which induces an increase of fluorescence (in the form of dissipation of light energy on the part of the chlorophyll) and heat emission (West, 2003).

In the next stage (visible expression of the symptom), the activity of the pathogens show a reduction in chlorophyll content in leaves, necrotic lesions or chlorotic

spread, which increase VIS reflectance, causing a displacement of the spectral signature in the position in the spectrum.

Specific spots and effects of browning and/or of senescence, furthermore, may manifest on the leaves (linked to the nature of the pathogen), and influence the areas of VIS and NIR due to a possible condition of water shortage.

At the foliage level, instead, the disease can cause thinning of the leaf surface density of the foliage which can be observed in the NIR area (Franke and Menz, 2007).

1.5.1 Hyperspectral RS of plant diseases

Spectroscopy

Spectroscopy is among one of the techniques most used in RS PDD, VIS and/or SWIR both imaging type or nonimaging. The regions of the visible and infrared of the electromagnetic spectrum are known to provide maximum information on the levels of physiological stress of plants (Muhammed, 2005) and, therefore, some of these wavelengths (specific of a disease) could be used for early detection (West, 2003), even before the symptoms are visible.

These techniques are considered promising for the monitoring of crop diseases. Below, some of the most recent and relevant spectroscopy-based techniques applied on different crops and measurement conditions are discussed.

Naidu et al. (2009) use spectral reflectant of the leaves to identify viral infections (in the field) in vine (*Vitisvinifera* L.) that cause grapevine leafroll disease, using a portable spectrometer to collect data from each leaf of the plant.

Yang et al. (2007) have studied brown planthoppers and leaf-folder infestations in rice plants. The conditions of infestation of plants were classified and further efforts have been conducted to identify the extent of the infestation, using spectroradiometric reflectance measurements (350-2400 nm) of data collected in field conditions.

A combined study on the differentiation of winter wheat diseases due to pathogens (powdery mildew and yellow rust) and insect infestation (wheat aphid) was recently carried out (Yuan et al., 2014).

The methods of early diagnosis of diseases are of particular interest (Malthus and Madeira, 1993, Rumpf et al., 2010), although their actual application in crop management was inconsistent between crops. The available studies have been conducted only on specific crops and therefore the results obtained cannot be generalized to other crops and/or locations.

Delalieux et al. (2007) used hyperspectral reflectance data (350-2500 nm) to detect apple scab caused by *Venturia inaequalis*. The study involved the identification of infected plants and selection of appropriate wavelengths to classify between healthy and infected leaves.

Wu et al. (2008) used hyperspectral refelctance data, in laboratory conditions,for the detection of fungal infections of eggplant infected by *Botrytis cinerea* before visible symptoms.

Ray et al. (2010) used spectral signatures of leaves acquired in the field to distinguish potato varieties, nitrogen levels, irrigation levels and the early detection of late blight disease by selecting optimal wavelengths.

Gualano et al. (2012) produced vegetation indices by the algebraic combinations of reflected or emitted energy values measured in the different bands of the electromagnetic spectrum. Measurements highlighted a difference in the spectral signatures of trees infected by *Citrus tristeza virus* (CTV) with respect to the CTV-negative trees; thus, specific indices were selected for the implementation of a detection algorithm, which was applied to a processed multispectral satellite image.

Alnaasan (2015) assessed the potential of the hyperspectral reflectance and a set of derived Spectral Vegetation Indices (SVIs), in the range of Visible-Near Infrared wavelength, to detect fire blight infections (*Erwinia amylovora* Burrill) on inoculated apple seedlings in glasshouse and on symptomatic trees in the field.

Imaging tools

Lately, hyperspectral imaging tools have been introduced for the evaluation of plant disease at leaf and canopy levels. Laboratory-based studies include for example Fusarium fungal infection and head blight disease in wheat (Bauriegel et al., 2011) and early detection of sugar beet diseases (Mahlein et al., 2012).

The possibility to differentiate infections in a certain stage of their development is of particular interest for the efficient operation of RS PDD (Mahlein et al., 2012).

In this regard, studies of yellow rust disease in the field of wheat plants (Bravo et al., 2003) tried to distinguish among wheat diseases and abiotic stress conditions (Moshou et al., 2004).

Reynolds et al. (2011) and Huang et al. (2007) used hyperspectral data in the field as well as in flight to evaluate the gravity of *Rhizoctonia* crown and root rot disease in sugarbeet and yellow rust in wheat, respectively. The hyperspectral data obtained in flight resulted in being better suited for RS PDD application both at farm and regional levels.

Zhang et al. (2003) used Airborne Visible Infrared Imaging Spectrometer (AVIRIS) data to identify the late blight disease on tomato crops, while Hillnhütter et al. (2011) studied soil pest-induced sugarbeet disease using two different airborne sensors (AISA and HyMap).

UAV (Unmanned Aerial Vehicle) technology represents a possible solution for the monitoring of plant health issues in the field, reducing the cost of data acquisition. However, even fewer scientific works have analysed the use of these platforms at a regional level for PDD applications. In that respect, Wheat streak mosaic was assessed using multispectral satellite data.

Wang et al. (2012) instead, analyze the multilevel capacity (from leaf-to-satellite observation) offered by RS to detect winter wheat stripe rust.

There is a possibility, still unexploited, to integrate the Spectroscopy based imaging technique (VIS and SWIR ranges) with fluorescence imaging. The application of this integrated method has so far been exploited to improve the detection of winter wheat yellow rust in the field and through use of UAV aerial platforms (Panigada et al., 2014).

Recently, the merger of airplane data from thermal sensors, fluorescence and hyperspectral optics has been successfully experimented on olive trees to evaluate those infected by Verticillum wilt (Calderón et al., 2013).

Statistical methods

One of the limitations of hyperspectral optical sensors is the vast amount and complexity of data gathered.

In order to effectively utilize the data from the optical sensors (imaging and non-imaging-based) for the identification and diagnosis of the disease, advanced methods of statistical analyses are essential. In general, tasks such as the following are required from RS data:

- *early identification* of a disease;
- *differentiation* between different diseases;
- *separation* between biotic and abiotic stress;
- *quantification* of the seriousness of the disease.

These requirements must be achieved with a higher accuracy or be equivalent to the requirements achieved from methods of traditional evaluation and with shorter response times.

Hyperspectral RS data analysis methods are continually introduced in the science of plant diseases and may be classified in four large-scale groups which include:

a. univariate techniques of correlation, regression and analysis of variance (ANOVA) of the presence of disease and/or its seriousness with the spectral response, in specific or subinterval narrowbands of the wavelengths of the spectrum (Huang et al., 2012);

b. univariate techniques of evaluation and identification of specific spectral indices of the vegetation by narrowband (SVIs) sensitive to the presence of the disease (Reynolds et al., 2011, Mahlein et al., 2013, Stilwell et al., 2013);

c. multivariate techniques of Data Mining applied to processing of spectral data for the reduction of dimensionality and extraction/selection of features (Grisham et al., 2010, Bauriegel et al., 2011);

d. techniques of Machine Learning and of Classification, parametric and non-parametric, supervised and unsupervised for cluster identification (classes of homogeneous groups) based on the presence/absence of the disease and/or of relative levels of seriousness (Moshou et al., 2012, Mahlein et al., 2012).

In this context, recent scientific researches have identified sensors and methods of *Data Mining* for the survey/collection of data, identification and the quantification of diseases in plants (Sankaran et al., 2010).

If the sensors are able to capture subtle optical properties of plants in different regions of the electromagnetic spectrum (also beyond the range of the visible), thereby allowing for the detection of early changes in plant physiology induced by forms of biotic and/or abiotic stress (colour of the tissue, shape the leaf, rate of transpiration, morphology and density of foliage, photosynthetic changes etc.), the possibility of better analysing the enormous wealth of information contained in the RS data is relied upon for the appropriate choice of the techniques of the processing statistics used.

For example, Chen et al. (2008), utilizing a portable spectroradiometer in field conditions, made use of a linear model of correlation between the levels of severity (LS) and the spectrum of the foliage to precisely identify the presence of Verticillium wilt in diseased cotton plants. The authors reported that the first derivative of the reflectance of the infrared, in the wavelength range comprised between 731 and 1317 nm, was more effective in predicting the presence of Verticillium wilt.

In addition to the statistical classification models, many nonimaging spectroscopy-based studies utilize vegetative indices (univariati methods) to evaluate the variance of spectral reflectance in various dichotomous conditions (healthy or diseased plants).

Some of the vegetative indices used for the diagnosis of plant diseases are reported in Table 1/1.

Table 1/1 - Major Vegetation indices employees in Plant Disease Detection.

Vegetation Index	***Equation***	***Reference***
Photochemical Reflectance Index (PRI)	$(R531 - R570) / (R531 + R570)$	Gamon et al. (1992)
Normalized Difference Vegetation Index (NDVI)	$(R800 - R670) / (R800 + R670)$	Rouse Jr et al. (1974)
Modified Chlorophyll (a and b) Absorption in Reflectance Index (MCARI)	$[(R701 - R670) -0.29 * (R701 - R550)] * (R701 / R670)$	Daughtry et al. (2000)
Chlorophyll Normalized Difference Index (ChlNDI)	$(R750 - R705) / (R750 + R705)$	Gitelson and Merzlyak (1994)
Water Index (WI)	$(R900) / (R970)$	Pen¯Uelas et al. (1995)
Plant Senescence Reflectance Index (PSRI)	$(R680 - R500) / R750$	Merzlyak et al. (1999)
Structure Insensitive Pigment Index (SIPI)	$(R800 - R445) / (R800 + R680)$	Pen¯Uelas et al. (1995)
Red Edge Position (REP)	$R700 + 40*[((R670 - R780) / 2) - R700] / (R740 - R700)$	Curran et al. (1995)
Yellow Index (YI)	$- (R580 + R668 - 2*(R624)) / (0.044\text{^}2)$	Adams et al. (1999)
Optimal Vegetation Index (Viopt)	$(1.45) \times (R8002 + 1)/(R670 + 0.45)$	Reyniers et al. (2006)
Pigment Specific Normalized Difference (PSNDa)	$(R_{800\ nm} - R_{680\ nm})/(R_{800\ nm} + R_{680\ nm})$	Blackburn (1998)
Pigment Specific Normalized Difference (PSNDb)	$(R_{800\ nm} - R_{635\ nm})/(R_{800\ nm} + R_{635\ nm})$	
Pigment Specific Normalized Difference (PSNDcar)	$PSNDcar = (R_{800\ nm} - R_{470\ nm}) /(R_{800\ nm} + R_{470\ nm})$	
Modified Simple Ratio (mSR)	$\left(\frac{R_{800\ nm}}{R_{670\ nm}} - 1\right) / \sqrt{\left(\frac{R_{800\ nm}}{R_{670\ nm}} + 1\right)}$	Chen (1996)
Anthocyanin Reflectance Index (ARI)	$(1/550nm) - (1/700nm)$	Gitelson et al. (2001)
Blue/Green Index (BIG2)	$R_{450\ nm} / R_{550\ nm}$	Zarco-Tejada et al. (2005)
Double Difference Index (DD)	$(R_{749\ nm} - R_{720\ nm}) - (R_{701\ nm} + R_{672\ nm})$	le Maire et al. (2004)

Normalized Phaeophytinization Index (NPQI)	$\frac{R_{415\ nm} - R_{435\ nm}}{R_{415\ nm} + R_{435\ nm}}$	Barnes et al. (1992)
Modified Chlorophyll Absorption Integral (mCAI)	$\frac{(R_{545\ nm} - R_{752\ nm})}{2}(752 - 545) - \left(\sum_{R_{545}}^{R_{752}} R \times 1.423\right)$	Laudien et al. (2003)
Red Edge Vegetation Stress Index (RVSI)	$\frac{R_{714\ nm} + R_{752\ nm}}{2 - R_{733\ nm}}$	Naidu et al. (2009)
Index-SPAD	$K \log_{10}\left[\frac{R_{940\ nm}}{R_{650\ nm}}\right]$	Unpublished
Lichtenthaler Indices (Lic)	$R_{440\ nm}/R_{690\ nm}$	Lichtenthaler et al. (1996)

Imaging techniques are an improvement compared to spectroscopic techniques. (Nonimaging spectroscopy approaches). The data from this type of sensors involves larger surfaces and provides information in the form of three-dimensional images. In hyperspectral imaging, the reflectance of each pixel is acquired for each range of wavelength of the electromagnetic spectrum, thereby generating large “volumes” of spectral data.

This type of data, therefore, requires statistical methods for the analysis of the images such as PCA, Spectral Angle Mapper (SAM) classification and Support Vector Machine (SVM) classification that improve the accuracy of detection of the disease.

Moshou et al. (2004) utilized a spectrograph to acquire spectral images from 460 to 900 nm to detect yellow rust in wheat. Statistical techniques such as Quadratic Discriminant Analysis (QDA), Self-Organizing Maps (SOM), and multilayer perceptrons (MLP) and based Artificial Neural Networks (ANN) were used to classify the diseased from healthy wheat plants. Information from four wavebands, namely 543, 630, 750 and 861 nm were used for the classification models. The classification accuracies based on QDA and neural network (MLP) for the discriminating of individual healthy plants were 92.0% and 98.9%, and diseased plants were 97.8% and 99.4%, respectively.

Li et al. (2012) studied the applicability of aerial hyperspectral imaging to detect greening of canopy in citrus plants. The analyses carried out made it possible to classify images through the Spectral Angle Mapping and Spectral Feature Fitting (SFF) techniques, although they did not provide high accuracy.

1.5.3 From univariati methods to multivariati statistics

In the context of RS applied to the vegetation, the simplest statistical model that is considered is fundamentally based on empirical relationships (correlation) between the spectral information available (bands or a transformation of same) and the properties of the vegetation under observation (univariate regression analysis).

The spectral characteristics extracted from RS data are mainly synthesized through vegetation indices, developed and calculated as a mathematical combination of several spectral bands (at least 2), linked through a mathematical structure ranging from simple or difference ratio, in the form of normalized difference which is slightly more complex (Jones and Vaughan, 2010).

With the advent of hyperspectral RS a great variety of indices of vegetation have been defined with narrowband (SVIs) sensitive to different biochemical and biophysical parameters (Pu and Gong, 2011).

Vegetation indices have always been supported by the scientific community because they are deemed beneficial. In fact, given the mathematical structure that defines them, they minimize the variability of the spectral reflectance (normalization) induced by external and internal factors such as lighting differences of the scene, reflectance of the soil, atmospheric dispersion, leaf angle distribution and foliage structure in relation to Sensor Acquisition Geometry.

Specific chlorophyll indices (PSNDa, PSNDb) and caretenoids (PSNDcar) have been developed in order to estimate the concentration of these pigments in plants (Blackburn, 1998).

Photochemical Reflectance Index (PRI) has been found to be a good indicator of photosynthetic efficiency (Gamon et al., 1992). In particular, if the plant is healthy, then through a physiological process called the “xanthophyll cycle”, carotenoids quickly dissipate excess light energy that can accumulate in excessive lighting conditions, to protect the photosynthetic apparatus (Demmig-Adams and Adams, 1996).

In the past few years many studies have focused on the ability of SVIs to measure specific diseases (Mahlein et al., 2013) and many studies have been conducted to improve the performance related to sensitivity and linearity with the biochemical or biophysical quantity which are associated (Ustin et al., 2009).

However, despite the development and the proposed amendments to the mathematical form or of the optimal wavelength indices (e.g. the issue of Off-Absorption-Center Waveband, Majeke et al. (2008), there is currently, no broad consensus regarding the best Spectral Vegetation Index (SVI) that is able to globally predict the most common forms of biotic stress (Main et al., 2011).

The modifications, in general, do not translate into substantial improvements of the performance index, because even if key parts of the answer sought can be highligted, they tend to be sensitive to minor errors or noise in the spectral measurement mode (Rivera et al., 2014).

As a result of this uncertainty there has been an increasing application of multivariate statistical methods that take advantage of the full spectrum of hyperspectral data. Statistical techniques such as Stepwise Multiple Linear Regression (SMLR), PCA, Canonical Correlation Analysis (CCA) and Partial Least Squares Regression (PLSR), mentioned in part in the previous paragraph, are among the most popular methods of data mining employed in many works on plant diseases.

The complete exploration of all the wavelengths of the spectral signature often reveals its usefulness in improving the estimation of the parameter, and resolving the problem of Off-Absorption-Center Waveband (especially at canopy scale).

SVIs, relying on centers and absorption intervals (especially In-Chlorophyll Centre Waveband: 640 – 680 nm, for indices sensitive to chlorophyll), and being constructed on a few bands, weaken their performance (sensitivity to parameter), due to the structure of the foliage (effect of the propagation of the leaf signal and diffusion at canopy level) (Asner, 1998).

This effect occurs even when it is required to classify the different species on the basis of SVIs in a broader spectrum such as Normalized Difference Vegetation Index, NDVI (Majeke et al., 2008), or, finally, when there is overlapping absorption characteristics that share the same chemical bonds (Kumar et al., 2001). For example, the strong O-H bond present in plant tissue is a component of the water absorption characteristic, of cellulose, sugar, starch and lignin.

Therefore, the use of spectral vegetation indices (constructed on a few bands) may not be sufficient to represent a specific biophysical parameter or biochemical properties compared to the use of multiple bands (optimal spectral analysis) or even all bands

(full spectral analysis) that would require a multivariable analysis which is able to better represent the vegetation property (Darvishzadeh, 2008) and explain the various sources of variability of the spectra.

1.6 Necessity and objective of the research

Xylella fastidiosa (*Xf*), named the "bacterium killer of olive trees" by the media, is the most severe phytosanitary emergency that Italy, other European countries and the Mediterranean region are facing in the last few years. It is a recent invasive bacterium of quarantine importance which is killing thousands of olive trees in Apulia, South of Italy, inducing the Olive Quick Decline Syndrome mainly in ancient trees.

Considering the economic, landscape and religious importance of the olive tree in Italy as in the whole Mediterranean region and the lack of efficient control measures, the impact of this infection is inestimable. Preventive measures based on early pathogen surveillance (before the onset of disease symptoms) are the most valuable for supporting NPPO in applying immediate phytosanitary measures to avoid the development and the spread of this pathogen. The monitoring of *Xf* is very difficult due to the great number of host species, the efficiency of spread by insect vectors and the huge economic and human resources needed.

With the aim of promptly identifying the presence *Xf*, soon after its discovery, various emergency methods have been adopted in Apulia to counter the spread of the bacterium in the whole territory. A demarcation area with different interventions was defined by the EU Commission (EU Decision 2015/789) to separate the infected zone and the buffer zone from the pathogen-free areas. In the buffer and pathogen-free zones the monitoring is focused on early identification of olive trees showing OQDS symptoms.

With the aim of promptly ascertaining the presence of Xylella in OQDS or symptomless olive trees "mass" diagnostic tools (fast, reliable and low cost) were soon developed (Djelouah et al., 2014, YASEEN et al., 2015) with different sampling methods.

Unfortunately, current techniques are not sufficiently advanced to survey large areas and to manage large numbers of samples for early pathogen detection after its first establishment in South of Apulia.

In the last few years several studies were carried out on the use of geomatic techniques in the framework of pathogen monitoring programmes. These techniques allow for the integration in one operational environment of heterogeneous spatial data which are from direct methods such as “the monitoring actions” or from indirect methods such as “the elaboration of remote sensing data’. Acquired data can be used for the implementation of forecasting models for the phytosanitary surveillance of the territory.

On this premise, the overall goal of this research is aimed at using the remote sensing technology for supporting the official programme for the monitoring of *Xf* in Apulia in the rapid identification of infected olive trees, either showing OQDS-like symptoms either at early stages of symptom development.

To this aim the research is presented as two scientific works (Chapters 2 and 3): the first entitled “Assisted analysis of aerial images at high geometric resolution for the identification of Olive Quick Decline Syndrome associated with the *Xylella fastidiosa* bacterium in Apulia” and the second entitled “Determining Optimal Hyperspectral Wavebands for detection of *Xylella fastidiosa* using Reflectance data: a Internal Clustering Criteria approach”.

In the first work the specific objective has been addressed at experimenting with the effectiveness of the manual technique of photointerpretation in GIS environment for the classification of phytosanitary information (OQDS-suspected olive trees) starting from aerial images (RGB and NIR) at high geometric resolution related to the outbreak areas of *Xf*. This remote sensing approach which was the only available approach at the time of the first pathogen outbreak in Apulia could soon be adopted in the official pathogen monitoring programme because it is affordable.

The second work is focused on the analysis of data from high spectral content (hyperspectral reflectance) of asymptomatic leaves of infected plants, for early detection of infection in specific sections of the spectral signature, through the development of a multivariate method of spectral precision.

This approach will allow for the selection or definition of optimal narrowband sensors (multispectral and/or hyperspectral) useful for the early detection of the presence of infection from remote platforms (aircraft, UAV).

As a method of reference a multivariate heuristic method (based on LLR^2, PCA and Stepwise Discriminant Analysis (SDA) was selected from literature (Thenkabail et al., 2004), that combines the Feature Selection Methods (FSM) and Full-Spectral-Analysis (FSA) methods, capable of addressing the high multicollinearity and dimensional nature of the hyperspectral data, and correlate "optimal" narrowbands to biochemical characteristics.

The specific objectives are therefore aimed at:

1. developing a combined general purpose detection method, interval PCA and Internal Clustering Validation (iPCA-ICV), for selection of excellent narrowbands;
2. identifying narrow spectral bands of the spectrum, suitable for distinguishing *Xf* infection from the early stages in cultivars with different degrees of susceptibility.

The first objective will address the critical aspects of the management and processing of hyperspectral data, the problem of collinearity of this type of data, the lack of statistical methods applicable to different phytosanitary issues and not related to the specific disease and/or culture, overfitting data of some supervised methods and management of outliers always present large amounts of data. The above will be achieved through the development, in Matlab environment, of a general purpose detection method of hyperspectral data browsing with many "collinear" variables, unsupervised, able to search among the cluster spectral signatures (observations) and provide a measure of separation. By these methods spectral bands of variable width will be selected, starting from the reflectance spectra of leaves infected and not infected by *Xf*. It is a combined approach of multivariate Principal Component Analysis and Cluster Analysis (CA) techniques.

In the second specific objective, two statistical methods of variable selection will be compared in order to discriminate leaves infected by *Xf* and to select specific narrowbands of reflectance spectra of olive leaves. A heuristic approach to variable selection,

used in literature (LLR^2, PCA and SDA) will be implemented in Matlab environment and compared with the combined general purpose detection method, proposed in this research (iPCA-ICV).

The discriminative ability of the selected wavelengths of the two methods was assessed by Generalized Discriminant Analysis (GDA) based on canonical correlation and error measurement type Leave-One-Out Cross-Validation (LOOCV), through confusion matrices.

CHAPTER 2: Assisted analysis of aerial images at high geometric resolution for the identification of "Olive Quick Decline Syndrome" associated with the Xylella fastidiosa bacterium in Apulia

2.1 Introduction

The discovery of quarantine pathogen *Xf* in olive trees (*Olea europea* L.) in southern Apulia has raised strong concerns, which are unfamiliar and critical in plant health emergency management, and which is unique due to its peculiarity.

Xf is a gram negative bacterium which reproduces in the xylem vessels of certain plants and blocks the conducting system of raw sap, provoking a series of physiological alterations capable of causing even the death of infected plants (Saponari et al., 2013).

In fact, the symptomatology observed in olive trees in Apulia has also attracted the presence of other parasitic agents which have further aggravated the plant health framework (Nigro et al., 2013).

The damage observed and diagnosed was derived from collection of woody material, and from phloem and xylem with blockage of the lymphatic vessels of affected plants.

This symptomological framework suggests the definition and characterization of the disease that has affected the olive trees in Apulia as OQDS (Saponari et al., 2013), which has frequently been associated with the presence of *Xf* (Figure 2/1).

Fig. 2/1 - Clear cut symptoms of OQDS in olive tree.

At this moment in Apulia, the presence of this formidable bacterium is relegated only to the province of Lecce (groves in Gallipoli, Galatina, Lecce, Trepuzzi, Copertino, Alezio, Taviano, Nardò, Ugento, Cursi, Morciano di Leuca, etc.).

The characterization and phylogenetic analyses conducted on the bacterium have highlighted the genetic arrangement of the Apulian strain to which the classification of *Xf* subspecies *pauca* CoDiRO strain has been assigned (Cariddi et al., 2014).

The new CoDiRO strain identified in Salento is genetically distinct from strains already noted in the same subspecies that attack coffee and citrus in Brazil.

In particular, on the basis of research conducted to date by scientific institutions in Apulia, apart from olive trees, which constitute the predominate crops in these areas together with grapevine (an important source of economic viability and cultural and land-

scape heritage), other relevant host species (fruit and ornamental), prove to be infected by *Xf*.

As a matter of fact, *Xf* has been detected in the province of Lecce in several species ofalmond, cherry, oleander, *Vinca minor*, *Polygala myrtifolia*, *Westringia fruticosa*, *Catharanthus roseus* etc.. Other possible plants that are host to the bacterium can be added to the list (which is still in the stage of experimental verification) and can further burden plant health conditions in the province of Lecce, which is already heavily compromised.

In actions to be conducted for the containment of *Xf* it is essential to effect constant monitoring of the implicated area in order to identify the symptoms of OQDS in olive tree species and to accurately measure the spread of the infection to implement the actions necessary to combat it.

It is, therefore, considered appropriate to carry out research that would be able to examine the possibility of providing for the management of classic monitoring from ground level, and the support of aircraft based remote sensing with the aim of identifying the symptoms of OQDS in olive trees from an elevated level.

It is, therefore, considered appropriate to carry out research that would be able to provide technical support to the conventional monitoring from ground level with aircraft remote sensor equipped with the aim of identifying the symptoms of OQDS in olive trees remotely.

This current effort necessitated the verification of the effectiveness of visual photointerpretation (Jones and Vaughan, 2010) within the GIS framework, to classify plant health information, starting from aerial images of high geometric resolution (in the visible and near infrared closeups), related to study areas of outbreaks in Salento where the presence of *Xf* was ascertained.

In the context of the proposed work, the methodology taken into consideration is best suited for a solid application of the remote sensored area, which is generally the most utilized method in forestry and environmental fields (Malthus et al., 2000).

2.2 *Materials and Methods*

The aerial scenes of the areas affected by *Xf* in the Salento region, in which symptoms of OQDS were evident, were analysed. In particular, the study was conducted in the outbreak area of Trepuzzi on the borders of the city of Squinzano (which includes a small rural area) (Figure 2/2).

The area examined presented flat morphological characteristics with significant expanses of olive tree groves distinguishable from the surrounding landscape.

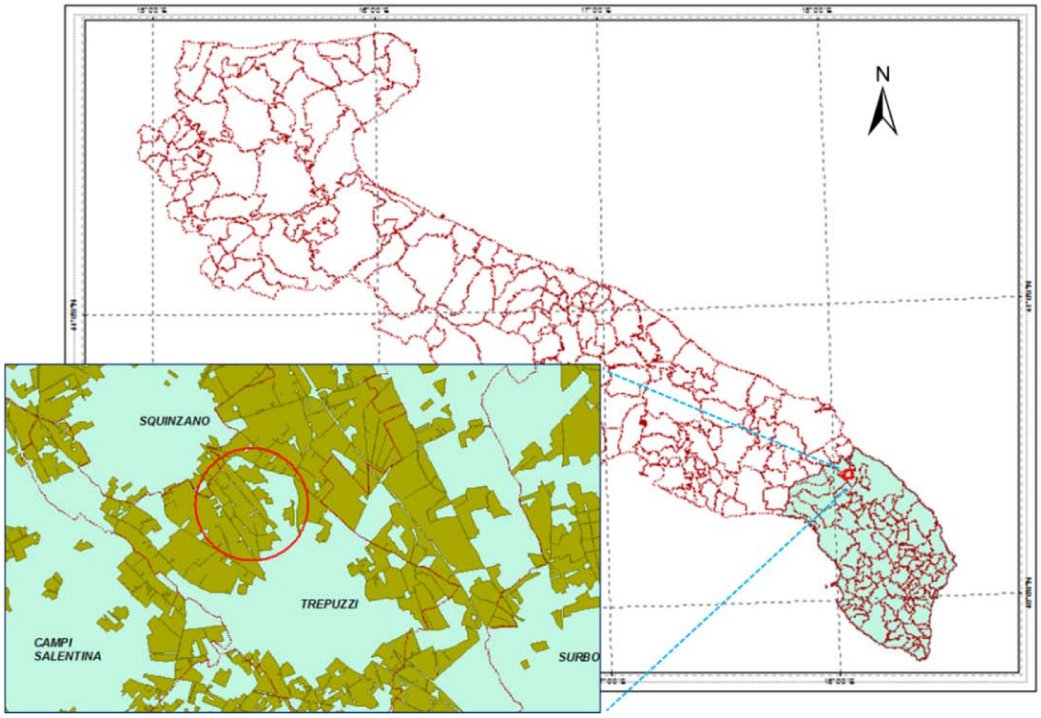


Fig. 2/2 - Subject of the study – the infected area circled in red. Olive groves are in green.

The raster data used to identify the symptoms of OQDS from elevated distance are composed of ten orthoimages in tiff format (uncompressed) resized to 10 cm from aerial photographs captured with digital camera DMC (Digital Mapping Camera of Inter-

graph Z/I), from an average height of about 750 metres above sea level (Ground Sample Distance from ground pixel of about 7 cm).

In particular, the aerial photographs in the visible and near infrared band (total implicated surface about 250 hectares) were acquired in March 2014 from the company SIT in Noci (BA) during a flight targeted at flying over certain areas affected bythe bacterium outbreak.

All images used were inspected and processed in the ArcMap 9.3 framework of Esri and generated the respective geospatial element points (shape files) of the OQDS plants identified (Figure 2/3).

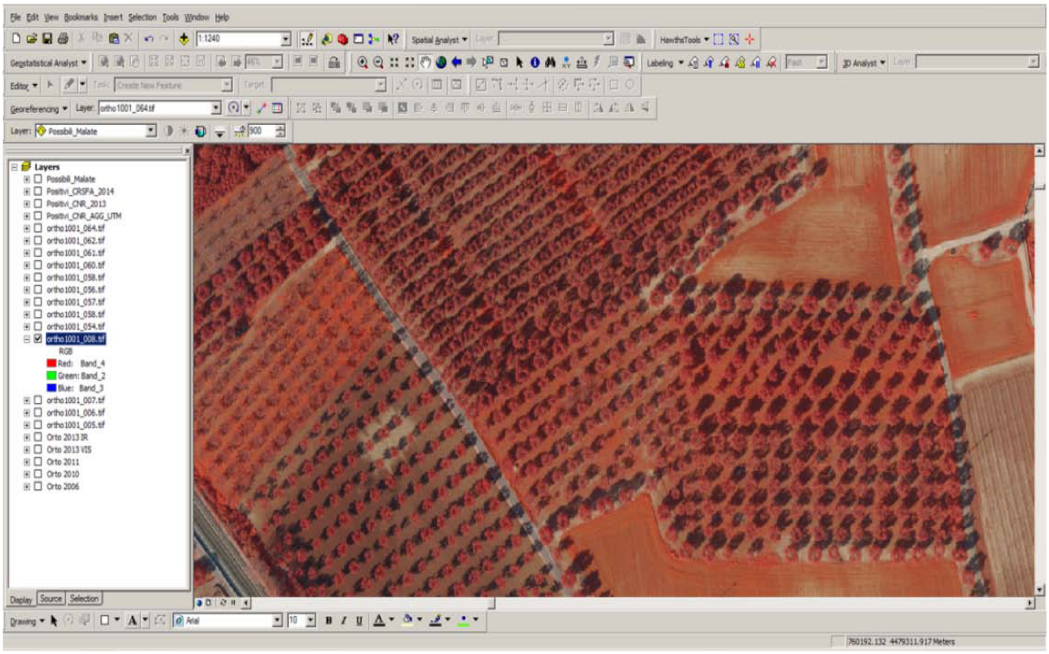


Fig. 2/3 - Images in false color, used in the photointerpretation phase manual, within the GIS framework.

To establish the presence of *Xf* in olive trees identified by means of the technique, serological (Enzyme-Linked ImmunoSorbent Assay, ELISA) and molecular (Polymerase Chain Reaction, PCR) (Loconsole et al., 2014) tests were carried out at the accredited laboratories of the CIHEAM - Mediterranean Agronomic Institute of Bari, in Va-

lenzano. In particular, the molecular test, which is more sensitive than the serological test, was applied to all samples that turned out to be serologically negative or dubious, to confirm the absence of the pathogen.

2.3 Results

In order to carry out the photointerpretation of the evidence of OQDS some key-sof photo reading were accessed and established on the orthoimages taken in March, on the basis of the gravity of OQDS symptoms. Among the typical symptomatologies primarily associated with the infection of *Xf* (Guario et al., 2013), the "leaf scorch", desiccation of shoots and branches and the reduced growth of the foliage must be taken into account (Figure 2/4).

Fig. 2/4 - Characteristic symptoms of OQDS on olive trees infected by *Xf*: a) "leaf scorch", b) desiccation of branches and c) Quick decline of the tree.

The above mentioned symptoms, in particular, become more evident as the infection spreads asymmetrically in the plants (even leading to death) and have been considered key objectives in this work.

In particular, with consideration to the pixel size from ground level (10 cm), a geodatabase was created to attribute the seriousness of OQDS symptoms at three levels (mild, medium and high) depending on the amount of drying observed in foliage.

Figure 2/5 shows the typical morphology of three photo types (in false colors), selected to be monitored for visual analysis.

The authors wish to draw attention to the more effective images in false colors to highlight the symptoms compared to the visible.

This is a well-founded justification for the minor influence of lighting conditions on image quality, which facilitates the recognition of symptoms, thanks to a greater contrast with the background (Spicciarelli and Arpaia, 1991).

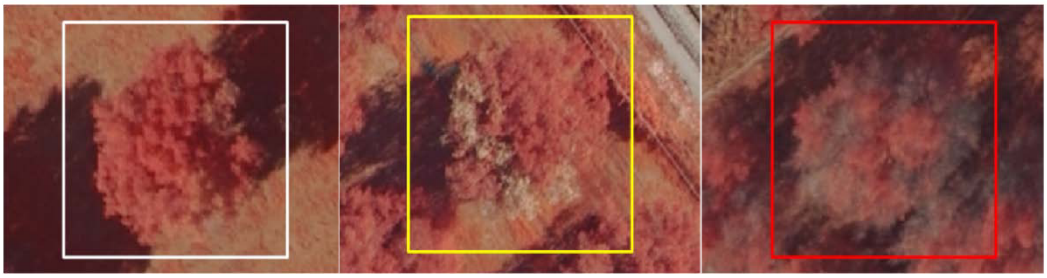


Fig. 2/5 - Key photo types selected for the three levels of gravity (white frame: mild; yellow frame: medium; red frame: high).

The edited results of the recognition process (in the infested area of Trepuzzi), obtained within the framework of ArcMap 9.3, has provided the opportunity to classify about 450 specimens of olive trees, and to specify the categories of gravity of OQDS (Figure 2/6).

Fig. 2/6 - OQDS classified olive trees.

In order to validate the technique, starting with all plants classified as OQDS in June 2014, 43 specimens were inspected in the field (catalogued TR1 – TR43). From the latter, 20 plants were selected as suitable samples from the group.

Laboratory results confirmed the presence of *Xf* in four samples of olive trees with symptoms ranging from mild to medium, thereby establishing an overall of 20% of infected plants and demonstrating the effectiveness of the method (Figure 2/7).

Furthermore, by comparison with the previous results of the official monitoring (supported by the Apulian Plant Protection Service), it was revealed that OQDS plants identified in the infected area, coincided, in part, with plants that tested positive for the bacterium (especially for the high gravity condition) as reported in the enlargement in Figure 2/7.

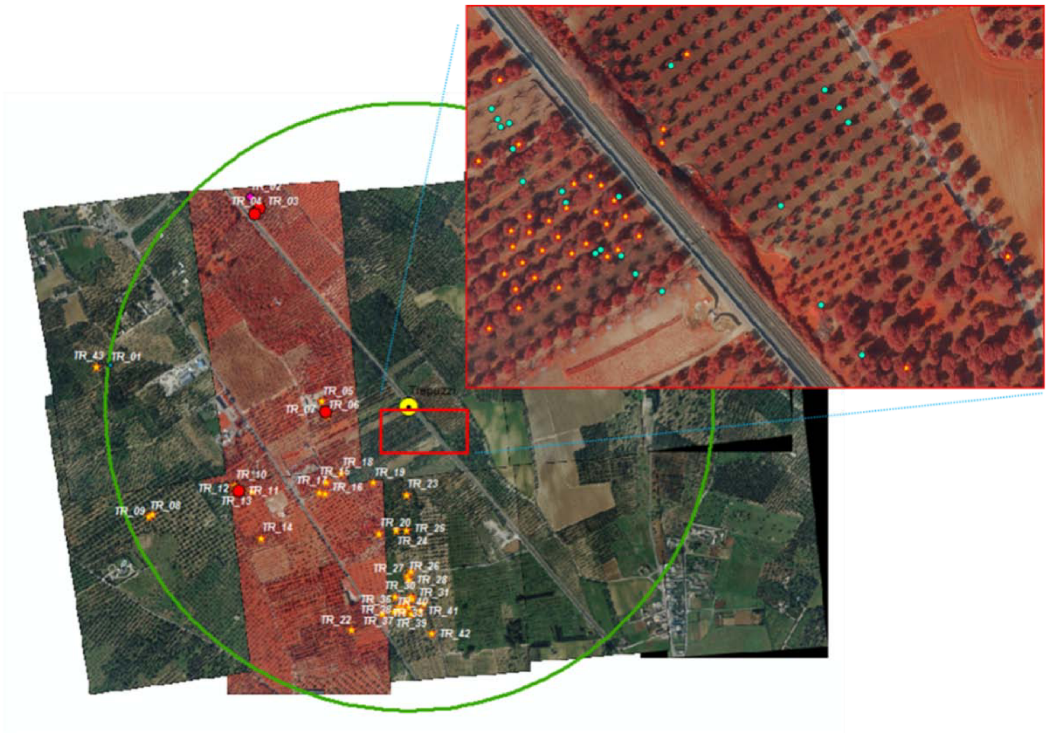


Fig. 2/7 - Photointerpretation of OQDS trees inspected in fields and which tested positive (circled in red) to serological and molecular testing in June 2014. In the enlarged area, trees that resulted positive from the official monitoring have been overlapped for contrasting effect (circled in cyan).

2.4 Conclusions

With this current work it was necessary to explore the potentiality of aerial remote sensing in the evaluation of the condition of the health of olive trees implicated with the problem of *Xf* experiencing the symptoms of OQDS with which it is highly associated.

In particular, the use of spatially defined images strengthened by the presence of the near infrared band in this manner has greatly facilitated the identification of signs of OQDS starting with key photo types which are well correlated to the expression of the disease.

Although further testing on larger areas should be conducted, the technique made it possible to reasonably achieve the objectives of the research (20% of the trees

identified and tested have shown infection of the bacterium) and establishes the premise to examine in depth and improve the methodology through the use of stereoscopic restitution in the GIS environment.

The perspective of the ground pixel resolution can also be improved with the use of Unmanned Aerial Vehicle (UAV). Research aimed in this direction will effectively allow for the better understanding of the symptoms of the disease at its initial stages (e.g. drying out of shoots), thereby allowing for an early intervention in areas of outbreak.

In contrast, the necessity to investigate vast areas of Apulia motivates the consideration of the use of aerial images of higher resolutions (or satellite images) and of "textbook" methods for processing derived data, capable of intercepting the symptoms of the disease timeously.

CHAPTER 3: Determining Optimal Hyperspectral Wavebands for detection of Xylella fastidiosa using Reflectance data: an Internal Clustering Criteria approach

3.1 Introduction

Xylella fastidiosa (Wells et al., 1987), is a negative gram bacterium, transmitted by xylem feeding insects (Janse and Obradovic, 2010) and provokes serious diseases in many species, including economically important crops, such as grapevine, citrus and almond.

In just California alone, Pierce's disease in vines, costs the wine grape industry 104 million euro per annum (Tumber et al., 2014). In the last few years, this pathogen has spread from its origin (the Americas), and has colonized new parts of the world. It has spread to Asia, where it has caused outbreaks of Pierce's disease and pear decline in Taiwan and in Iran (EPPO, 2013a); (Amanifar et al., 2014).

In Europe, the pathogen was first detected in Kosovo, as the causative agent of Pierce's disease of vine (Berisha et al., 1998), and then in 2013 in southern Italia (Apulia), through Olive Quick Decline Syndrome (OQDS), where it infected thousands of hectares of olive (EPPO, 2013b); (Nigro et al., 2013, Saponari et al., 2013, Martelli et al., 2016).

The resistance of the cultivars to the bacterium is an important aspect in the management strategies of any plant disease. This aspect has been well studied in the case of *Xf* in vine and citrus; however, it still remains to be ascertained in the case of olives.

Even though field observations, in some hotbeds in Apulia, suggest the possible resistance in some new Italian cultivars, (Boscia et al., 2014), further studies should be

conducted on the subject. Unfortunately, the pathogen agent is noted for its "uncertain presence" within the xylem vessels of plants that it colonizes (in addition to its wide range of hosts); this characteristic manifests both at plant spatial level (presence of mottled foliage) (Zhang, 2008) as well as at temporal level (differential bacterial count title of a few seasons) (EPPO, 2004b).

These aspects make the control of the bacterium even more difficult and for these reasons there is a strong need for methods of detection, and accurate and efficient monitoring.

Conventional methods for determining the presence of the pathogen are based on destructive laboratory measures related to the plant tissue sampled in the field (Loconsole et al., 2014).

A newly introduced detection method is Direct Tissue Blot Immuno Assay (DTBIA), suitable and validated in the case of *Xf* in Apulian vine. The application of this techique in situ has proved to be reliable in detecting infection (Djelouah et al., 2014), but moreover, it is useful because it provides an early diagnostic tool for monitoring infection at a large scale, without the need to handle infectious material with the risk of infection.

Even though other in situ diagnostic techniques (more sensitive to the bacterium) are available, (YASEEN et al., 2015), there remains the cost of essential human resources and of the implementation time of the monitoring plan on large scale, which is necessary for the detection and elimination of the source of infection.

The potential of hyperspectral RS PDD in the field of plant protection has been studied at length (Sankaran et al., 2010, Calderón et al., 2013, Mahlein et al., 2013, Calderón et al., 2015). The analysis and development of remote sensors for the measurement of spectral reflectance has created new opportunities for a rapid, non destructive, and relatively inexpensive approach for the estimation of biochemical and biophysical properties associated with vegetation stress.

Spectral reflectance is defined as the ratio of radiant flux reflex to incident solar reflex, when plants exposed to pathogens activate defense responses, which in the initial stages react with physiological mechanisms such as reduction in the rate of photosynthesis, which induces an increase of fluorescence and emission of heat (Erdle et al., 2011).

Specific spots and effects of browning and/or of senescence may further more manifest on the leaves (associated to the type of pathogen agent) affecting the VIS and

NIR areas due to a possible condition of water shortage. In a subsequent phase (visible expression of the symptom), the activity of the pathogens shows a reduction in the chlorophyll content of leaves, with widespread necrotic or chlorotic lesions that increase VIS reflectance causing a displacement of the position of the specral signature in the spectrum.

Of recent, various remote sensing systems have been introduced for the evaluation of diseases of the plants at leaf and foliage level. Laboratory-based studies include Fusarium fungal infection and head blight disease in wheat (Bauriegel et al., 2011), early detection of sugar beet diseases (Mahlein et al., 2012), and detection of Cercospora leaf spot, sugar beet rust, and powdery mildew on sugarbeet leaves (Mahlein et al., 2013).

Reynolds et al. (2011) and Huang et al. (2007) have used, both in fields and in flight, hyperspectral data to evaluate the gravity of *Rhizoctonia* crown and root rot disease in sugarbeet and yellow rust in wheat, respectively. The hyperspectral data obtained in flight tests are more suitable for RS PDD applications both at farm and regional levels.

Zhang et al. (2003) utilize AVIRIS data to identify mildew disease in tomato crops, while Hillnhütter et al. (2011) study soil pest-induced sugarbeet disease utilizing two different airborne sensors (AISA e HyMap).

Unmanned Aerial Vehicle technologies represent a possible solution for the monitoring of plant health issues in the field compared to conventional photo-reconnaissance aircraft (small dimensions, low mass, slow flight speed etc.), by reducing the cost of data acquisition and supplying images with very high spatial resolution in near real time (Ballesteros et al., 2014, Herwitz et al., 2004, Pelosi et al., Torres-Sánchez et al., 2014). However, even fewer scientific works have analyzed the use of these platforms at a regional level for PDD applications, especially in relation to the availabilty of efficient hyperspectral sensors.

Given the opportunity to conduct continuous low cost sampling, hyspectral RS has in the past few years focused on the possibilty of selecting “ad hoc”, the length of the narrow wave sensitive to specific physiological variations for early diagnosis of the disease (Malthus and Madeira, 1993, Laudien et al., 2003, Rumpf et al., 2010).

The available studies on the statistical models used, however, show inconsistent results, because they are conducted on specific crops, and therefore cannot be generalized with other crops and/or locations.

With reference to hyperspectral RS imaging techniques, the continuous spectrum of each pixel provides very narrow (<10 nm) and contiguous spectral band, compared to multispectral systems that acquire images in a few broad bands (> 50 nm) (Hansen and Schjoerring, 2003).

Narrowband hyperspectral sensors (integrated with thermal and fluorescence sensors), are therefore, able to provide a significant improvement on the ability of accurate estimates of biophysical characteristics, caused by biotic stress, compared to traditional broadband sensor (Shi et al., 2015, Calderón et al., 2015).

Hyperspectral data containing hundreds and even thousands of wavelengths, narrow and contiguous, (high-dimensional data), although containing richer information of that of the multispectral data, constitute a real challenge for their analysis and selection of the most suitable multivariate statistical models (Rinaldi et al., 2015).

Finding effective solution statistics becomes a precondition to fully exploit the information potential of hyperspectral data. The elimination (dropping) of redundant bands is a necessary and indispensable choice for the realization of optimal specialized sensors, that focus and collect data related to forms of stress (e.g. such as radiometers) on board UAV/air platforms for PDD applications.

The spectral characteristics extracted from RS data are primarily summarized through vegetation indices (commonly applied), developed and calculated as a mathematical combination of spectral bands, arranged through a mathematical structure that goes from the simple or difference ratio, to a slightly more complex normalized difference form (Jones and Vaughan, 2010).

Specific chlorophyll indices (PSNDa, PSNDb) and carotenoids (PSNDcar) have been developed in order to estimate the concentration of these pigments in plants (Blackburn, 1998). PRI has been found to be a good indicator of photosynthetic efficiency (Gamon et al., 1992).

Different studies in the last few years have focused on the ability of SVIs to detect specific diseases and many have been conducted to improve performances related to the sensitivity and linearity with biochemical or biophysical quantities that correlate (Ustin et al., 2009). However, despite the development and the proposed amendments of mathematical formsor of the optimal wavelength index (e.g. problem of Off-Absorption-Center Waveband, Majeke et al. (2008)), at present there is no broad consensus with regards to

the best SVIs that are able to globally predict the most common forms of biotic stress. Because of these concerns, there has been a growing application of multivariate statistical methods that takes advantage of the full spectrum of hyperspectral data.

The complete exploration of all the wavelengths of the spectral signature often reveals its usefulness in improving the estimate of the parameter, thereby resolving the problem of *Off-Absorption-Center Waveband* (especially at foliage level).

SVIs, in effect, relying on centers and intervals of absorption (e.g. In-Chlorophyll Centre Waveband: 640–680 nm, for indices sensitive to cholrophyll) and constructed on a few bands, weaken their performance (sensitivity to parameter), due to the structure of the foliage (effect of propagation of the signal leaf and spread of the canopy level) (Asner, 1998).

This effect occurs even when it is required to classify the different species on the basis of SVIs of a wider spectrum such as Normalized Difference Vegetation Index (NDVI) (Blackburn, 2007); (Majeke et al., 2008) or when there are overlapping absorbment characteristics that share the same chemical bonds (Kumar et al., 2001). For example, the strong O-H bond present in plant tissue is a component of the water absorption characteristic, of cellulose, sugar, starch, and lignin.

Therefore, the use of spectral vegetation indices (constructed on a few bands) may not be sufficient to represent a specific biophysical parameter or biochemical properties compared to the use of multiple bands (Optimal spectral analysis) or even all bands (Full spectral analysis) that would require a multivariable analysis which is able to better represent the vegetation property (Darvishzadeh et al., 2008) and explain the various sources of variability of the spectra.

To reduce the dimension of hyperspectral data and eliminate redundancy, hyperspectral data mining and Dimension Reduction techniques are applied through two of the following approaches:

- *Feature Selection Methods* (FSM) approach;
- *Full-Spectral-Analysis* (FSA) approach.

The *FSM* (also known as *optimal-spectral-analysis* method) translates into the search for a subset of wavelengths appropriate to the original, while with the FSA method all wavelengths are used and it includes methods of characteristic extraction that create combinations of new functions different from the original wavelengths (function of the processing space), as the main components (Bajwa and Kulkarni, 2011).

One method of spectral characteristics selection (FSM or FSA) aims at selecting "optimal" narrowbands with the objective of capturing those that contain the most information "sought" (the form of biotic stress), which is of fundamental importance to maximize the discrimination in the group (classification between healthy and diseased plants.

However, the optimal bands obtained are not always "invariant" in the spectrum (in time and space), but can vary depending on the phenological stage, types of cultivars within the same species, climatic conditions, providing a different "spectral contribution" in different portions of the spectral signature, throughout the growth cycle. This is very important in terms of design and optimal use of future multi-super spectral sensors with optimized bandwidth (10-50 bands max), dedicated to the recognition and monitoring of diseases (Verrelst et al., 2012).

There is no single approach for selecting the best narrowbands (Thenkabail et al., 2004) among the numerous statistical models of supervised or unsupervised classification methods, but there is an objective necessity to understand which is the most appropriate statistical method for the selection.

The FSM methods of *Multiple Linear Regression* (MLR), such as the *Stepwise Multiple Linear Regression* (SMLR) represent the most widely used standard regressive procedure not only to establish relationship between data from the spectral signature and characteristics of interest of the crops, but also to select the wavelength correlated with the properties of the crop.

However, given the high multicollinearity nature of hyperspectral data (the adjacent bands are similar), the SMLR technique has been widely criticized as vulnerable because of the problem of overfitting (Curran, 1989). Grossman et al. (1996), in effect, have shown that the bands selected by the SMLR method are not related to known absorption bands, as they were selected in similar studies. In other words, when the number of wavelengths (p) is large compared to the number of samples of plant observations (n), the method tends to exaggerate "the goodness of fit" because of regression coefficients that

are not bounded and are highly distorted. This increases the risk of selecting irrelevant wavebands, simply because the noise model is related to chemical response of the chemical prediction model calibration.

Multivariate statistical methods such as *Principal Component Analysis*, *Principal Component regression* (PCr), *Partial Least Squares Regression*and *Stepwise Discriminant Analysis,* are FSA methods, where their main objective is to reduce the space of the original wavelengths within a space of a few latent variables (t<p), where the new variables, called principle components (pcs), pls factors or variable root, are simply the linear combination of the original variables (Bajwa and Kulkarni, 2011).

PCA and SDA are two non supervised methods. While in PCA, the minimum threshold of the variance is preset to determine the optimal number of pcs, the SDA classification technique allows for the selction of variables that maximize the distinction between classes. In addition, to identify the optimal parameters of discrimination between the different samples (better separation between classes), SDA uses multivariate measures of separability, such as*Wilks' Lambda, F-Value* and average canonical correlation as criteria to identify the narrowbands. The supervised PLSR model is a powerful statistical tool used in chemometrics for predictive purposes and less used for the selection of variable importance.

The FSM and FSA methods presented for the analysis of hyperspectral RS PDD data, demonstrate advantages and disadvantages.

Given the redundancy, the collinearity and high-dimensional hyperspectral data, FSM models guarantee on the one hand, a careful selection of bands related to the biochemical/biophysical parameter (good interpretation, in terms of physiological importance associated with the selected wavelength), and on the other hand, the use of latent variable transformation (principle components, factors etc.) that make better use of all the spectral information available (guaranteed by the FSA methods).

Therefore, it can be concluded that both the FSA and FSM methods are equally valuable for RS PDD data analysis, and the need arises to be able to use them simultaneously.

Thenkabail et al. (2004), through a heuristic approach, combine the salient features of the two methods, by selecting "optimal" narrowbands related to biochemical/biophysical characteristics.

The authors, pinpoint with their method i) bands with minimal correlation (to solve the problem of collinearity and the correlation between portions of the spectral signature, i.e. eliminate similar or redundant information), ii) those with the highest informative contribution and finally iii) the bands able to better identify a property among many others (that correlate better with a "specific feature").

These properties have been modeled through the analysis of correlation between all bands (LLR^2), PCA and SDA. The three methods used provide complementary and additional information.

This combined heuristic approach has been successfully used to also discern the effect of nitrogen availabilty and to differentiate between different conditions of water regimes and discern plant diseases (Ray et al., 2010).

Cluster Analysis is one of the most widely used techniques for exploratory data analysis, with applications that include image processing, voice processing, information retrieval and web applications.

Data mining in agriculture is a relatively new field of research and the use of cluster analysis has only just been introduced.

Worner et al. (2013), apply the first neutral Kohonen Self-Organizing Map (SOM) algorithm for the prioritizing of risk of plant pests originating from the introduction of insect pests at territorial level.

Other examples of applications in agriculture reported in the literature concern the use of the technique in the epidemiological field, in crop management, estimates of precipitation on crops and others.

Clustering is defined as the partitioning of a set of data into groups (cluster), such that the points of the same group are similar, while the points of the different groups must be dissimilar (von Luxburg, 2007). This basic rule is critical, both for the design of new algorithms, as well as for the evaluation of the classification results.

The Cluster validity provides tools to validate the quality of the results of clustering algorithms applied, through the use of validity indices. In general, these are classified as internal indices, based on intrinsic information to the data and external indices based on prior knowledge about the data (Rendón et al., 2011).

3.2 *Materials and methods*

This section introduces the area of study and data, the spectral transformations, multivariate models and the process of model validation.

3.2.1 *Study Area*

The study area selected for research falls within the Apulia region, in an outbreak area of *Xf* in Gallipoli (in the province of Lecce) situated in southern Italy (Figure 3/1).

Fig. 3/1 - The study area: small olive orchard located in Gallipoli, Province of Lecce.

The study was conducted in a commercial olive orchard (18° 2' 12,2172" N, 40° 3' 48,9132" W) of about 0.5 hectares, a homogeneous irrigated area, planted in 1998. Main cultivars were: (cv.) Cellina di Nardò, (cv.), Leccino (cv.) and Frantoio, arranged in a

regular matrix of 6x5 m and distributed in similar percentages (approximately 53 trees-from a total of 174 trees). Other varieties were also present, including 12 trees of the Cima di Meli cultivar and 1 tree of the Carolea cultivar.

The varieties selected for the study were (cv.) Cellina di Nardò, an old variety present in almost all of the southern Apulia territory, and (cv) Leccino, a recently introduced variety (approximately 40 years old) mainly present in intensive orchards in various areas in Apulia. Both cultivars in the orchard were planted in the same year, and cultivated under the same agronomic conditions.

3.2.2 Plant sampling collection

For each of the varieties considered, 5 adult trees were selected for the study and chosen from adjacent rows as shown in Figure 3/2.

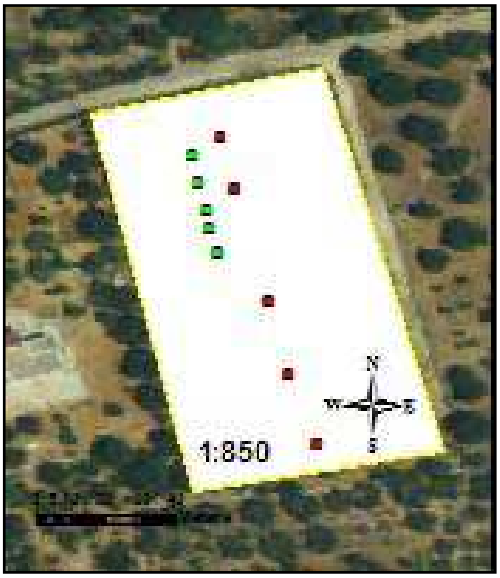


Fig. 3/2 - Localization of the plant subjects according to the cultivar.

Symptoms of infection and of twig die back in the branches were visible in almost all trees in the orchard (Figure 3/3) with the exception of very few specimens. The same symptoms were also noted in the surrounding orchards.

Fig. 3/3 - Leaf scorching and twig and branch die-back observed in the orchard.

The sampling was carried out in April by collecting twigs from 10 trees according to a rational sampling scheme.

The sampling was carried out in two days close together in April 2015 (DoY 103 and DoY 105; Day of Year), initially taking sprigs of plant material according to a rational scheme from 5 trees of the Cellina di Nardò variety and subsequently from another 5 trees of the Leccino variety (all 10 plants were symptomatic).

In particular, for each plant, a branch of foliage was selected (main branch) and subdivided into 3 levels according to the height. Subsequently, 4 adult branches (13-20 cm) from each level were taken at random far apart from the points that showed OQDS symptoms (if present). Each branch was finally divided into three subsections: basal, me-

dium and apical. Each subsection, with 2-5 leaves, was considered as a single sample (4-7 cm) to be subjected to spectral acquisitions (Figure 3/4).

Fig. 3/4 - Sampling scheme.

The association of each sample to the relative results was achieved through individual codification, able to distinguish between plant, cultivar, sampling canopy level, twig sample effected at any time.

Therefore, all this information was associated to the spectral signature (for example, “20150410 _CN1L1B3”: measurement date 20150410, plant 1, variety Cellina di Nardò, level 1, branch 3).

Since each sample was taken from the third subsection of the main branch, all individual samples generated were placed in small plastic bags, labelled with the same code, but updated with the letters A, C or M, according to the section of the twig: C = basal, M = medium and, A = apical (for example, “20150410 _CN1L1B3_A“: measurement date 20150410, plant 1, variety Cellina di Nardò, level 1, branch 3, apical section).

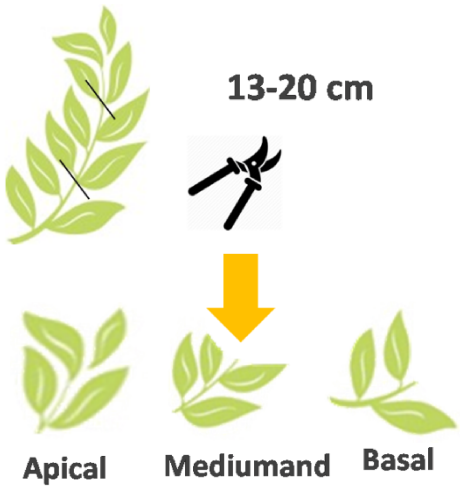


Fig. 3/5 - Collected twigs are divided into 3 samples (on the right) with specific codes.

All individual samples were collected, codified and subjected to analysis for assessing the presence or absence of *Xf*. Serological (ELISA) and molecular tests (conventional PCR) were performed at the accredited laboratory of the Mediterranean Agronomic Institute of Bari (CIHEAM/MAIB) to confirm the positivity or negativity of the tested samples after the acquisition of spectral signatures (see following paragraph).

In particular, for the slightly larger samples (6-7 cm) and with a higher number of leaves (4-5 leaves), analyses were conducted for each single leaf, in order to verify the "intensity of the infection" in each section of the twig.

3.2.3 Hyperspectral RSdata collection

The spectral signatures of attached leaves (upper surface) were obtained for all the samples gathered in plastic bags (encoded).

The spectral reflectances were collected from a portable spectroradiometer FieldSpec (ASD Inc., Boulder, CO, USA) through a specific acquisition interface made up of an optic fiber connected to a Plant-probe (Figure 3/6).

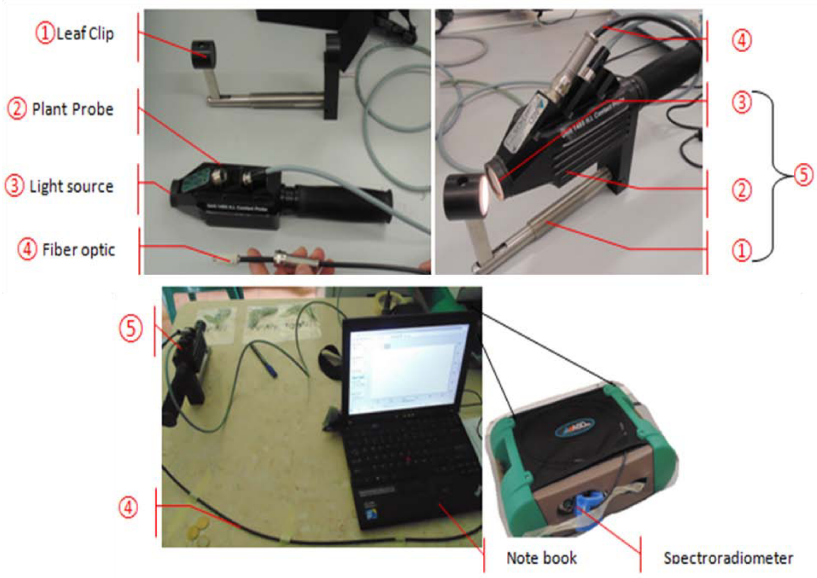


Fig. 3/6 - Acquisition scheme.

This proximal sensing tool allowed for hyperspectral reflectance data to be collected in the VIS-NIR range comprised between 350-1000 nm (FWHM: 3 nm) and in the SWIR range, between 1000-1830 nm (FWHM: 10 nm). The spectral channels are subsequently subsampled and interpolated, supplying spectral signatures of reflectance with spectral resolution of 1 nm at the output.

The plant-probe fibre optic spectroradiometer acquisition system, positioned on a workbench, was firmly anchored to the flat surface, to avoid optical absorption effects, present as a result of movements of the optical fibre in the measurement state. For every single sample processed (acquisition of the spectral signature of the associated leaves), an automatic procedure of calibration and reference was performed (internal to the instrument: white reference and light optimization) which guaranteed homogeneous and normalized spectra able to capture subtle changes of the spectrum for all measured leaves.

The reflectance data were collected during the course of the day, immediately after the codification of the vegetation samples in the field, maintaining the spectrum-leaf sample association.

So, for example, for the spectral signature acquired from the first leaf (defined 000000 by spectroradiometer) of sample 20150410_CN1L1B3, the spectral signature was codified as: "20150410_CN1L1B3_000000". All the spectral signatures obtained were stored by the spectroradiometer management software in a proprietary format (.asd).

3.2.4 Spectral pre-processing:Savitzky-Golay filter

The Hyperspectral RS data were appropriately processed before being analyzed by statistical models.Through the proprietary softwareViewSpecPro® ASD the raw data of the original reflectance were exported in a text format (American Standard Code for Information Interchange, ASCII) and imported into a routine Matlab R2011b (MathWorks, Inc., USA).

The Matlab script has corrected and filtered parts of the spectral signature that showed distortions or presence of instrumental noise (related to the measure). In particular, the "noise bands" close to near ultraviolet, were terminated (new range: 400-1830 nm) and in all the remaining bands a moving Savitzky-Golay with a frame size of 15 data points (2nd degree polynomial) was applied.

Mathematically, the filter operates simply as a weighted sum of neighbouring values as follows:

$$x_{j*} = \frac{1}{N} \sum_{h=-k}^{k} c_h x_{j+h}$$

Equation 3/1 - Savitzky Golay polynom.

Where x_{j*} as the new value, N is a normalizing coefficient, k is the number of neighbour values at each side of j and c_h are pre-computed coefficients, that depends on the chosen polynomial order and degree.

The routine in the end, corrected small spikes and offsets of misalignment of linear reflectance (at 1000 nm), due to the different nature of the VIS and SWIR detectors that couple around that wavelength (ASD Technical Guide 3rd Ed.).

The data thus processed, for both varieties, were collected in matrices of reflectance values (dataset) in different wavelengths and saved in excel format.

3.2.5 Dimensionality reduction of hyperspectral reflectance data

The statistical analyses of spectral reflectance data have been limited to subintervals of constant width of the original range which is considered noise-free (400-1830 nm). In this new spectral subdivision, which reduces the predictive variables (averages of the corresponding reflectance), the average value of each subinterval was considered the center of a narrowband with a width of around 10 nm.

The reason for reducing the variables is justified by the strong collinear nature of hyperspectral data. Previous studies, in fact, have shown that the near wavelength often provides similar information, thus becoming redundant (Thenkabail et al., 2004). The choice of the width of the subintervals, instead, was subjected to spectral resolutions of RS hyperspectral sensors available today. In effect, the first space-borne hyperspectral sensor, Hyperion, on board the EO-1 platform, has a spectral range of 400-2500 nm, withbroad band of 10 nm (Thenkabail et al., 2004).

Lastly, the problem of convergence of some multivariate algorithms applied to very large data sets, implemented in most of commercial statistical programs, is not to be underestimated. The 1430 original bands have been reduced to 143 new bands, each one with constant width of around 10 nm.

The two datasets were structured in the following manner: the columns represent the predictor variables (averages of the spectral reflectance in the central band) and the lines represent the spectral signatures of infected or healthy leaves (cases or observations). Basic statistical exploratory analysis, in conclusion, was performed on the two datasets to check the condition of normality (asymmetry and kurtosis) and the presence of outlier.

For this purpose, the Grubbs test and Henze-Zirkler algorithm (Mecklin and Mundfrom, 2004) were selected and the relative routine was implemented in Matlab Software.

3.2.6 Statistical methods: selection of optimal spectral bands

In this paragraph two multivariate statistical methods will be presented. They are able to select narrow spectral bands, starting from the reflectance spectra of leaves infected by *Xf*, of the two varieties of olive trees under study (Cellina di Nardò and Leccino). For each of the two methods, one of which is proposed by the author, a brief description of the procedure follows and explains how it was applied in this study.

3.2.6.1 Heuristic approach with combined models

For optimal narrowbands we mean a subgroup of narrowbands, of all bands of radiometric magnitude (reflectance in our case), that illustrate among themselves: i) minimum correlation; ii) high information content and iii) high discriminative power of a target characteristic (Thenkabail et al., 2004). These three properties are modelled in the following way:

- Lambda-Lambda model (LLR^2);
- Principal Component Analysis model;
- Stepwise Discriminant Analysis model.

The combined application of statistical models defines the Heuristic Approach model (HAM).

LLR^2model

This model seeks the Pearson linear correlation (elevated to the square), r^2, between pairs of wavelengths λi and λj (related to reflectances $R(\lambda i)$ and $R(\lambda j)$) (with i = 1, 2, p; j = 1, 2, p; with p = number of total wavelengths of the spectral signature).

Every single λi wavelength "is correlated" with every other λj wavelength. Byiterating this operation across the spectrum, a positive matrix (p x p) of r^2 coefficients that have values between 0 and 1 is constructed, which are visible on a graph.

A very high correlation between the two wavelengths (high value of r^2) indicates similar or redundant information. The lowest correlation areas instead (low value of r^2), indicate that the wavelengths concerned contain only unique information (Thenkabail et al., 2004, Jain et al., 2007).

The LLR^2 model was implemented in Matlab R2011b (MathWorks, Inc., USA) using the Statistics Toolbox library. Starting from the matrix of the correlation coefficient (squared) of the 143 narrowbands constructed (dimension 143x143), only the r^2 values less than or equal to a post correlation threshold equal to 0.005 were considered for analyses, with a more conservative level of significance α (equal to 0.01). The reasons for the choice of the threshold value have been dictated by common sense, because they are not comparable with those available in literature (Jain et al., 2007, Ray et al., 2010).

Principal Component Analysis model

PCA is a multivariate statistical technique used in the exploratory analyses of data. The aim of PCA is: i) to extract the most important information from the data set; ii) to compress its original dimensions, allowing for a simplified view in the transformed level and iii) to highlight "hidden structures" in the original data (observations and variables) if present.

PCA analysis can be conducted both through the data covariance matrix as well as through correlation (essentially normalized data). To use the covariance matrix, the predictor variables must have the same order of magnitude and no significant difference between their variances. If the data do not meet this condition, the use of the covariance matrix produces an increase of the variance and a greater allocation of weights, with consequent errors of the results.

In this study, the PCA model was implemented in Matlab R2011b (MathWorks, Inc., USA) using the Fathom Toolbox library (Jones, 2014).

The PCA analysis has been applied to the covariance matrix of the 143 reflectance media to obtain new variables able to explain the maximum information content of the originating data. In order to select the excellent bands that distinguish the leaves infected by *Xf*, only the principal components were maintained which explained more than 97% of the total variance. For these, the variables of *loading* that contributed with greater weight were selected. Within each main component extracted, therefore, all those bands

with the highest loading (in absolute value) were selected through a specific routine implemented in Matlab. The literature suggests a number of bands equal to five (Jain et al., 2007).

However, in the examples, application of the method related more to macroscopic aspects of a discrimination process (classification). So for this study, which analyzes precise asymptomatic leaves, it was decided to select all bands with the highest loading.

Stepwise discriminant analysis model

Discriminant Analysis is a multivariate statistical technique that is commonly used as a powerful classification approach for *data mining*, as it uses multiple quantitative attributes to discriminate single classification variables. A discriminant model, also known as a classification criterion, is determined by a measure of generalised squared distance. The classification criterion can be based on either the individual within-group covariance matrices or the pooled covariance matrix; it also accounts for the prior probabilities of the groups.

The Stepwise techniqueis a widely used method for the selection of variables in discriminant analysis. This method uses the combination of two algorithms: at each step (of analysis) it retains or eliminates a variable predictor, in function of the significance of their discriminatory capacity (using Wilks' Lambda, Pillai trace and average squared canonical correlation).

The SDA technique was used with success to distinguish plant species of different cultures through hyperspectral RS (Thenkabail et al., 2004). However, only the Wilks' Lambda values were indicative of the discriminatory power of spectral bands between the types of species.

Several researchers have highlighted three basic problems inherent in the use of stepwise methodologies: i) incorrect degrees of freedom; ii) sampling error capitalization and iii) failure to select the best subset of variables of a given size. They presented harsh criticisms for applications of these techniques.

The discrimination power of variables can be evaluated by means of the Wilks' Lambda which is defined as:

$$\Lambda = \frac{|W|}{|W + B|}$$

Equation 3/2 - Wilks' Lambda parameter.

where *W* is the within sum of squares and cross-products matrix and accounts for the average within class variability; *B* is the between sum of squares and cross-products matrix and accounts for the average between class variability. The Wilks' Lambda is related to the likelihood ratio criterion and ranges between 0 and 1, where values close to 0 indicate that the class means are different. Consequently, variables with the lowest Wilks' Lambda values can be retained in the classification model as optimal variables for separating the considered classes.

In light of these observations to determine the bands that maximize the discrimination between the leaves (property required by the heuristic model), it was decided to evaluate the “discriminant power” directly by applying the Wilks' Lambda test (measure of separability between bands) starting from knowledge of the categorical variable, infected and healthy plants on the basis of laboratory results.

The Wilks' Lambda test was implemented in Matlab R2011b (MathWorks, Inc., USA) using the Classification Toolbox library, version4.2 (Ballabio and Consonni, 2013).

The procedure was applied to the dataset of reflectance average of leaves and to the dichotomous vectors of the results, to identify the bands with the higher degree of separability between the originating bands (a conservative level of significance α equal to 0.01 was set for *Λ*).

In light of these observations, to determine the bands that maximize the discrimination between leaves (property required by the heuristic model), it was decided to evaluate the “discriminatory power” by directly applying Wilks' Lambda test (measure of separability between bands) starting from the knowledge of the categorical variable, infected and healthy trees, on the basis of laboratory results.

Wilks' Lambda test was conducted in Matlab using the Classification Toolbox version 4.2 library.

The procedure was applied to the average reflectance dataset of leaves and to the dichotomous vector results, to identify the bands with the higher degree of separability among the originating bands (for Λ a conservative level of significance α equal to 0.01 was set).

3.2.6.2 Combined general purpose detection method: interval PCA Internal Clustering Validation (iPCA-ICV)

From a detailed analysis of the literature the following critical aspects emerge for the management and the processing of Hyperspectral RS Data:

- data contains hundreds and even thousands of narrow and collinear wavelengths: problem of convergence of multivariate algorithms;
- the statistical methods employed for the early diagnosis of the disease are not general purpose: they are specific to the disease studied and are not generalizable to other crops and/or locations;
- statistical methods for selecting variables such as supervised classifiers tend to lean towards overfitting; a major complexity of the model does not correspond to a “good” classification; a large amount of training data is necessary;
- the supervised models are sensitive to anomalous data.

In light of what has emerged, an unsupervised *general purpose detection method* of hyperspectral data browsing with many collinear variables is proposed, which is able to search for clusters between spectral signatures and furnish a measure of separability, and provide a discrimination measure, for selection of bands with variable width, starting from reflectance spectral signatures of leaves infected by *Xf* and healthy leaves.

This deals with an approach combined with the PCA and Cluster Analysis mutlivariate techniques which follow the following logical steps:

- *Step-1*: apply proximity measures to detect outliers starting from processed post data;
- *Step-2*: subdivide the range of the hyperspectral data into a selected number of equal intervals;
- *Step-3*: apply the PCA analysis in each interval. It is assumed that the maximum information content is directly correlated to the maximum variance of the data. Retain the PCs that explain at least 80% of the variability of the total variation;
- *Step-4*: construct a geometric relationship (Euclidean) between the predictor variables and observations, on the basis of vectors from laboratory results at the level of the most informative PCs;
- *Step-5*: apply Internal Clustering Validation measures to select the "optimal" spectral bands.

All *iPCA-ICV* operating routines were developed and implemented in Matlab.

iPCA-ICV model: step 1 - outlier identification

In exploratory data analysis, a first step is to identify moderate or extreme "outliers" that perhaps should receive special scrutiny and possible action. This is accomplished by a ranking of data points according to "outlyingness", which then also can be used to determine boundaries delineating a "middle half" or "middle 75%" of the data set. For example, many practical classification problems are imbalanced; i.e., at least one of the classes constitutes only a very small minority of the data.

Proximities describe the similarity or dissimilarity between items or objects in a numerical way. Proximities are typically used for analyzing and visualizing hidden similarity structures by graphical displays or when looking for homogeneous clusters in a set of data (Van Mechelen et al., 2004). The proximity measure can for example be used to detect outliers. For each sample, the average squared proximity to the other samples in the class is calculated. The raw outlier measure is the number of samples divided by this average proximity (that is, the raw outlier measure will be large if the squared average proximity is small). The median and absolute deviation are calculated for the raw outlier measures of each class, and the raw measure for each sample in that class is then nor-

malized by subtracting the median from the raw score and then dividing by the absolute deviation.

The outlyingness measure, described by Breimanand Cutler (Breiman, 2006), was implemented in Matlab R2011b (MathWorks, Inc., USA) using the Fathom Toolbox library (Jones, 2014). In general, a value greater than 10 indicates that it is in the presence of an outlier. Unlike the analysis required by the heuristic model, the model proposed directly analyzed the post processed data sets, with no further verification on normality of data.

iPCA-ICV model: step 2-4 interval PCA

In a previous paragraph, the salient principles of the PCA technique have already been summarily described. It is worth mentioning that the decomposition of the original X matrix by PCA results in two matrices known as *scores* and *loadings*. The scores (t) correspond to the coordinates of the projection of the samples onto each PC and those that have similar scores will cluster together. The loadings (p) describe the way in which the original variables are linearly combined to new variables and define the direction of the principal components (Abdi and Williams, 2010).

The graphical visualization of the scores (scores plot) can be very informative as it reveals the inherent clustering, trends and outliers in a data matrix. The graphical visualization of the loadings (loadings plot) describes the influence of the measured variables in the model plane, and the relation among them. The direction of the measured variables corresponds to the observed patterns in the scores plot. The graphical output of the scores and loadings matrices can be visualized separately, as scores plot and loadings plot, or in the same graph as biplots.

The biplot is a two-dimension data visualization method that overlays samples (scores) and variables (loadings) of a data matrix, highlighting the relations existing within samples, within variables and between samples and variables. It provides a useful tool of data analysis and leads to capturing the most relevant features in the multivariate data set (clustering and correlations among variables) (Bro and Smilde, 2014).

For this part of the model development the Matlab iToolbox was used (Leardi and Nørgaard, 2004, Nørgaard, 2005), developed by the author for exploratory investigations

of data with many collinear variables (normal used in chemometrics), which make up an important part of the model.

The iTools library applies PCA analysis (globally and locally), permits the subdivision in subintervals and generates the graphs of the analysis results.

Since this concerns the absorption spectra of chemical substances similar in complexity and structure to the nature of a spectral signature obtained by a spectroradiometer, it was considered useful to use the code already developed, making the appropriate changes and corrections.

iPCA-ICV model: step 5-Internal Clustering Validation

Clustering, one of the most important unsupervised learning problems, is the task of dividing a set of objects into clusters such that objects within the same cluster are similar while objects in different clusters are distinct.

Clustering validation, which evaluates the goodness of clustering results, has long been recognized as one of the vital issues essential to the success of clustering applications.

External clustering validation and internal clustering validation are the two main categories of clustering validation. Unlike external validation measures, which use external information not present in the data, internal validation measures only rely on information in the data. The internal measures evaluate the goodness of a clustering structure without respect to external information.

As the goal of clustering is to make objects within the same cluster similar and objects in different clusters distinct, internal validation measures are often based on the following two criteria: compactness and separation.

Compactness: it measures how closely related the objects in a cluster are. A group of measures evaluate cluster compactness based on variance. Lower variance indicates better compactness. Also, there are numerous measures estimate the cluster compactness based on distance, such as maximum or average pairwise distance, and maximum or average center-based distance.

Separation: It measures how distinct or well-separated a cluster is from other clusters. For example, the pairwise distances between cluster centers or the pairwise minimum distances between objects in different clusters are widely used as measures of separation.

Most indices consider both of the evaluation criteria (compactness and separation) in the way of ratio or summation Index = (a * Separation) / (b * Compactness) where *a* and *b* are weights, or differences.

One category of internal indexes is based on these properties, and examples of this type are given by Dunn (1973), Davies and Bouldin (1979), Caliński and Harabasz (1974) and Silhouette index (Rousseeuw, 1987).

Two indices were considered in the *iPCA-ICV* model as evaluation parameters of the validity of the grouping results between infected and healthy leaves within each subinterval (narrowband): Calinski-Harabasz index and Silhouette index.

The Calinski-Harabasz index (CH) evaluates the cluster validity based on the average between and within cluster sum of squares. Assumes positive integer greater than 1. Two groups are well separated if very high values are assumed: the higher the value, the more obvious is separation and unity between the two groups.

The Silhouette index (S) validates the clustering performance based on the pairwise difference of between and within-cluster distances. In addition, the optimal cluster number is determined by maximizing the value of this index. Assumes values including between -1 and 1. The Silhouette value ranges from -1 to 1. A high Silhouette value indicates that is well-matched to its own cluster, and poorly-matched to neighbouring clusters. If most points have a high Silhouette value, then the clustering solution is appropriate. If many points have a low or negative Silhouette value, then the clustering solution may have either too many or too few clusters. The Silhouette clustering evaluation criterion can be used with any distance metric.

The two internal indices, CH and S were calculated with Matlab R2011b (MathWorks, Inc., USA) using the Cluster Validity Analysis Platform library - CVAP (Wang et al., 2009).

3.2.6.3 Heuristic Approach Model vs interval PCA Internal Clustering Validation

Classification among groups of infected leaves asymtomatic and healthy leaves, for both varieties of olive trees (with different susceptibility), on the basis of optimal band spectral profile selected from HAM and iPCA-ICV models, constitutes an important aspect of the research, given the different natures of the two classification approaches.

Therefore, to provide an objective assessment of performance a method of Generalised Discriminant Analysis based on a dissimilarity matrix to test for differences in prior groups of multivariate observations has been applied. This consists of using classical multidimensional scaling to obtain a low dimensional representation of the data for which Euclidean distances approximate the original dissimilarities. The scores in this representation are used in discriminant analysis for groups giving tests based on the canonical correlations (Anderson and Robinson, 2003). Classification is then achieved in the usual manner of discriminant analysis. Namely, we consider the Euclidean distance from the new point to each of the group centroids in the canonical space and classify where this function is minimised. A measure of the apparent error rate for this classification method has been obtained using the leave-one-out cross-validation method (Lachenbruch and Mickey, 1968), through a confusion matrix. The strategy is to leave one of the observations out of the analysis, then allocate that observation into one of the groups, as if it were a “new” observation, as described above. The proportion of incorrect allocations is then a measure of the error.

The GDA-LOOCV classification process was calculated with Matlab R2011b (MathWorks, Inc., USA) using the Fathom Toolbox library (Jones, 2014).

3.3 Results

This paragraph presents the results in the following sequence:

3.3.1 Laboratory analysis

Table 3/1 shows the overall statistics of the field and laboratory results, related to the variety, the day of sampling, the number of trees selected, number of branches sampled, the number of single samples extracted (small sprigs with 2-5 leaves) and the infection rate obtained by the outcome of the laboratory analysis of *Xf*.

Table 3/1 - Statistics related to samples taken from the field and the related tests for the assessment of *Xf*.

Variety	DoY	Trees [count]	Branches [count]	Single Samples [count]	Percentage [%]	
					Xf infected	*Non infected*
Cellina di Nardò	**103**	**5**	**60**	**180**	**85.7**	**14.3**
Leccino	**105**	**5**	**60**	**180**	**19.1**	**80.9**
Total		**10**	**120**	**360**		

3.3.2 Comparison of the cultivars: infection rate

The laboratory results obtained, for the total number of samples (360), were compared statistically using the two cultivars, utilizing One Way ANOVA Staistica 7 (StatSoft, Inc., USA). The evidence of the difference in the infection of the two varieties is represented in Figure 3/7.

Apparently, the Cellina di Nardò variety appears to be more prone to *Xf* infection. Future research will surely provide answers; however, at the end of this research, this is

the best result that could be expected for the contemporaneity of two target types of olive trees to different infection.

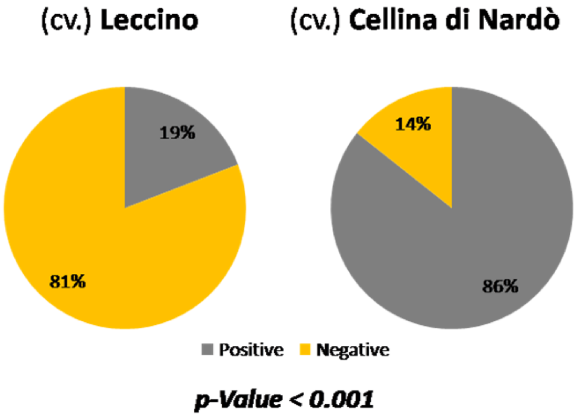


Fig. 3/7 - *Xylella fastidiosa* Infection rate according to the cultivar studied.

3.3.3 Spectral analysis

During the study period, 100% of the samples gathered were subjected to spectral surveys. As previously mentioned, the branches were chosen so as to cover entirely and evenly the whole canopy area (for all trees). Each branch was then further divided into three subparts and the selection of these single samples was carried out on the basis of the availabilty of leaves (minimum 2 per sample) and the quality of the sample (healthy-looking leaves).

However, the Leccino variety turned out to be in full vegetative development compared to the other (different phenological stage), and this entailed the sampling of parts with very young leaves. All leaves were tested (results are reported in Table 3/2) and spectral signatures were obtained, but about 20% of the total samples of Leccino (those not yet mature) were not subjected to statistical analyses. This resulted in a recalculation of the infection rate based on the total availability of positive and negative spectra

of Leccino leaves. In other words, only 585 spectra of Leccino were analyzed, of which 450 resulted negative (76.9%) and 135 positive (23.1%). It is necessary to clarify that for each single sample (basal, medium or apical part), only about 90% of the total plant material was tested for *Xf* (sum of 90% of the leaves present). Thereis no 1 to 1 match between the spectral signature of a single leaf and the respective laboratory results. It is reasonable to assume, given the high percentage of leaves tested for eachbranch (more than 90%), that the "infected" and "notinfected" property of the single sample is the same for every single leaf attached to it. Despite a small error, it is assumed that the result of the outcome of the group of leaves tested for *Xf* is the same for all leaves that make up the single sample. From a statistical point of view this does not involve great loss of information seeing as there is a large number of reflectance spectra available.

Table 3/2 reports the number of spectral signatures acquired in the field (DoY 103 e DoY 105) and the respective leaves of the two laboratory processed varieties.

Table 3/2 - Spectral signatures and related leaves tested for *Xf* in the sampling period.

Variety	DoY	Spectral signature [count]	Leaf [count]	
			Xf infected	*Non infected*
Cellina di Nardò	**103**	**795**	**681**	**114**
Leccino	**105**	**585**	**135**	**450**
Total		**1380**		

The spectral curve obtained at leaf level of the two species of olive trees wereacquired under controlled conditions (light and temperature). This has limited the reflectances to be excluded from the analysis of the models, however, their pre-processing is thus required as described in paragraph 3.2.3.

Figure 3/8 shows the variability detected by the spectral reflectances of 795 leaves from Cellina di Nardò (A) and and 585 leaves from Leccino (B).

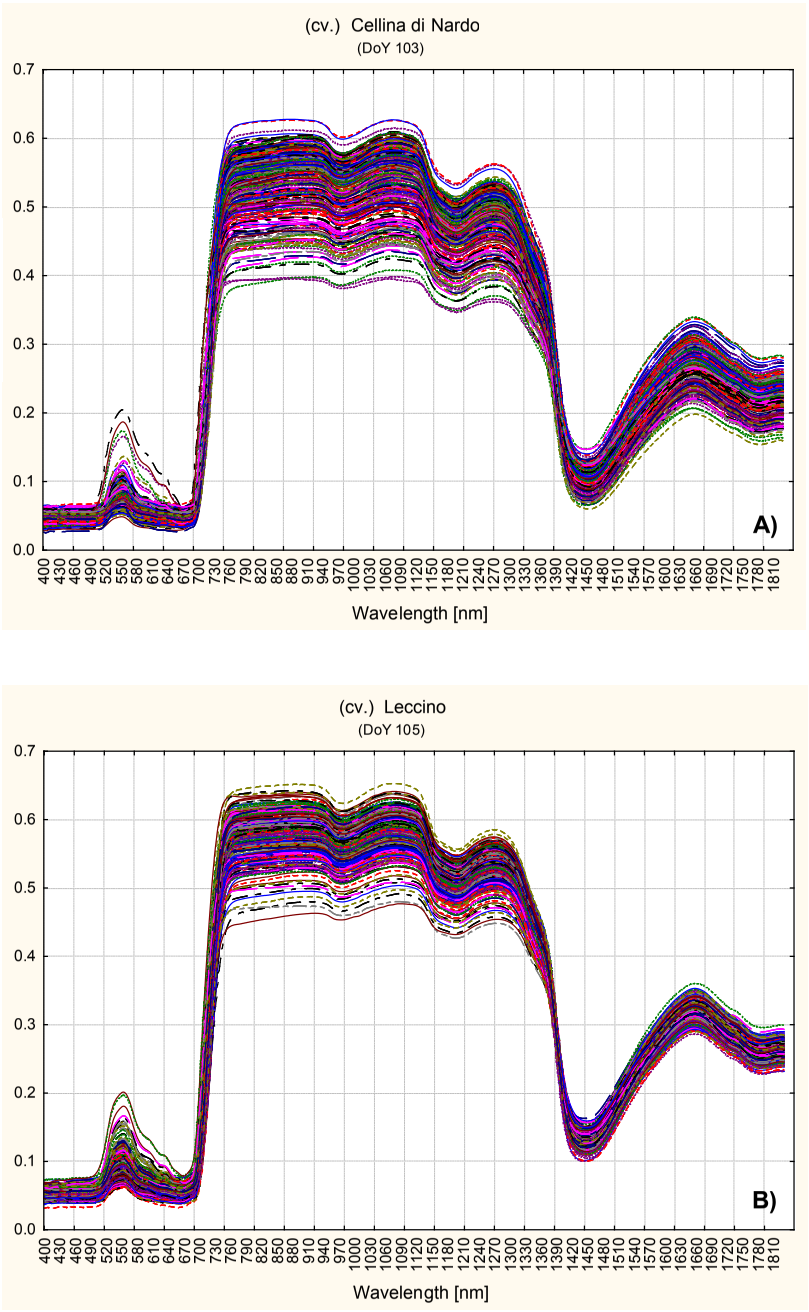


Fig. 3/8 - Field hyperspectral measurement (n=1380): (cv.) Cellina di Nardò (A) and (cv.) Leccino (B).

For both varieties, the lowest variation is found in the visible region (VIS) (~ 400-700 nm), especially in blue (~450-495 nm), red (~650-680 nm) and in the infrared region of shortwave (SWIR) for the Leccino variety (~1300-1830 nm). Light variation, instead, occurs in the green region (~520-610 nm), probably linked to the presence of younger leaves. In VIS, the reflectance varies greatly, towards the near infrared region (NIR) (~800-1300 nm), as a result of light diffusion with in the internal structure of the leaf. The red-edge area (~680-730 nm) is quite narrow for both species.

Two water absorption regions are evident in the NIR around 970 nm and 1170 nm. Moving to the longest wavelengths (SWIR), two deep absorption characteristics around 1449 nm and 1966 nm are observed, mainly caused by water that is still in the leaf, and to a lesser extent, by ligno cellulosic compounds in the mesophyll (their absorption characteristics are identified in this region).

The difference between the two spectra is clear, both in shape and in dispersion (especially in NIR) but above all it is worth highlighting the “apparently healthy” nature of the spectra obtained (asymptomatic, in the presence of *Xf* infection).

3.3.4 Selection of optimal spectral bands through Heuristic Approach model (HAM): LLR², PCA and Wilks' Lambda methods

All Matlab libraries that implement statistical methods contained in the HAM model have been implemented in one main code.

A logical intersection condition was adopted in the developed code to select the optimal narrowbands, from those with low correlation, greater information and high discriminative potential.

This choice appeared to be the most obvious, although the literature has adopted the solution of "increased occurrence" bands (Thenkabail et al., 2004, Jain et al., 2007, Ray et al., 2010). Although it is shareable because it is linked to the low flexibility of software statistics adopted by the authors, this approach highlights inconsistencies on the precision that they then impose on the minimum/maximum selection threshold of the "property" of the relationship, information or discrimination applied by three of the HAM model methods.

The HAM model, in the Matlab framework, was carried out with the following set parameters: number of sub-intervals equal to 143, maximum variance explained by principal components equal to 97% (of total variability), maximum correlation allowed between the variables equal to 0.005 (p value < 0.01) and the separation ability of leaves infected or not infected by *Xf* in the 143 narrowband range, with a < 0.01 p value.

Even though the 143 narrowbands showed a normal univariate distribution (Shapiro-Wilk test), the Henze-Zirkler test demonstrated that the Cellina di Nardò and Leccino datasets are not distributed normally.

The HAM procedure, however,was performed without further recommendations (Ray et al., 2010, Stellacci et al., 2016, Jain et al., 2007, Thenkabail et al., 2004).

Figure 3/9, shows the matrix of the combination of the 143 bands, generated by the Matlab code, with the lowest significant correlation (almost black sections), or minimum redundancy of information, for both varieties.

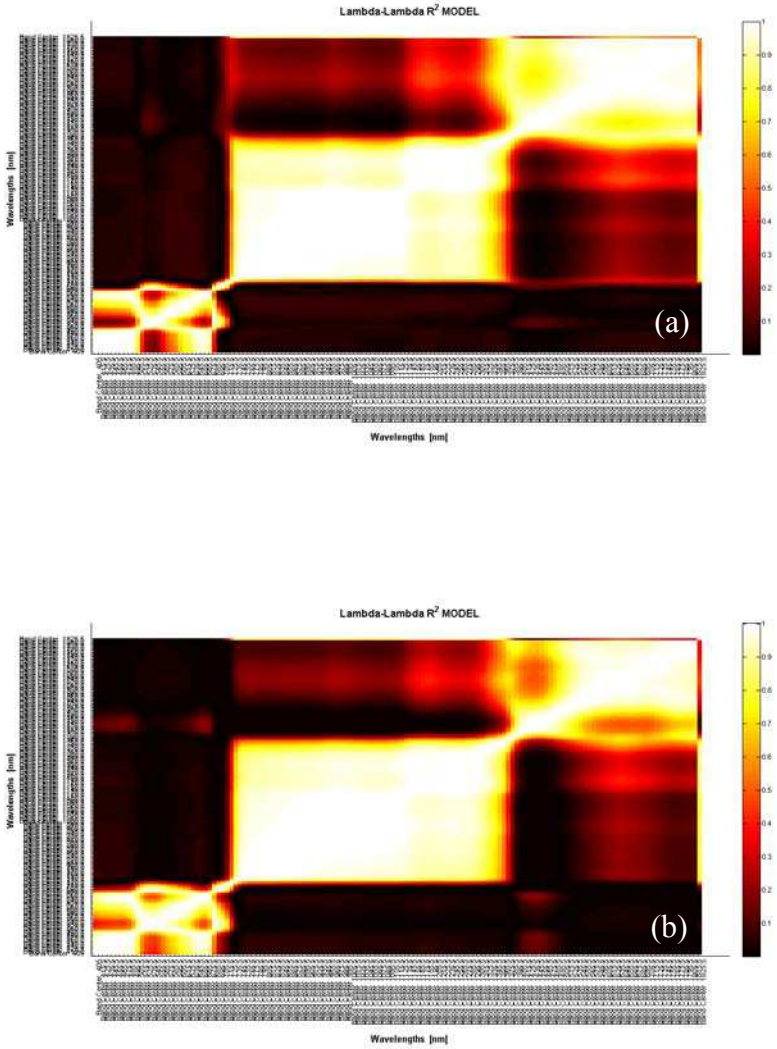


Fig. 3/9 - Correlation matrix LLR^2, relative to both varieties, in function with the 143 narrowbands: (a) Cellina di Nardò and (b) Leccino.

The unique matrices of the 143 bands are instead reported in Figure 3/10 for both varieties.

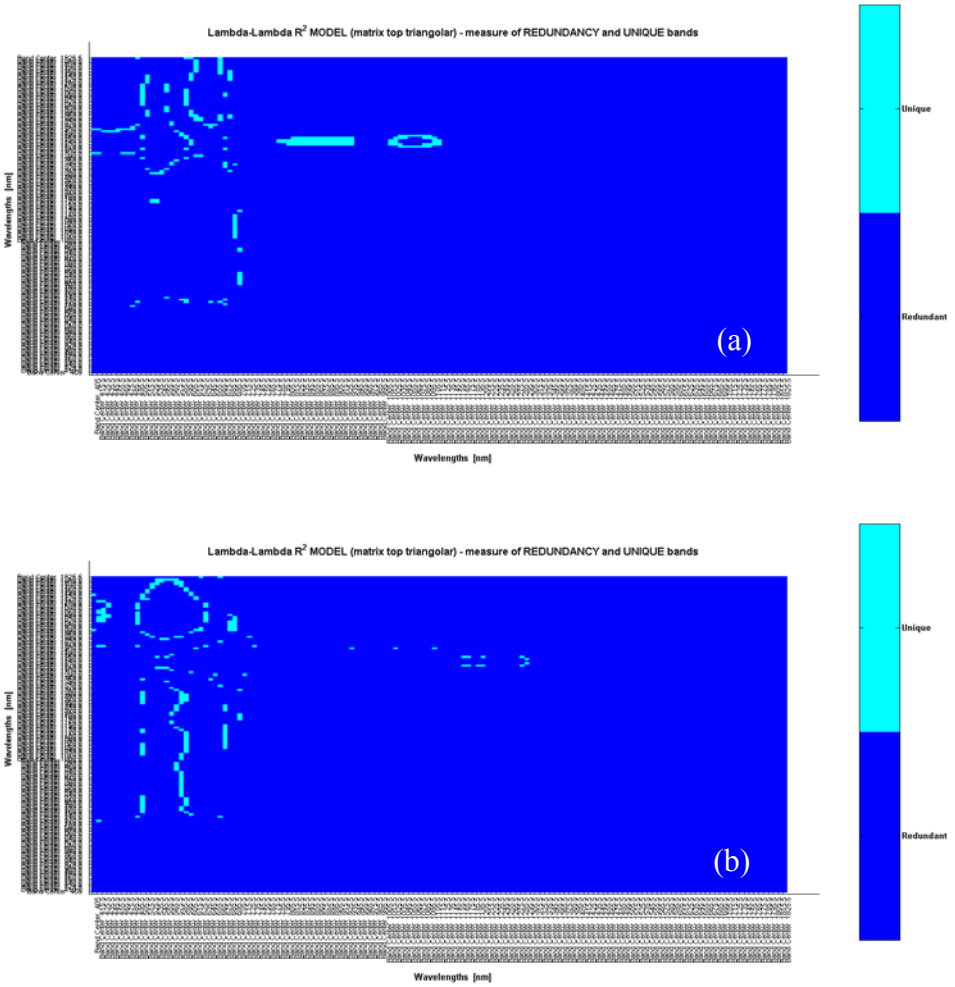


Fig. 3/10 - UNIQUE correlation matrix (LLR^2), relative to both varieties, in funzione with the 143 narrow-bands: (a) Cellina di Nardò and (b) Leccino.

The section in cyan represents the combination of bands with lowest correlation coefficient squared ($r^2 < 0.005$, p value < 0.01) or unique bands, compared with those of higher correlation (highlighted in blue). The two varieties have different unique bands. With regards to the PCA analyses, for both varieties, the threshold set to 97% retained the first two or three main components (PC1, PC2 and PC3) which explains the cumula-

tive percentage of the total tax variance. At the time of acquisition of the spectral reflectance, the olive leaves appeared to be apparently healthy. Therefore, the variability between the state of infection and non infection was minimum (it must be noted that the sampling design was meant to maximize the effect of early recognition of *Xf*). This justifies the retention of the noisier principal components in the code (e.g. PC3 or PC4).

Regarding the separation capacity among the 143 narrowbands defined, implemented by the Wilks' Lambda tests Figure 3/11 further highlights the difference between the two varieties.

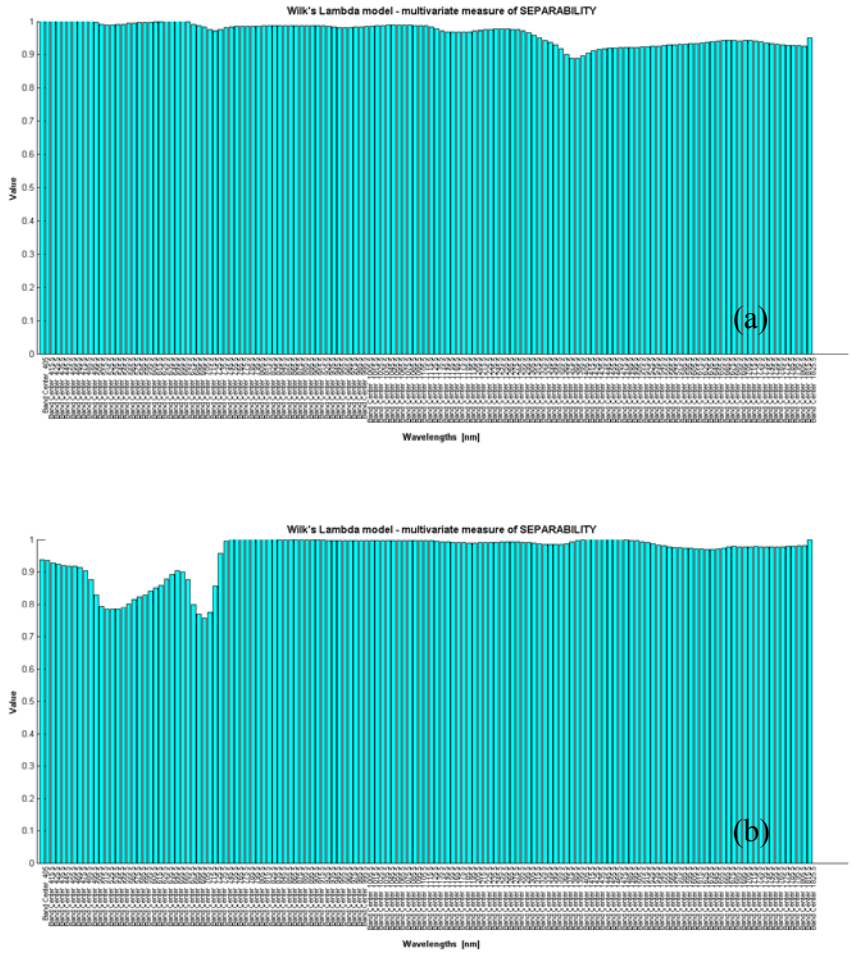


Fig. 3/11 - Spectrum of separable potentials between healthy and infected leaves of both varieties: (a) Cellina di Nardò and (b) Leccino.

The two graphs generated by the program demonstrate areas of the spectrum with different discriminating power. While for Cellina di Nardò the SWIR bands appear to prevail (area of water absorption, lignin and cellulose, mainly), Leccino tends to mainly differentiate healthy and infected leaves in the visible region.

The excellent narrowbands selected for early detection of *Xf* (synthesis and overlapping patterns of *LLR*^{2}*, PCA* and Wilks' Lambda), obtained by the implementaion of HAM; are reported in Figure 3/12 for both varieties.

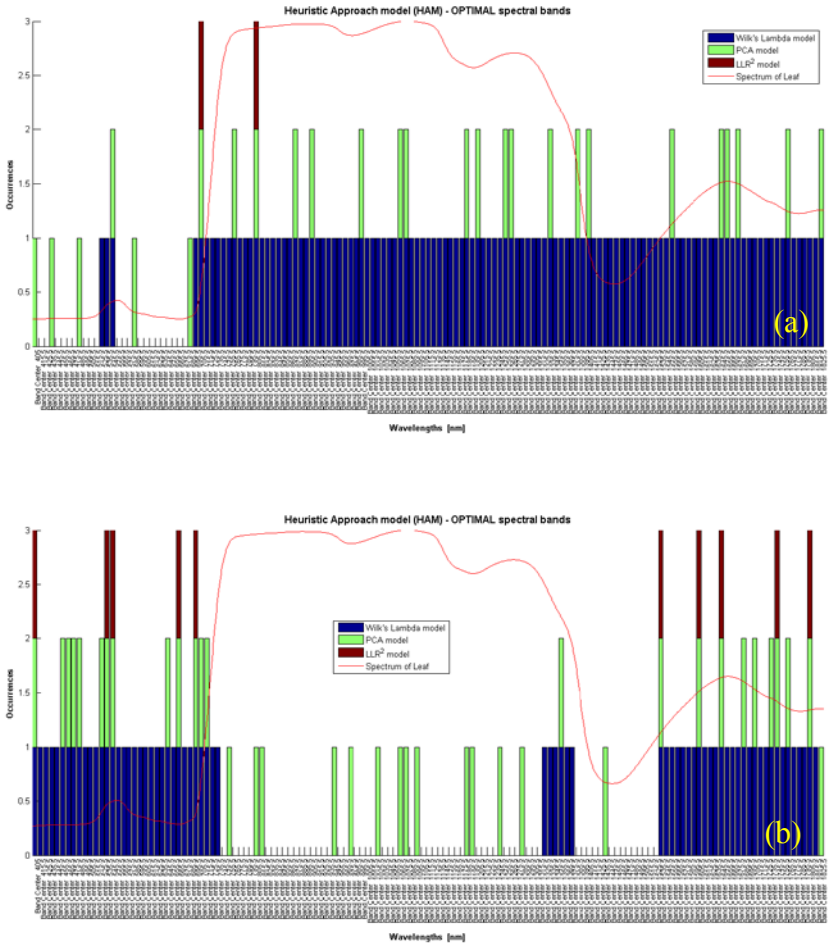


Fig. 3/12 - Excellent narrowbands selected from HAM model to select leaves infected or not infected by *Xf*, superimposed on the relative spectral signature average: (a) Cellina di Nardò and (b) Leccino.

The bars in blue represent the bands that are capable of better separating the disease, those in green are the bands with the most information and the brown bars represent the bands that are non redundant.

The excellent bands are those represented by the superposition of the three effects. According to the HAM model, *Xf* can be distinguished through the spectral reflectance of the leaves, in the narrowbands (10 nm) centered at 405, 535.5, 545.5, 665.5, 695.5, 1535.5, 1605.5, 1645.5, 1745.5, 1805.5 nm, for Cellina di Nardò and 705.5, 805.5 nm for the Leccino variety.

The two varieties show different conditions of *Xf* detection. It is only in the vicinity of the red-edge that they show similar conditions of recognition.

3.3.5 Selection of optimal spectral bands through a combined general purpose detection method: iPCA-ICV

The general purpose detection model, iPCA-ICV, proposed in this research, allows for the outlier management, and to create subdivision conditions from the original spectrum in subintervals in the desired width, to apply PCA to each of them, to constructa Euclidean metric and apply an evaluation measure on groupings (infected and non infected categories).

Many of these actions have been implemented in the model from libraries, others have been improved and some entirely developed according to need. In particular, for a consistency in the results a portion of the code that implements the PCA has been aligned with the theoretical basis (Legendre and Legendre, 1998) implemented in the library Fathom (Jones, 2014). Additional minor adjustments concerned improvements in the management of the input dataset formats and in the graphical display of some typical parameters of the PCA (explanation of variance etc.).

A Euclidean metric was developed and integrated in the iTools model to evaluate the centroids of the two groups (infected and non infected) and to calculate the classification distance not available in the library.

The iPCA-ICV model, in the Matlab framework, was executed following the following parameters: number of subintervals equal to 143 and an outlyingness threshold of

5 (conservative conditions compared to 10, proposed in the literature. Compare paragraph 3.2.6.2).

Figure 3/13, shows the effect of outlyingness function on datasets of both varieties with a threshold equal to 5 (displayed in blue horizontally).

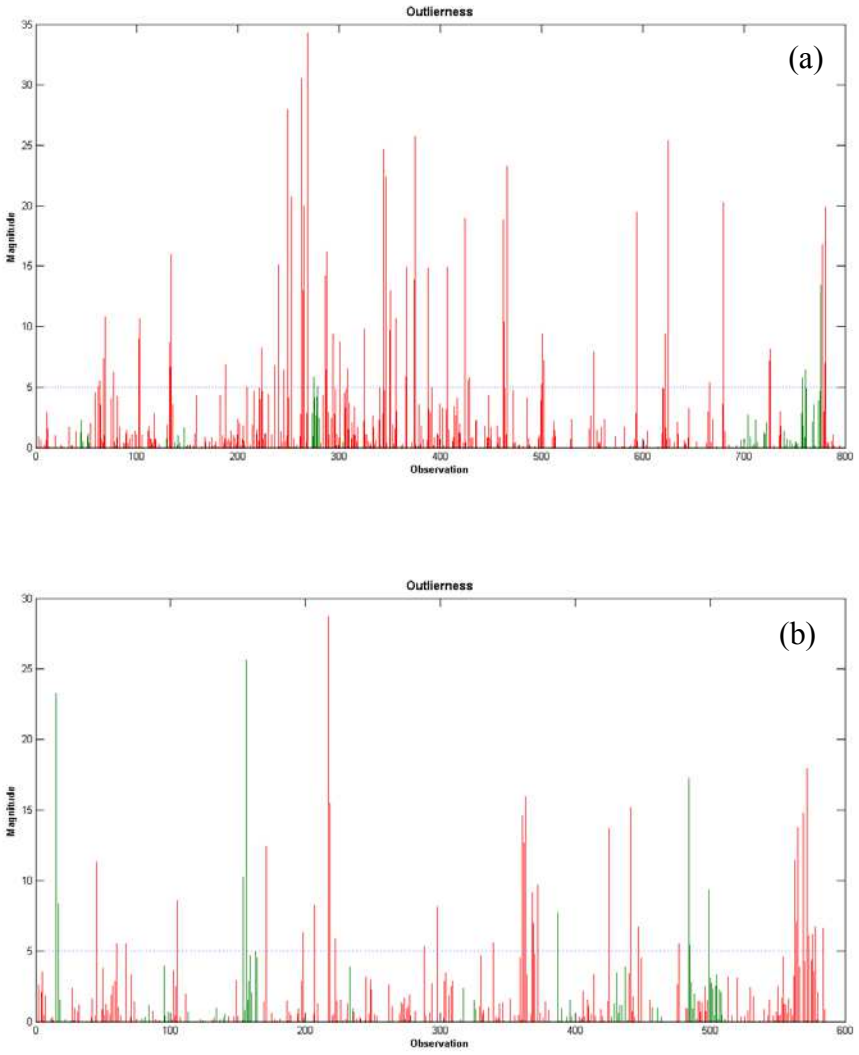


Fig. 3/13 - Selection of the observations (spectral signatures of the leaves) deemed “extreme cases”, relative to the two varieties, through the dissimilarity function. The observations that exceed 5 are discarded from the model: (a) Cellina di Nardò and (b) Leccino.

The spectral reflectance variabilty highlighted in paragraph 3.3.3 is reflected from a point of view of outliers.

Once put into action,for both varieties, the iPCA-ICV, generates, in the processed level of the principal components PC1 - PC2, the outputs of an global analysis (less important), conducted across the reflectance spectrum and a local analysis conducted in all of the 143 subintervals of the average reflectance (Figure 3/14).

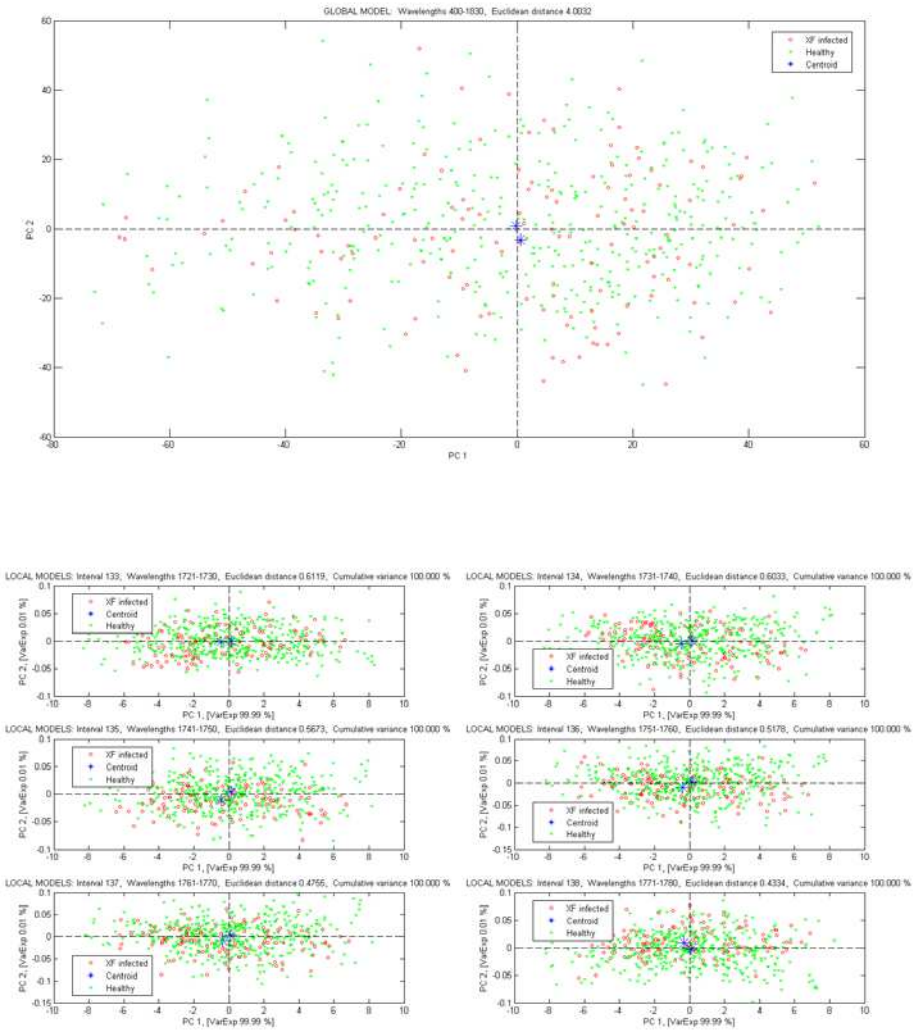


Fig. 3/14 - Global and local analyses of PCA relative to the Leccino variety.

From the table two groupings (classes of presence of pathogen) in red (leaves infected by *Xf*) and green (non infected leaves) are deduced.

Two blue star shaped points identify the centroid of the two classes, which provide through their distance, a measure of the separation between the groups. At this point the function of PCA at local level is clear.

In a very narrow range (10 nm), in which the reflectance collinearity is strong, PCA naturally creates two groups that explain almost 100% of the total variance of the range.

In the HAM method this translated into a single value: the average reflectance value. The availabilty of having two inherent clusters, strongly linked to the presence of infection (vectors of laboratory results), permits the application of methods and techniques of cluster analysis.

In this case, the algorithm that groups the best data is not sought (e.g. k-means) but the separation condition imposed by health of the leaves is taken advantage of.

The collinearity associated with each interval is related to the section of the spectral signature typical to the variety.

The principal components levels generated by the model can highlight this strong relationship, as can clearly be seen from the charts of the scores. It is easy to infer that the narrowbands capable of early detection of *Xf* are deduced from analysis of internal indices.

Calinski-Harabasz index and Silhouette index average, applied to groupings identified by PCA in each range (Figure 3/15).

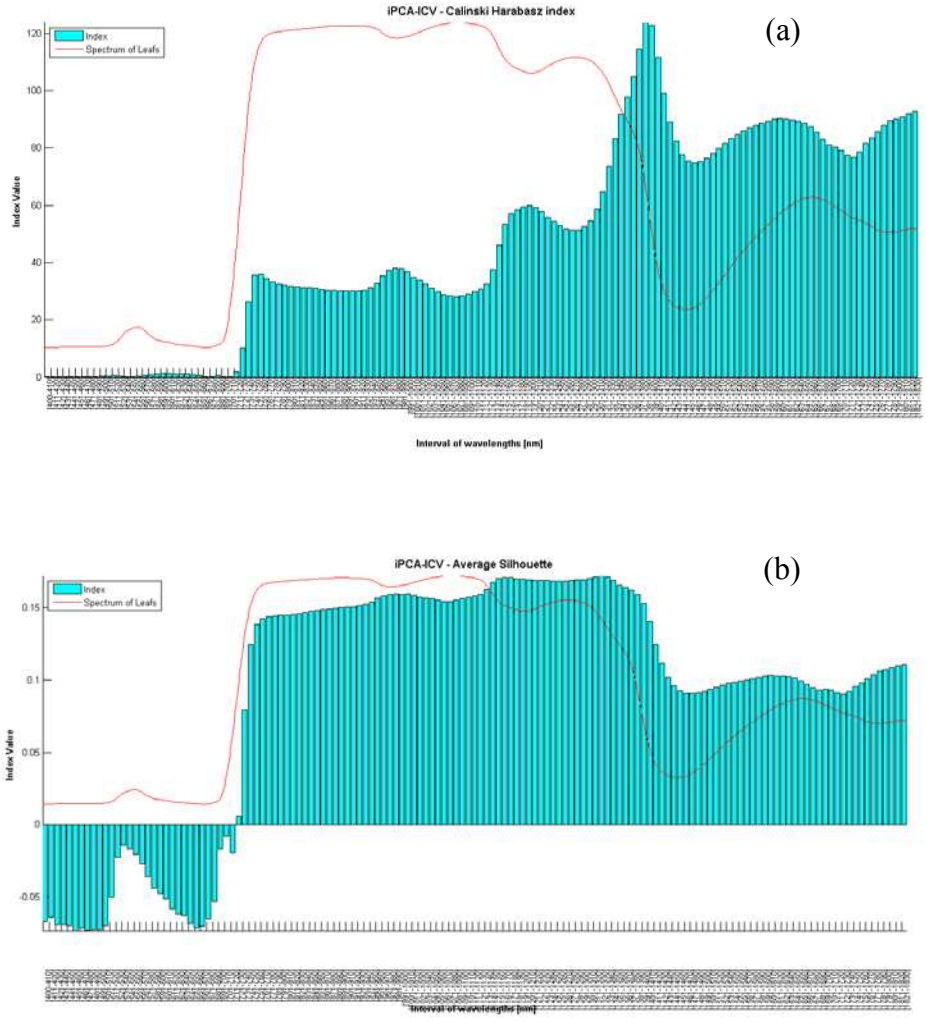


Fig. 3/15 - Internal indices overlap the average spectral signature used to identify the excellent bands related to the Cellina di Nardò variety: (a) Calinski-Harabasz index, (b) Silhouette index.

The best narrowbands for *Xf* detection are calculated from iPCA-ICV, on the basis of an objective function that extracts the maximum for the Calinski-Harabasz envelope index (high-performing in terms of Separation).

According to the iPCA-ICV model, *Xf* can be distinguished through spectral reflectance of the olive leaf, in the bands (751-760 nm), (901-910 nm), (971-980 nm), (1191-1200 nm), (1381-1390 nm), (1601-1610 nm), (1821-1830 nm), for Cellina di Nardò and in the bands (451-460 nm), (481-490 nm), (671-680 nm), (721-730 nm), (1141-1150 nm), (1321-1330 nm), (1451-1460 nm), (1571-1580 nm), (1721-1730 nm) for the Leccino variety. Figure 3/16 compares the CH index for both varieties.

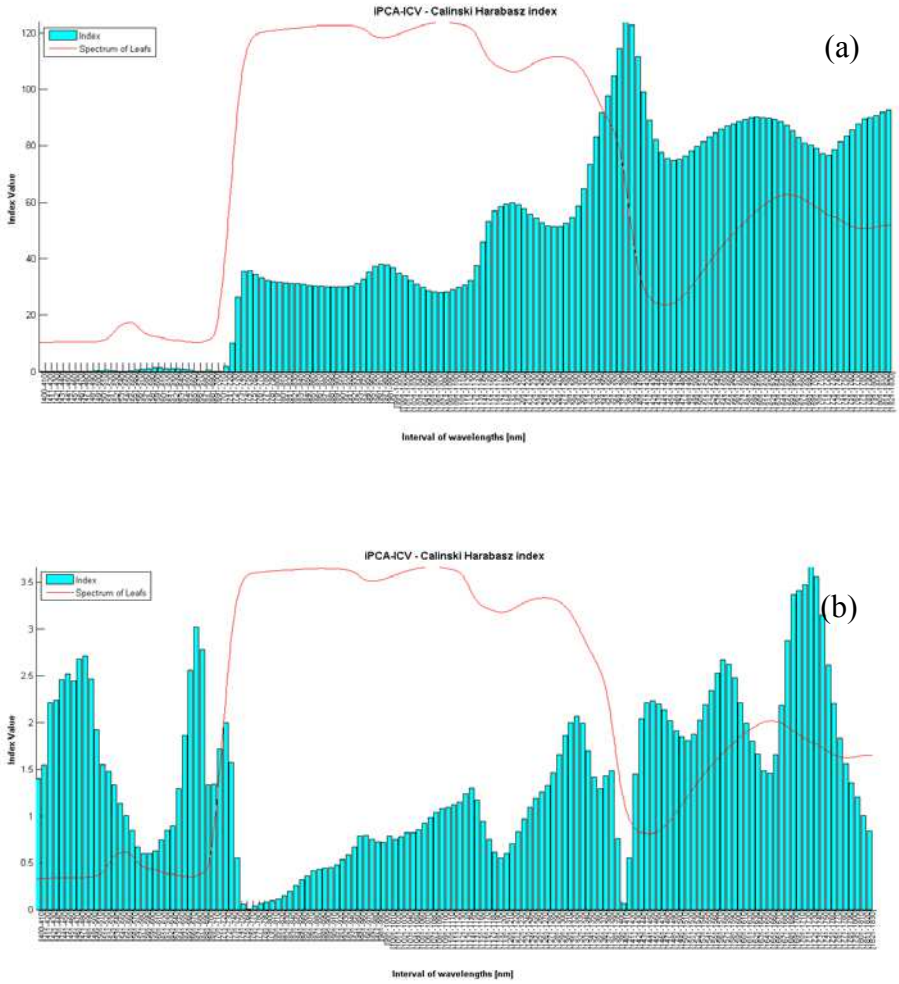


Fig. 3/16 - Spectrum of Calinski-Harabasz index related to both varieties superimposed on the spectral signatures: (a) Cellina di Nardò and (b) Leccino.

The two varieties show different conditions for *Xf* detection, combined in some ranges. In relative terms, the smallest value of the CH index calculated for Cellina di Nardò (approx. 40) is evaluated at about 10 times more compared to the biggest value for Leccino (approx. 3.5). In Figure 3/17 the average Silhouette index between the two varieties is compared.

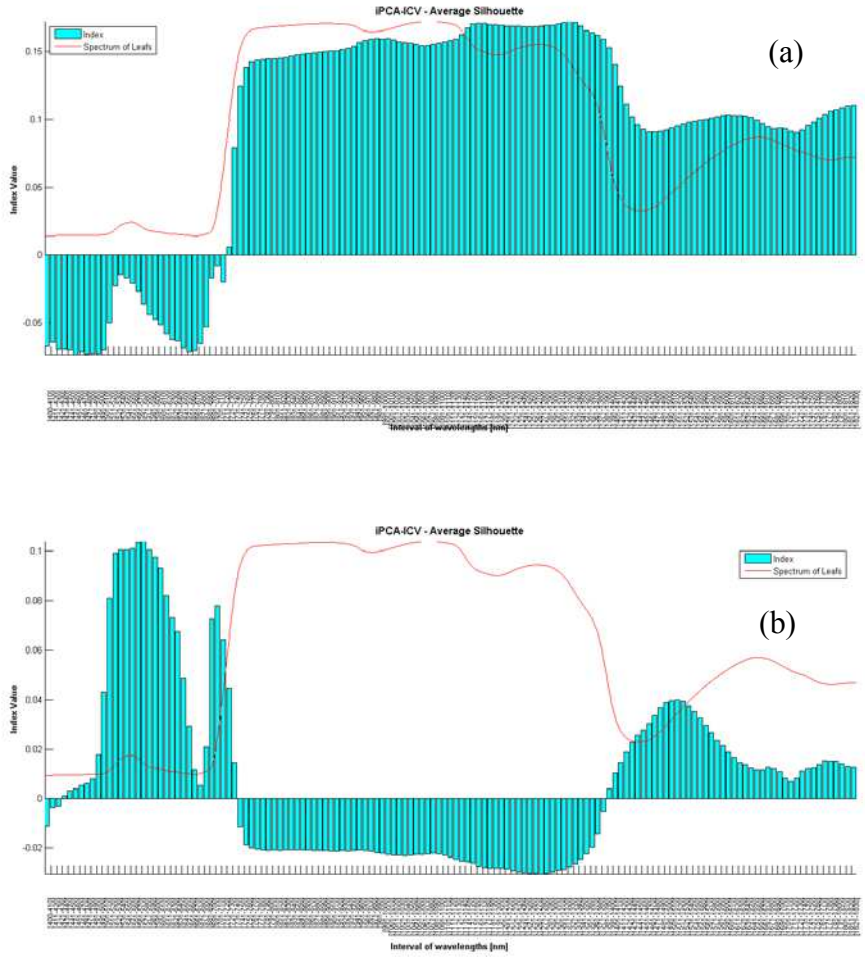


Fig. 3/17 - Spectrum of average Silhouette index related to both varities superimposed on the spectral signatures: (a) Cellina di Nardò and (b) Leccino.

The two spectra are comparable in the distant part of SWIR.

3.3.5 Best overall model: optimal band identification

The results of the Generalized Discriminant Analysis based on canonical correlation, GDA-LOOCV (through a measurement of error such as *leave-one-out cross-validation*), for both methods and both varieties, are reported in Table 3/3 and 3/4.

Table 3/3 - Confusion matrix derived from GDA-LOOCV applied to the bands selected through the (i) HAM (LLR^2, PCA, Wilks' Lambda) and (ii) iPCA-ICV.

Generalized Discriminant Analysis Leave-One-Out Cross-Validation-GDA-LOOCV Confusion Matrix (%)					
		HAM		*iPCA-ICV*	
		Xf infected	Non infected	*Xf* infected	Non infected
Cellina di Nardò	*Xf* infected	**58.1**	**41.9**	**76.2**	**23.8**
	Non infected	**45.9**	**54.1**	**22.9**	**77.1**
Leccino	*Xf* infected	**10.2**	**89.8**	**10.2**	**89.8**
	Non infected	**74.4**	**25.6**	**74.4**	**25.6**

Table 3/4 - Cross-validation error rates (%) calculated through a confusion matrix derived from GDA-LOOCV applied to the bands selected through the (i) HAM (LLR^2, PCA, Wilks' Lambda) and (ii) iPCA-ICV. The number of bands used.

Generalized Discriminant Analysis Leave-One-Out Cross-Validation-GDA-LOOCV		
	HAM (LLR^2, PCA, Wilks' Lambda)	*iPCA-ICV*
Cellina di Nardò	**42.27 (10)**	**23.70 (7)**
Leccino	**22.02 (2)**	**22.02 (9)**

The results of the GDA-LOOCV analyses (Table 3/4) show, in general, a rather high Error Rate. This may be due to the hypothesis of work: there is no 1 to 1 correspondence among the spectral signature of single leaves and the respective laboratory result (compare paragraph 3.3.3).

In general, the maximum discriminating power was obtained from the iPCA-ICV model proposed for both varieties (error rate equal to 23.7% and 22.02% respectively for cv. Cellina di Nardò and cv. Leccino).

To support the performance of the results shown in the table, a discriminant analysis was conducted on the narrowbands selected from two models, for each variety. The discrimination results are shown in Figure 3/18 and are quantified by Wilks' Lambda (Table 3/5).

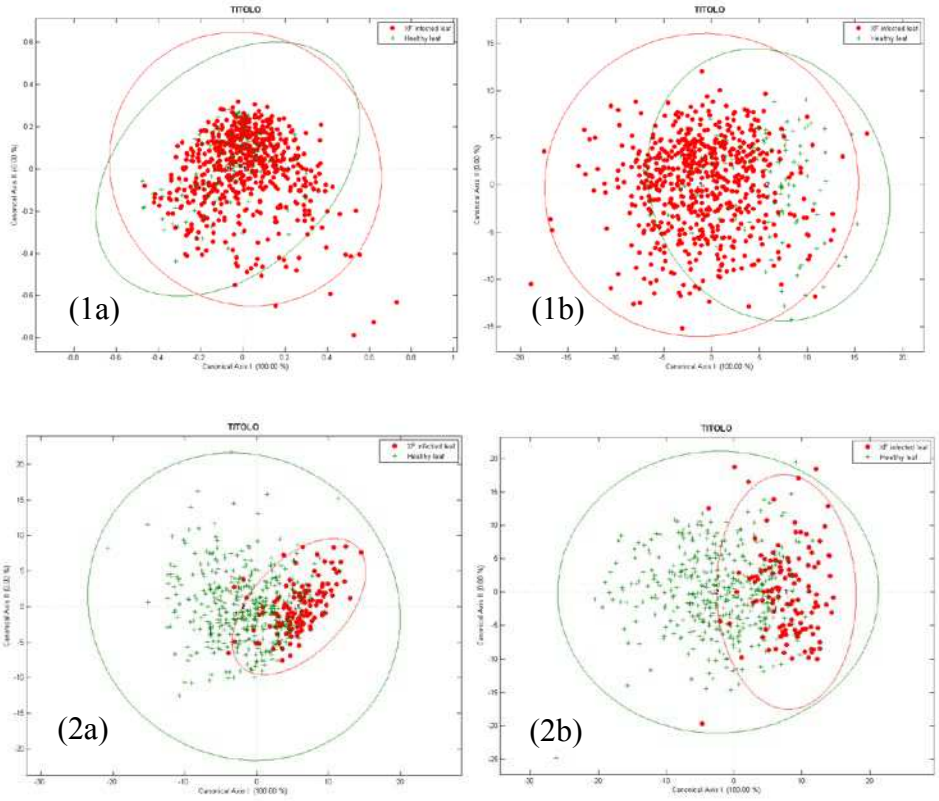


Fig. 3/18 - Canonical Discriminant Analysis applied to the bands selected by the two models: (1a) HAM on Cellina di Nardò, (2a) HAM on Leccino, (1b) iPCA-ICV on Cellina di Nardò, (2b) iPCA-ICV on Leccino.

Table 3/5 - Wilks' Lambda calculated through Canonical Discriminant Analysis applied to the bands selected through the (i) HAM (LLR^2, PCA, Wilks' Lambda) and (ii) iPCA-ICV. The number of bands used.

	HAM (LLR^2, PCA, Wilks' Lambda)	iPCA-ICV
	Wilks' Lambda	
Cellina di Nardò	**0.9727** *(10)*	**0.7874***(7)*
Leccino	**0.6718***(2)*	**0.6634***(9)*

3.4 Conclusions

The problem of dimensionality reduction of hyperspectral data, selecting a subset of bands from the signal, is a key aspect for efficiently and completely capitalizing on the potential of hyperspectral data.

An important aspect of this processis to maintain all the key wavelengths, that contain significant information without losing the discriminative power, downlining the statistical method which it processes (Thenkabail et al., 2012). Choosing the right data is therefore, an important aspect of Data Mining.

The research conducted in this direction has shown that the early symptoms of the *Xylella fastidiosa* infection can be detected using hyperspectral reflectance measurements acquired from olive leaves. The identification of the presence of the pathogen, however, cannot be generalized between the different varieties. The bands used in this research were found to depend on the model and on the variety.

For the Leccino variety, two different methodologies selected bands that correlate in proximity in the blue area, the red and SWIR from a distance. In particular, the closest bands are found in the range comprised between 660 and 690 nm and 1720 and 1750 nm.

For the Cellina variety, the methods selected bands that overlapless, the heuristic method, HAM, identified a band in the red-edge and one in the immediate plateau (805 nm), while the proposed iPCA-ICV method was able to reveal all the water absorption levels (971-980 nm, 1191-1200 nm and 1381-1390 nm) further with in the SWIR spectrum.

The bands selected with both methods showed coherence with the level of infection detected in the asymptomatic leaves of the two varieties (19.1% in Leccino and 85.7% in Cellina di Nardò).

A lower level of infection, probably, was expressed, from the spectral reflectance aspect, through disorders mainly in the visible region (at the turn of the peaks of chlorophyll absorption in the blue and red regions); levels of higher infection, instead, that manifested in the more susceptible variety, highlight a condition closer to the manifestation of the symptoms, expressed through increasedloss of water content of the plant, together with structural changes of the leaf (SWIR region according to Thenkabail et al., 2014).

The iPCA-ICV method proposed was high-performing compared to the heuristic model selected from the literature (HAM), for both varieties, also for its characteristic of supplying n easily interpretable graphical representation and coherence with the development of the symptoms of the disease.

With regards to the HAM model, unstructured with respect to the spectral signature (to minimize the complexity of the data), iPCA-ICV follows through values taken on from Calinski-Harabasz index or from Silhouette index the maximum difference between classes (distance), if it exists, through the natural separation between gruops in the spectrum (infected and non infected for every narrowband).

The separation value and that of the compactness of the indices showed to be different and directly proportional to the level of infection found in the two cultivars studied.

It would be an advantage to be able to select specific bands for two reasons: i) it helps to select the wavelength more correlated to the phenomena which the object of the study and ii) allows for the development strategies to select the wavelengths for the realization of specific sensors, such as those on board new generation satellites or UAV, in particular in the application of precision agriculture.

CONCLUSIONS

The Mediterranean area is home to a wide biodiversity of olive plants which must be protected for social, economic and environmental reasons from potential genetic erosion caused by the recent phytosanitary emergency of *Xylella fastidiosa,* recently introduced in Apulia, South of Italy. Indeed, this pathogen is severely affecting olive trees with inestimable impact for the whole Mediterranean region. The lack of direct control measures emphasizes the need for early surveillance and detection of the infection on a large scale.

This approach will support the official programme of NPPO in the rapid adoption of quarantine measures such as the eradication of infected trees for preventing the spread of the pathogen in virgin areas.

To this aim the activities conducted in this research have allowed for the evaluation of the potentiality of remote and proximal sensing in early surveillance and detection of *Xf*. In particular, the focus was on the use of tools, to assess plant health of the olive trees implicated in the disease, and to accurately identify the suspect trees.

Among the different methodologies, in the present work, assisted photointerpretation was employed as a rapid method for recognizing visible symptoms of OQDS in olive trees and a study was conducted on the analysis of spectral data for early detection of the infection in specific portions of the spectral signature so as to choose or define sensors (multispectral and/or hyperspectral) aimed at early detection of the presence of infection through the use of remote platforms (aircraft, UAV).

Given the state of emergency, assisted photointerpretation was the primary tool used to begin to provide information, in a short period of time about the potential presence of symptoms attributable to OQDS in areas still considered harmless. The use

of images sospatially defined and strengthened by the presence of the NIR band, greatly facilitated the identification of signs of OQDS starting from key photo types, well correlated to the expression of symptoms of the disease.

The recognition process in the outbreak in Trepuzzi has allowed for the classification of about 450 specimens of olive trees, specifying the class of gravity of OQDS.

Data analyses shows that 430 olive trees were classified in the map as trees with OQDS symptoms. Of these, in about 20% of the trees, the presence of Xf was confirmed with light to medium gravity; in the rest of the 80% of trees signs of deterioration caused by (i) diseases such as leprosy, (ii) shortage/lack of agronomic practices, (iii) burning by fire and, finally, (iv) the presence of dead trees, were detected. In all these cases the depletion of the foliage, detected through assisted photointerpretation, made it difficult to distinguish between trees that were symptomatic of OQDS and others.

The presence of *Xf* in the plants implicated in the study was supported by laboratory tests. Moreover, the comparison with the previous results of monitoring, showed that OQDS trees identified in the hotbed, coincided in part with plants that tested positive for the bacteria (especially for the high gravity condition).

Therefore, with the photointerpretation technique, olive growing areas that are healthy can be quickly distinguished from areas that are suspect or compromised, limiting to the latter, diagnostic investigations to be carried out in the field with visual surveys and any sampling and analysis in the laboratory.

In this way precise monitoring can be adopted by directing inspections at suspect sites on plants well-georeferenced, allowing for the optimization of both human and financial resources. Moreover, photointerpretation carried out on images acquired in different periods can: (i) provided for comprehensive picture on the evolution of space-time of incidence and severity of the infection and (ii) can provide information on plants to be sampled even in periods subsequent to pruning on the basis of symptoms photo interpreted before pruning.

Because of the state of emergency prevailing in the Phytosanitary Service of the Apulia Region, photointerpretation developed in this work has permitted the quick detection, remotely and from geolocation, photo of the types of trees that show symptoms of drying (OQDS-similar), there by becoming to valid tool in official monitoring of *Xf*. In

greatly facilitating additions to the work of the inspectors in the field, photointerpretation has supported the territorial epidemiological analysis on the infection for the implementation of appropriate containment measures.

The problem of dimensionality reduction of hyperspectral data, selecting a subset of bands from the signal, is a key aspect for efficiently and completely capitalizing on the potential of hyperspectral data.

An important aspect of this process is to maintain all the key wavelengths, that contain significant information without losing the discriminative power, downlining the statistical method which it processes (Thenkabail et al., 2012). Choosing the right data is therefore, an important aspect of Data Mining.

The research conducted in this direction has shown that the early symptoms of the *Xylella fastidiosa* infection can be detected using hyperspectral reflectance measurements acquired from olive leaves. The identification of the presence of the pathogen, however, cannot be generalized between the different varieties.

The bands used in this research were found to depend on the model and on the variety.

For the Leccino variety, two different methodologies selected bands that correlate in proximity in the blue area, the red and SWIR from a distance. In particular, the closest bands are found in the range comprised between 660 and 690 nm and 1720 and 1750 nm.

For the Cellina variety, the methods selected bands that overlapless, the heuristic method, HAM, identified a band in the red-edge and one in the immediate plateau (805 nm), while the proposed iPCA-ICV method was able to reveal all the water absorption levels (971-980 nm, 1191-1200 nm and 1381-1390 nm) further with in the SWIR spectrum.

The bands selected with both methods showed coherence with the level of infection detected in the asymptomatic leaves of the two varieties (19.1% in Leccino and 85.7% in Cellina di Nardò). A lower level of infection, probably, was expressed, from the spectral reflectance aspect, through disorders mainly in the visible region (at the turn of the peaks of chlorophyll absorption in the blue and red regions); levels of higher infection, instead, that manifested in the more susceptible variety, highlight a condition closer to the manifestation of the symptoms, expressed through increasedloss of water content of the

plant, together with structural changes of the leaf (SWIR region according to Thenkabail et al., 2014).

The iPCA-ICV method proposed, show edit was high-performing compared to the heuristic model selected from the literature (HAM), for both varieties, also for its characteristic of supplying an easily interpretable graphical representation and coherence with the development of the symptoms of the disease. With regards to the HAM model, unstructured with respect to the spectral signature (to minimize the complexity of the data), iPCA-ICV follows through values taken on from Calinski-Harabasz index or from Silhouette index the maximum difference between classes (distance), if it exists, through the natural separation between gruops in the spectrum (infected and non infected for every narrowband).

The separation value and that of the compactness of the indices showed to be different and directly proportional to the level of infection found in the two cultivars studied.

It would be an advantage to be able to select specific bands for two reasons: i) it helps to select the wavelength more correlated to the phenomena which the object of the study and ii) allows for the development strategies to select the wavelengths for the realization of specific sensors, such as those on board new generation satellites or UAV, in particular in the application of precision agriculture.

REFERENCES

ABDI, H. & WILLIAMS, L. J. 2010. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics,* 2**,** 433-459.

ADAMS, M. L., PHILPOT, W. D. & NORVELL, W. A. 1999. Yellowness index: An application of spectral second derivatives to estimate chlorosis of leaves in stressed vegetation. *International Journal of Remote Sensing,* 20**,** 3663-3675.

ALMEIDA, R. P. P., BLUA, M. J., LOPES, J. R. S. & PURCELL, A. H. 2005. Vector Transmission of Xylella fastidiosa: Applying Fundamental Knowledge to Generate Disease Management Strategies. *Annals of the Entomological Society of America,* 98**,** 775-786.

ALNAASAN, Y. 2015. Hyperspectral discrimination of fire blight infection in apple and pear, and molecular typing of some Mediterranean isolates of its causal agent Erwinia amylovora.

AMANIFAR, N., TAGHAVI, M., IZADPANAH, K. & BABAEI, G. 2014. Isolation and pathogenicity of Xylella fastidiosa from grapevine and almond in Iran. *Phytopathologia Mediterranea,* 53**,** 318.

ANDERSON, M. J. & ROBINSON, J. 2003. Generalized discriminant analysis based on distances. *Australian & New Zealand Journal of Statistics,* 45**,** 301-318.

ASNER, G. P. 1998. Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment,* 64**,** 234-253.

BAJWA, S. & KULKARNI, S. 2011. Hyperspectral data mining In: PS Thenkabail, JG Lyon, A. Huete, editors, Hyperspectral remote sensing of vegetation and agricultural crops. CRC Press, New York.

BALLABIO, D. & CONSONNI, V. 2013. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Analytical Methods,* 5**,** 3790-3798.

BALLESTEROS, R., ORTEGA, J. F., HERNÁNDEZ, D. & MORENO, M. A. 2014. Applications of georeferenced high-resolution images obtained with unmanned

aerial vehicles. Part I: Description of image acquisition and processing. *Precision Agriculture,* 15**,** 579-592.
BARET, F., HOULÈS, V. & GUÉRIF, M. 2007. Quantification of plant stress using remote sensing observations and crop models: the case of nitrogen management. *Journal of Experimental Botany,* 58**,** 869-880.
BARNES, J. D., BALAGUER, L., MANRIQUE, E., ELVIRA, S. & DAVISON, A. W. 1992. A reappraisal of the use of DMSO for the extraction and determination of chlorophylls a and b in lichens and higher plants. *Environmental and Experimental Botany,* 32**,** 85-100.
BAURIEGEL, E., GIEBEL, A., GEYER, M., SCHMIDT, U. & HERPPICH, W. B. 2011. Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Computers and Electronics in Agriculture,* 75**,** 304-312.
BERISHA, B., CHEN, Y., ZHANG, G., XU, B. & CHEN, T. 1998. Isolation of Peirce's disease bacteria from grapevines in Europe. *European Journal of Plant Pathology,* 104**,** 427-433.
BLACKBURN, G. A. 1998. Spectral indices for estimating photosynthetic pigment concentrations: A test using senescent tree leaves. *International Journal of Remote Sensing,* 19**,** 657-675.
BLACKBURN, G. A. 2007. Hyperspectral remote sensing of plant pigments. *Journal of Experimental Botany,* 58**,** 855-867.
BOCK, C. H., POOLE, G. H., PARKER, P. E. & GOTTWALD, T. R. 2010. Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. *Critical Reviews in Plant Sciences,* 29**,** 59-107.
BOLTON, D. K. & FRIEDL, M. A. 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology,* 173**,** 74-84.
BOSCIA, D., SAPONARI, M., PALMISANO, F., LOCONSOLE, G., MARTELLI, G. & SAVINO, V. Field observations on the behaviour of different olive cultivars in response to Xylella fastidiosa infections. Prooceedings "International Symposium on the European Outbreak of Xylella fastidiosa in Olive", Gallipoli–Locorotondo, Italy, 2014.
BRAVO, C., MOSHOU, D., WEST, J., MCCARTNEY, A. & RAMON, H. 2003. Early Disease Detection in Wheat Fields using Spectral Reflectance. *Biosystems Engineering,* 84**,** 137-145.
BREIMAN, L. 2006. randomForest: Breiman and Cutler's random forests for classification and regression. URL http://stat-www. berkeley. edu/users/breiman/RandomForests, R package version.
BRO, R. & SMILDE, A. K. 2014. Principal component analysis. *Analytical Methods,* 6**,** 2812-2831.

CALDERÓN, R., NAVAS-CORTÉS, J. & ZARCO-TEJADA, P. 2015. Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. *Remote Sensing,* 7**,** 5584.

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 Sensing of Environment,* 139**,** 231-245.

CALIŃSKI, T. & HARABASZ, J. 1974. A dendrite method for cluster analysis. *Communications in Statistics,* 3**,** 1-27.

CAMPBELL, J. B. & WYNNE, R. H. 2011. *Introduction to remote sensing*, Guilford Press.

CARIDDI, C., SAPONARI, M., BOSCIA, D., DE STRADIS, A., LOCONSOLE, G., NIGRO, F., PORCELLI, F., POTERE, O. & MARTELLI, G. 2014. Isolation of a Xylella fastidiosa strain infecting olive and oleander in Apulia, Italy. *Journal of Plant Pathology,* 96**,** 425-429.

CASAS, A., RIAÑO, D., USTIN, S. L., DENNISON, P. & SALAS, J. 2014. Estimation of water-related biochemical and biophysical vegetation properties using multitemporal airborne hyperspectral data and its comparison to MODIS spectral response. *Remote Sensing of Environment,* 148**,** 28-41.

CHEN, B., LI, S.-K., WANG, K.-R., WANG, J., WANG, F.-Y., XIAO, C.-H., LAI, J.-C. & WANG, N. 2008. Spectrum Characteristics of Cotton Canopy Infected with Verticillium Wilt and Applications. *Agricultural Sciences in China,* 7**,** 561-569.

CHEN, J. M. 1996. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing,* 22**,** 229-242.

CHO, M. A., SKIDMORE, A. K. & SOBHAN, I. 2009. Mapping beech (Fagus sylvatica L.) forest structure with airborne hyperspectral imagery. *International Journal of Applied Earth Observation and Geoinformation,* 11**,** 201-211.

COLWELL, R. N. 1956. *Determining the prevalence of certain cereal crop diseases by means of aerial photography*, University of California.

CURRAN, P. J. 1989. Remote sensing of foliar chemistry. *Remote Sensing of Environment,* 30**,** 271-278.

CURRAN, P. J., WINDHAM, W. R. & GHOLZ, H. L. 1995. Exploring the relationship between reflectance red edge and chlorophyll concentration in slash pine leaves. *Tree Physiology,* 15**,** 203-206.

DARVISHZADEH, R. Hyperspectral remote sensing of vegetation parameters using statistical and physical models. 2008. ITC.

DARVISHZADEH, R., SKIDMORE, A., SCHLERF, M., ATZBERGER, C., CORSI, F. & CHO, M. 2008. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS Journal of Photogrammetry and Remote Sensing,* 63**,** 409-426.

DAUGHTRY, C. S. T., HUNT JR, E. R. & MCMURTREY III, J. E. 2004. Assessing crop residue cover using shortwave infrared reflectance. *Remote Sensing of Environment,* 90**,** 126-134.

DAUGHTRY, C. S. T., WALTHALL, C. L., KIM, M. S., DE COLSTOUN, E. B. & MCMURTREY III, J. E. 2000. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment,* 74**,** 229-239.

DAVIES, D. L. & BOULDIN, D. W. 1979. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PAMI-1**,** 224-227.

DE GENNARO, B. & ROSELLI, L. 2013. La filiera olivicola-olearia pugliese: struttura, organizzazione e competitività. La rivista di scienza dell'alimentazione.

DE JONG, S. M. & VAN DER MEER, F. D. 2007. *Remote sensing image analysis: Including the spatial domain*, Springer Science & Business Media.

DELALIEUX, S., VAN AARDT, J., KEULEMANS, W., SCHREVENS, E. & COPPIN, P. 2007. Detection of biotic stress (Venturia inaequalis) in apple trees using hyperspectral data: Non-parametric statistical approaches and physiological implications. *European Journal of Agronomy,* 27**,** 130-143.

DELLA COLETTA-FILHO, H., FRANCISCO, C. S., LOPES, J. R. S., DE OLIVEIRA, A. F. & DA SILVA, L. F. D. O. 2016. First report of olive leaf scorch in Brazil, associated with Xylella fastidiosa subsp. pauca. *Phytopathologia mediterranea,* 55**,** 130.

DEMMIG-ADAMS, B. & ADAMS, W. W. 1996. The role of xanthophyll cycle carotenoids in the protection of photosynthesis. *Trends in Plant Science,* 1**,** 21-26.

DJELOUAH, K., FRASHERI, D., VALENTINI, F., D'ONGHIA, A. M. & DIGIARO, M. 2014. Direct tissue blot immunoassay for detection of Xylella fastidiosa in olive trees. *Phytopathologia Mediterranea,* 53**,** 559.

DUNN, J. C. 1973. A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters.

ERDLE, K., MISTELE, B. & SCHMIDHALTER, U. 2011. Comparison of active and passive spectral sensors in discriminating biomass parameters and nitrogen status in wheat cultivars. *Field Crops Research,* 124**,** 74-84.

FERERES, E., ORGAZ, F. & GONZALEZ-DUGO, V. 2011. Reflections on food security under water scarcity. *J Exp Bot,* 62**,** 4079-86.

FRANKE, J. & MENZ, G. 2007. Multi-temporal wheat disease detection by multi-spectral remote sensing. *Precision Agriculture,* 8**,** 161-172.

GAMON, J. A., PEÑUELAS, J. & FIELD, C. B. 1992. A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment,* 41**,** 35-44.

GITELSON, A. & MERZLYAK, M. N. 1994. Spectral Reflectance Changes Associated with Autumn Senescence of Aesculus hippocastanum L. and Acer platanoides L.

Leaves. Spectral Features and Relation to Chlorophyll Estimation. *Journal of Plant Physiology,* 143**,** 286-292.

GITELSON, A. A., MERZLYAK, M., ZUR, Y., STARK, R. & GRITZ, U. 2001. Non-destructive and remote sensing techniques for estimation of vegetation status.

GRISHAM, M. P., JOHNSON, R. M. & ZIMBA, P. V. 2010. Detecting Sugarcane yellow leaf virus infection in asymptomatic leaves with hyperspectral remote sensing and associated leaf pigment changes. *Journal of Virological Methods,* 167**,** 140-145.

GROSSMAN, Y. L., USTIN, S. L., JACQUEMOUD, S., SANDERSON, E. W., SCHMUCK, G. & VERDEBOUT, J. 1996. Critique of stepwise multiple linear regression for the extraction of leaf biochemistry information from leaf reflectance data. *Remote Sensing of Environment,* 56**,** 182-193.

GUALANO, S., SANTORO, F., DJELOUAH, K. & D'ONGHIA, A. M. PROXIMAL AND REMOTE SENSING IN THE MONITORING OF CITRUS TRISTEZA VIRUS (CTV) INFECTED TREES: PRELIMINARY RESULTS. 2012. International Society for Horticultural Science (ISHS), Leuven, Belgium, 641-646.

GUARIO, A., NIGRO, F., BOSCIA, D. & SAPONARI, M. 2013. Disseccamento rapido dell'olivo, cause e misure di contenimento. *L'Informatore Agrario,* 46**,** 51-54.

GUERSCHMAN, J. P., HILL, M. J., RENZULLO, L. J., BARRETT, D. J., MARKS, A. S. & BOTHA, E. J. 2009. Estimating fractional cover of photosynthetic vegetation, non-photosynthetic vegetation and bare soil in the Australian tropical savanna region upscaling the EO-1 Hyperion and MODIS sensors. *Remote Sensing of Environment,* 113**,** 928-945.

HAELTERMAN, R. M., TOLOCKA, P. A., ROCA, M., GUZMÁN, F. A., FERNÁNDEZ, F. D. & OTERO, M. L. 2015. First presumptive diagnosis of Xylella fastidiosa causing olive scorch in Argentina. *Journal of Plant Pathology,* 97.

HANSEN, P. M. & SCHJOERRING, J. K. 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sensing of Environment,* 86**,** 542-553.

HERWITZ, S. R., JOHNSON, L. F., DUNAGAN, S. E., HIGGINS, R. G., SULLIVAN, D. V., ZHENG, J., LOBITZ, B. M., LEUNG, J. G., GALLMEYER, B. A., AOYAGI, M., SLYE, R. E. & BRASS, J. A. 2004. Imaging from an unmanned aerial vehicle: agricultural surveillance and decision support. *Computers and Electronics in Agriculture,* 44**,** 49-61.

HILLNHÜTTER, C., MAHLEIN, A. K., SIKORA, R. A. & OERKE, E. C. 2011. Remote sensing to detect plant stress induced by Heterodera schachtii and Rhizoctonia solani in sugar beet fields. *Field Crops Research,* 122**,** 70-77.

HOPKINS, D. L. & PURCELL, A. H. 2002. Xylella fastidiosa: Cause of Pierce's Disease of Grapevine and Other Emergent Diseases. *Plant Disease,* 86**,** 1056-1066.

HUANG, J., LIAO, H., ZHU, Y., SUN, J., SUN, Q. & LIU, X. 2012. Hyperspectral detection of rice damaged by rice leaf folder (Cnaphalocrocis medinalis). *Computers and Electronics in Agriculture,* 82**,** 100-107.

HUANG, J. F. & BLACKBURN, G. A. 2011. Optimizing predictive models for leaf chlorophyll concentration based on continuous wavelet analysis of hyperspectral data. *International Journal of Remote Sensing,* 32**,** 9375-9396.

HUANG, W., LAMB, D. W., NIU, Z., ZHANG, Y., LIU, L. & WANG, J. 2007. Identification of yellow rust in wheat using in-situ spectral reflectance measurements and airborne hyperspectral imaging. *Precision Agriculture,* 8**,** 187-197.

HUNT, G. 1980. Electromagnetic Radiation: The Communication Lind in Remote Sensing [In: Remote Sensing in Geology;(Eds.): Siegal, B. S. and Gillespie, AR] John Wiley & sons. *New York,* 702.

JACQUEMOUD, S. & USTIN, S. L. Leaf optical properties: A state of the art. 8th International Symposium of Physical Measurements & Signatures in Remote Sensing, 2001. 223-332.

JAIN, N., RAY, S. S., SINGH, J. P. & PANIGRAHY, S. 2007. Use of hyperspectral data to assess the effects of different nitrogen applications on a potato crop. *Precision Agriculture,* 8**,** 225-239.

JANSE, J. D. & OBRADOVIC, A. 2010. XYLELLA FASTIDIOSA: ITS BIOLOGY, DIAGNOSIS, CONTROL AND RISKS. *Journal of Plant Pathology,* 92**,** S35-S48.

JONES, D. 2014. The Fathom Toolbox for MATLAB: software for multivariate ecological and oceanographic data analysis. College of Marine Science, University of South Florida, Tampa, FL, USA. *Petersburg, FL, USA*.

JONES, H. G. & VAUGHAN, R. A. 2010. *Remote sensing of vegetation: principles, techniques, and applications*, Oxford university press.

KOKALY, R. F., ASNER, G. P., OLLINGER, S. V., MARTIN, M. E. & WESSMAN, C. A. 2009. Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment,* 113, Supplement 1**,** S78-S91.

KUMAR, L., SCHMIDT, K., DURY, S. & SKIDMORE, A. 2001. Imaging Spectrometry and Vegetation Science. *In:* MEER, F. D. V. D. & JONG, S. M. D. (eds.) *Imaging Spectrometry: Basic Principles and Prospective Applications.* Dordrecht: Springer Netherlands.

LACHENBRUCH, P. A. & MICKEY, M. R. 1968. Estimation of error rates in discriminant analysis. *Technometrics,* 10**,** 1-11.

LAUDIEN, R., BARETH, G. & DOLUSCHITZ, R. Analysis of hyperspectral field data for detection of sugar beet diseases. Proceedings of the EFITA Conference, Debrecen, Hungary, 2003. 375381.

LE MAIRE, G., FRANÇOIS, C. & DUFRÊNE, E. 2004. Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment,* 89**,** 1-28.

LEARDI, R. & NØRGAARD, L. 2004. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of chemometrics,* 18**,** 486-497.

LEE, K. S., COHEN, W. B., KENNEDY, R. E., MAIERSPERGER, T. K. & GOWER, S. T. 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment,* 91**,** 508-520.

LEGENDRE, P. & LEGENDRE, L. 1998. Numerical ecology: second English edition. *Developments in environmental modelling,* 20.

LI, X., LEE, W. S., LI, M., EHSANI, R., MISHRA, A. R., YANG, C. & MANGAN, R. L. 2012. Spectral difference analysis and airborne imaging classification for citrus greening infected trees. *Computers and Electronics in Agriculture,* 83**,** 32-46.

LICHTENTHALER, H. K., LANG, M., SOWINSKA, M., HEISEL, F. & MIEHÉ, J. A. 1996. Detection of Vegetation Stress Via a New High Resolution Fluorescence Imaging System. *Journal of Plant Physiology,* 148**,** 599-612.

LOCONSOLE, G., POTERE, O., BOSCIA, D., ALTAMURA, G., DJELOUAH, K., ELBEAINO, T., FRASHERI, D., LORUSSO, D., PALMISANO, F. & POLLASTRO, P. 2014. Detection of Xylella fastidiosa in olive trees by molecular and serological methods. *Journal of Plant Pathology,* 96**,** 7-14.

LÓPEZ-ESCUDERO, F. J. & MERCADO-BLANCO, J. 2011. Verticillium wilt of olive: a case study to implement an integrated strategy to control a soil-borne pathogen. *Plant and Soil,* 344**,** 1-50.

MAHLEIN, A.-K. 2015. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease,* 100**,** 241-251.

MAHLEIN, A.-K., OERKE, E.-C., STEINER, U. & DEHNE, H.-W. 2012. Recent advances in sensing plant diseases for precision crop protection. *European Journal of Plant Pathology,* 133**,** 197-209.

MAHLEIN, A. K., RUMPF, T., WELKE, P., DEHNE, H. W., PLÜMER, L., STEINER, U. & OERKE, E. C. 2013. Development of spectral indices for detecting and identifying plant diseases. *Remote Sensing of Environment,* 128**,** 21-30.

MAIN, R., CHO, M. A., MATHIEU, R., O'KENNEDY, M. M., RAMOELO, A. & KOCH, S. 2011. An investigation into robust spectral indices for leaf chlorophyll estimation. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66**,** 751-761.

MAJEKE, B., VAN AARDT, J. A. N. & CHO, M. A. 2008. Imaging spectroscopy of foliar biochemistry in forestry environments. *Southern Forests: a Journal of Forest Science,* 70**,** 275-285.

MALTHUS, T. J. & MADEIRA, A. C. 1993. High resolution spectroradiometry: Spectral reflectance of field bean leaves infected by Botrytis fabae. *Remote Sensing of Environment,* 45**,** 107-116.

MALTHUS, T. J., YOUNGER, C. J., BIRCHWOODS, H. & LITTLEBURN, M. Remotely sensing stress in street trees using high spatial resolution data. second international geospatial information in agriculture and forestry conference, 2000.

MANZER, F. E. & COOPER, G. R. 1967. B646: Aerial Photographic Methods of Potato Disease Detection.

MARTELLI, G. P., BOSCIA, D., PORCELLI, F. & SAPONARI, M. 2016. The olive quick decline syndrome in south-east Italy: a threatening phytosanitary emergency. *European Journal of Plant Pathology,* 144**,** 235-243.

MECKLIN, C. J. & MUNDFROM, D. J. 2004. An appraisal and bibliography of tests for multivariate normality. *International Statistical Review,* 72**,** 123-138.

MERONI, M., ROSSINI, M. & COLOMBO, R. 2010. Characterization of leaf physiology using reflectance and fluorescence hyperspectral measurements. *Optical observation of vegetation properties and characteristics. Research Signpost, Trivandrum***,** 165-187.

MERZLYAK, M. N., GITELSON, A. A., CHIVKUNOVA, O. B. & RAKITIN, V. Y. U. 1999. Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum,* 106**,** 135-141.

MIRZAIE, M., DARVISHZADEH, R., SHAKIBA, A., MATKAN, A. A., ATZBERGER, C. & SKIDMORE, A. 2014. Comparative analysis of different uni- and multi-variate methods for estimation of vegetation water content using hyper-spectral measurements. *International Journal of Applied Earth Observation and Geoinformation,* 26**,** 1-11.

MOSHOU, D., BRAVO, C., WEST, J., WAHLEN, S., MCCARTNEY, A. & RAMON, H. 2004. Automatic detection of 'yellow rust' in wheat using reflectance measurements and neural networks. *Computers and Electronics in Agriculture,* 44**,** 173-188.

MOSHOU, D., GRAVALOS, I., BRAVO, D. K. C., OBERTI, R., WEST, J. S. & RAMON, H. 2012. Multisensor Fusion of Remote Sensing Data for Crop Disease Detection. *In:* THAKUR, J. K., SINGH, S. K., RAMANATHAN, A. L., PRASAD, M. B. K. & GOSSEL, W. (eds.) *Geospatial Techniques for Managing Environmental Resources.* Dordrecht: Springer Netherlands.

MUHAMMED, H. H. 2005. Hyperspectral Crop Reflectance Data for characterising and estimating Fungal Disease Severity in Wheat. *Biosystems Engineering,* 91**,** 9-20.

MUTANGA, O. & SKIDMORE, A. K. 2004. Narrowband vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing,* 25**,** 3999-4014.

NAIDU, R. A., PERRY, E. M., PIERCE, F. J. & MEKURIA, T. 2009. The potential of spectral reflectance technique for the detection of Grapevine leafroll-associated virus-3 in two red-berried wine grape cultivars. *Computers and Electronics in Agriculture,* 66**,** 38-45.

NAVARRO-CERRILLO, R. M., TRUJILLO, J., DE LA ORDEN, M. S. & HERNÁNDEZ-CLEMENTE, R. 2014. Hyperspectral and multispectral satellite sensors for mapping chlorophyll content in a Mediterranean Pinus sylvestris L. plantation. *International Journal of Applied Earth Observation and Geoinformation,* 26**,** 88-96.

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. *Applied and Environmental Microbiology,* 69**,** 7319-7327.

NIGRO, F., BOSCIA, D., ANTELMI, I. & IPPOLITO, A. 2013. Fungal species associated with a severe decline of olive in southern Italy. *Journal of Plant Pathology,* 95**,** 668.

NØRGAARD, L. 2005. iToolbox Manual.

NUNNEY, L., ORTIZ, B., RUSSELL, S. A., RUIZ SÁNCHEZ, R. & STOUTHAMER, R. 2014. The Complex Biogeography of the Plant Pathogen Xylella fastidiosa: Genetic Evidence of Introductions and Subspecific Introgression in Central America. *PLOS ONE,* 9**,** e112463.

NUTTER, F. W., VAN RIJ, N., EGGENBERGER, S. K. & HOLAH, N. 2010. Spatial and Temporal Dynamics of Plant Pathogens. *In:* OERKE, E.-C., GERHARDS, R., MENZ, G. & SIKORA, R. A. (eds.) *Precision Crop Protection - the Challenge and Use of Heterogeneity.* Dordrecht: Springer Netherlands.

OERKE, E. C. 2006. Crop losses to pests. *The Journal of Agricultural Science,* 144**,** 31-43.

PANIGADA, C., ROSSINI, M., MERONI, M., CILIA, C., BUSETTO, L., AMADUCCI, S., BOSCHETTI, M., COGLIATI, S., PICCHI, V., PINTO, F., MARCHESI, A. & COLOMBO, R. 2014. Fluorescence, PRI and canopy temperature for water stress detection in cereal crops. *International Journal of Applied Earth Observation and Geoinformation,* 30**,** 167-178.

PELOSI, F., CASTALDI, F. & CASA, R. Operational unmanned aerial vehicle assisted post-emergence herbicide patch spraying in maize: a field study. *Precision agriculture '15.*

PEN¯UELAS, J., FILELLA, I., LLORET, P., MUN¯OZ, F. & VILAJELIU, M. 1995. Reflectance assessment of mite effects on apple trees. *International Journal of Remote Sensing,* 16**,** 2727-2733.

PORCAR-CASTELL, A., TYYSTJÄRVI, E., ATHERTON, J., VAN DER TOL, C., FLEXAS, J., PFÜNDEL, E. E., MORENO, J., FRANKENBERG, C. & BERRY, J. A. 2014. Linking chlorophyll a fluorescence to photosynthesis for remote sensing

applications: mechanisms and challenges. *Journal of experimental botany*, eru191.

PU, R. & GONG, P. 2011. Hyperspectral remote sensing of vegetation bioparameters. *Advances in environmental remote sensing: Sensors, algorithms, and applications*, 101-142.

RAMOELO, A., SKIDMORE, A. K., CHO, M. A., MATHIEU, R., HEITKÖNIG, I. M. A., DUDENI-TLHONE, N., SCHLERF, M. & PRINS, H. H. T. 2013. Non-linear partial least square regression increases the estimation accuracy of grass nitrogen and phosphorus using in situ hyperspectral and environmental data. *ISPRS Journal of Photogrammetry and Remote Sensing,* 82**,** 27-40.

RAY, S. S., SINGH, J. & PANIGRAHY, S. 2010. Use of hyperspectral remote sensing data for crop stress detection: ground-based studies. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Science,* 38.

RENDÓN, E., ABUNDEZ, I., ARIZMENDI, A. & QUIROZ, E. 2011. Internal versus external cluster validation indexes. *International Journal of computers and communications,* 5**,** 27-34.

REYNIERS, M., WALVOORT, D. J. J. & DE BAARDEMAAKER, J. 2006. A linear model to predict with a multi-spectral radiometer the amount of nitrogen in winter wheat. *International Journal of Remote Sensing,* 27**,** 4159-4179.

REYNOLDS, G. J., WINDELS, C. E., MACRAE, I. V. & LAGUETTE, S. 2011. Remote Sensing for Assessing Rhizoctonia Crown and Root Rot Severity in Sugar Beet. *Plant Disease,* 96**,** 497-505.

RINALDI, M., CASTRIGNANÒ, A., DE BENEDETTO, D., SOLLITTO, D., RUGGIERI, S., GAROFALO, P., SANTORO, F., FIGORITO, B., GUALANO, S. & TAMBORRINO, R. 2015. Discrimination of tomato plants under different irrigation regimes: analysis of hyperspectral sensor data. *Environmetrics,* 26**,** 77-88.

RIVERA, J., VERRELST, J., DELEGIDO, J., VEROUSTRAETE, F. & MORENO, J. 2014. On the Semi-Automatic Retrieval of Biophysical Parameters Based on Spectral Index Optimization. *Remote Sensing,* 6**,** 4927.

ROUSE JR, J. W., HAAS, R., SCHELL, J. & DEERING, D. 1974. Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication,* 351**,** 309.

ROUSSEEUW, P. J. 1987. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics,* 20**,** 53-65.

RUMPF, T., MAHLEIN, A. K., STEINER, U., OERKE, E. C., DEHNE, H. W. & PLÜMER, L. 2010. Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance. *Computers and Electronics in Agriculture,* 74**,** 91-99.

SANKARAN, S., MISHRA, A., EHSANI, R. & DAVIS, C. 2010. A review of advanced techniques for detecting plant diseases. *Computers and Electronics in Agriculture,* 72**,** 1-13.

SAPONARI, M., BOSCIA, D., NIGRO, F. & MARTELLI, G. P. 2013. Identification of DNA sequences related to Xylella fastidiosa in oleander, almond and olive trees exhibiting leaf scorch symptoms in Apulia (southern Italy). *Journal of Plant Pathology,* 95**,** 668.

SAPONARI, M., LOCONSOLE, G., CORNARA, D., YOKOMI, R. K., DE STRADIS, A., BOSCIA, D., BOSCO, D., MARTELLI, G. P., KRUGNER, R. & PORCELLI, F. 2014. Infectivity and Transmission of Xylella fastidiosa by Philaenus spumarius (Hemiptera: Aphrophoridae) in Apulia, Italy. *Journal of Economic Entomology,* 107**,** 1316-1319.

SHI, T., WANG, J., LIU, H. & WU, G. 2015. Estimating leaf nitrogen concentration in heterogeneous crop plants from hyperspectral reflectance. *International Journal of Remote Sensing,* 36**,** 4652-4667.

SOMERS, B., COOLS, K., DELALIEUX, S., STUCKENS, J., VAN DER ZANDE, D., VERSTRAETEN, W. W. & COPPIN, P. 2009. Nonlinear Hyperspectral Mixture Analysis for tree cover estimates in orchards. *Remote Sensing of Environment,* 113**,** 1183-1193.

SPICCIARELLI, R. & ARPAIA, S. 1991. Photographic experiences with" false-colour" and" infrared" film for localization of some unhealthy conditions in a citrus grove of Basilicata. *Informatore Fitopatologico (Italy)*.

STEINER, U., BUERLING, K. & OERKE, E.-C. 2008. Sensor use in plant protection. *Gesunde Pflanzen,* 60**,** 131-141.

STELLACCI, A., CASTRIGNANÒ, A., TROCCOLI, A., BASSO, B. & BUTTAFUOCO, G. 2016. Selecting optimal hyperspectral bands to discriminate nitrogen status in durum wheat: a comparison of statistical approaches. *Environmental monitoring and assessment,* 188**,** 1-15.

STILWELL, A. R., HEIN, G. L., ZYGIELBAUM, A. I. & RUNDQUIST, D. C. 2013. Proximal sensing to detect symptoms associated with wheat curl mite-vectored viruses. *International Journal of Remote Sensing,* 34**,** 4951-4966.

THENKABAIL, P. S., ENCLONA, E. A., ASHTON, M. S. & VAN DER MEER, B. 2004. Accuracy assessments of hyperspectral waveband performance for vegetation analysis applications. *Remote Sensing of Environment,* 91**,** 354-376.

THENKABAIL, P. S., KNOX, J. W., OZDOGAN, M., GUMMA, M. K., CONGALTON, R. G., WU, Z., MILESI, C., FINKRAL, A., MARSHALL, M. & MARIOTTO, I. 2012. Assessing future risks to agricultural productivity, water resources and food security: how can remote sensing help? *Photogrammetric Engineering and Remote Sensing,* 78**,** 773-782.

THENKABAIL, P. S., MARIOTTO, I., GUMMA, M. K., MIDDLETON, E. M., LANDIS, D. R. & HUEMMRICH, K. F. 2013. Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 6**,** 427-439.

TORRES-SÁNCHEZ, J., PEÑA, J. M., DE CASTRO, A. I. & LÓPEZ-GRANADOS, F. 2014. Multi-temporal mapping of the vegetation fraction in early-season wheat fields using images from UAV. *Computers and Electronics in Agriculture,* 103**,** 104-113.

TUMBER, K. P., ALSTON, J. M. & FULLER, K. 2014. Pierce's disease costs California $104 million per year. *California Agriculture,* 68.

USTIN, S. L., GITELSON, A. A., JACQUEMOUD, S., SCHAEPMAN, M., ASNER, G. P., GAMON, J. A. & ZARCO-TEJADA, P. 2009. Retrieval of foliar information about plant pigment systems from high resolution spectroscopy. *Remote Sensing of Environment,* 113, Supplement 1**,** S67-S77.

VAN MECHELEN, I., BOCK, H.-H. & DE BOECK, P. 2004. Two-mode clustering methods: astructuredoverview. *Statistical methods in medical research,* 13**,** 363-394.

VAN WITTENBERGHE, S., VERRELST, J., RIVERA, J. P., ALONSO, L., MORENO, J. & SAMSON, R. 2014. Gaussian processes retrieval of leaf parameters from a multi-species reflectance, absorbance and fluorescence dataset. *Journal of Photochemistry and Photobiology B: Biology,* 134**,** 37-48.

VERRELST, J., ALONSO, L., CAMPS-VALLS, G., DELEGIDO, J. & MORENO, J. 2012. Retrieval of Vegetation Biophysical Parameters Using Gaussian Process Techniques. *IEEE Transactions on Geoscience and Remote Sensing,* 50**,** 1832-1843.

VON LUXBURG, U. 2007. A tutorial on spectral clustering. *Statistics and Computing,* 17**,** 395-416.

WANG, H., GUO, J. & MA, Z. 2012. Monitoring Wheat Stripe Rust Using Remote Sensing Technologies in China. *In:* LI, D. & CHEN, Y. (eds.) *Computer and Computing Technologies in Agriculture V: 5th IFIP TC 5/SIG 5.1 Conference, CCTA 2011, Beijing, China, October 29-31, 2011, Proceedings, Part III.* Berlin, Heidelberg: Springer Berlin Heidelberg.

WANG, K., WANG, B. & PENG, L. 2009. CVAP: Validation for Cluster Analyses. *Data Science Journal,* 8**,** 88-93.

WELLS, J. M., RAJU, B. C., HUNG, H.-Y., WEISBURG, W. G., MANDELCO-PAUL, L. & BRENNER, D. J. 1987. Xylella fastidiosa gen. nov., sp. nov: Gram-Negative, Xylem-Limited, Fastidious Plant Bacteria Related to Xanthomonas spp. *International Journal of Systematic and Evolutionary Microbiology,* 37**,** 136-143.

WEST, J. 2003. Effects of heat-stress on production in dairy cattle. *Journal of dairy science,* 86**,** 2131-2144.

WORNER, S., GEVREY, M., ESCHEN, R., KENIS, M., PAINI, D., SINGH, S., WATTS, M. & SUITER, K. 2013. Prioritizing the risk of plant pests by clustering methods; self-organising maps, k-means and hierarchical clustering. *NeoBiota,* 18**,** 83.

WU, D., FENG, L., ZHANG, C. & HE, Y. 2008. Early Detection of Botrytis cinerea on Eggplant Leaves Based on Visible and Near-Infrared Spectroscopy. 51.

YANG, C.-M., CHENG, C.-H. & CHEN, R.-K. 2007. Changes in Spectral Characteristics of Rice Canopy Infested with Brown Planthopper and Leaffolder. *Crop Science,* 47**,** 329-335.

YASEEN, T., DRAGO, S., VALENTINI, F., ELBEAINO, T., STAMPONE, G., DIGIARO, M., #039 & ONGHIA, A. M. 2015. *On-site detection of Xylella fastidiosa in host plants and in "spy insects" using the real-time loop-mediated isothermal amplification method*.

YUAN, L., HUANG, Y., LORAAMM, R. W., NIE, C., WANG, J. & ZHANG, J. 2014. Spectral analysis of winter wheat leaves for detection and differentiation of diseases and insects. *Field Crops Research,* 156**,** 199-207.

ZARCO-TEJADA, P. J., BERJÓN, A., LÓPEZ-LOZANO, R., MILLER, J. R., MARTÍN, P., CACHORRO, V., GONZÁLEZ, M. R. & DE FRUTOS, A. 2005. Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment,* 99**,** 271-287.

ZHANG, J. 2008. *Bacterial leaf scorch Xylella fastidiosa Wells et al. and its potential insect vectors in pin and red oaks in central New Jersey*, ProQuest.

ZHANG, M., QIN, Z., LIU, X. & USTIN, S. L. 2003. Detection of stress in tomatoes induced by late blight disease in California, USA, using hyperspectral remote sensing. *International Journal of Applied Earth Observation and Geoinformation,* 4**,** 295-310.

ZHAO, D., HUANG, L., LI, J. & QI, J. 2007. A comparative analysis of broadband and narrowband derived vegetation indices in predicting LAI and CCD of a cotton canopy. *ISPRS Journal of Photogrammetry and Remote Sensing,* 62**,** 25-33.

Abstract

The Olive Quick Decline caused by *Xylella fastidiosa* in Italy is seriously threatening olive trees.

In this study remote sensing was applied for the identification and early detection of infected trees which should be eradicated. The photo interpretation of aerial images of high geometrical resolution was used to recognize and classify about 20% symptomatic/infected trees. Moreover, the early identification of *Xylella* infection was achieved by using hyperspectral reflectance data of leaf samples from infected olive cvs. *Cellina di Nardò* (the most susceptible) and *Leccino* (the less susceptible). In order to locate the optimal bands that identify the infection from the earliest stages of development, a heuristic approach from literature (LLR2, PCA and Wilk's Lambda) and a new combined general purpose detection method (interval PCA Internal Clustering Validation, iPCA-ICV) were compared.

Using both methods, it was possible to distinguish leaves infected by *Xylella fastidiosa* and to select specific narrow bands. In particular, both methods agree with the VIS regions (close to blue and to red) and with the SWIR, as regions of the spectrum with greater emphasis on the discrimination of *Leccino* (451-490 nm, 671-680 nm, 1451-1460 nm, 1571-1580 nm, 1721-1730 nm), the variety that is less affected by the infection (23.1%), while, for *Cellina* (85.7% of positive findings), the two methods produced contrasting results. iPCA-ICV mainly identifies the water absorption bands around 970, 1200 and 1400 nm, compared to the reference method that identifies the bands around 705 and 805 nm, as determiners in the identification of *Xylella*. In particular, and though to a lesser extent, the proposed method identified discriminating bands in the vicinity of the water absorption peaks for the less susceptible variety (approxiamately at 1200 nm).

The best discriminatory power was obtained from iPCA-ICV for both varieties (error rates from 23.7% and from 22.02% respectively for cv. *Cellina di Nardò* and cv. *Leccino*), compared to the reference method (error rates equal to 42.47% and 22.02%).

The identification of critical regions of the spectrum, thus, represents the first logical step towards the development of robust stress indicators based on hyperspectral images. The techniques for band selection, furthermore, are extremely useful, not only to improve the power of predictive models, but also for the interpretation of data or the design of specific sensors (Pest Disease Detection).