Department of Agriculture, Forestry, Nature, Energy Science and Technology (DAFNE)

CIHEAM-Mediterranean Agronomic Institute of Bari (MAIB)

HYPERSPECTRAL DISCRIMINATION OF FIRE BLIGHT INFECTION IN APPLE AND PEAR, AND MOLECULAR TYPING OF SOME MEDITERRANEAN ISOLATES OF ITS CAUSAL AGENT *Erwinia amylovora*

A dissertation submitted in fulfillment of the requirement for the degree of

Doctor of Philosophy

Plant Protection -Cycle XXVII- S.S.D. AGR/12

by

Yaseen ALNAASAN

Coordinator and supervisor: Professor Leonardo VARVARO

Co-Supervisor: Anna Maria D'ONGHIA

Advisers:

Franco SANTORO

Franco VALENTINI

Alfredo FABI

29 May 2015

Summary

Fire blight, caused by the phytopathogenic enterobacterium *Erwinia amylovora*, is a serious disease of mainly apple and pear, and more than 200 commercially, ornamentally, and ecologically important rosaceous plant species grown in Italy and in other regions in the world. The lack of commercially effective control compounds, other than antibiotics which became prohibited almost worldwide, and the fast spread of fire blight make the disease prevention extremely important. Currently, the best fire blight management strategy is a combination of strict quarantine measurements, sanitation, cultural practices, chemical and biological sprays, and most importantly the vigilant regular monitoring. However, large-scale monitoring is difficult due to the huge number of trees to be inspected. *E. amylovora* is a genetically monomorphic pathogen, though its different stains have different pathogenicity, severity, and antibiotic resistance. Studying the genetic diversity of *E. amylovora* is therefore extremely important to track the infection sources, map the distribution of the severe strains, study the evolutionary, and improve resistant varieties against this quarantine pest.

The first aim in this study was to discriminate between healthy and fire blight-infected apple and pear plants using their hyperspectral reflectance, and to develop classification models capable of detecting the infection even in the asymptomatic plants. Multivariate analysis showed that hyperspectral data can effectively discriminate between healthy and infected (symptomatic and asymptomatic) apple and pear. Two classification models, Partial Least Squares-Discriminant Analysis (PLS-DA) and Artificial Neural Networks (ANNs), were developed in this study. These non-parametric models successfully classified the studied plants into healthy and fire blight-infected with high accuracy not only in the glasshouse but also in the open field.

The second aim in this dissertation was to genotype different strains of *E. amylovora*, collected from apple and pear in different regions mainly Mediterranean, using an updated set of polymorphic Variable Number of Tandem Repeats (VNTR) loci. Three new polymorphic loci were identified in the current study and added to other six loci, defined previously, to establish a set of nine VNTR loci to be used in a Multiple-Locus VNTR Analysis (MLVA). The improved MLVA approach recognized 38 haplotypes and seven Clonal Complexes from the studied strains, and demonstrated the high genetic diversity among the Mediterranean strains and among the Italian strains in particular suggesting the multiple introduction events of the disease.

Keywords: *Erwinia amylovora*, Hyperspectral remote sensing, non-parametric classification models, Spectral Vegetation Indices, Partial Least Squares-Discriminant Analysis, Artificial Neural Networks, Variable Number of Tandem Repeats, Multiple-Locus VNTR Analysis.

Riassunto

Il colpo di fuoco, causato dall'enterobatterio fitopatogeno *Erwinia amylovora*, è una seria malattia principalmente di pero e melo, e di più di 200 specie di piante rosacee importanti dal punto di vista commerciale, ornamentale, ed ecologico, che crescono in Italia ed in altre regioni nel mondo. La mancanza sul mercato di formulati di controllo efficaci, diversi dagli antibiotici che sono diventati proibiti quasi in tutto il mondo, e la veloce diffusione del colpo di fuoco rende la prevenzione della malattia estremamente importante. Attualmente, la miglior strategia di gestione del colpo di fuoco è una combinazione di severe misure di quarantena, sanitizzazione, pratiche colturali, spray chimici e biologici, e maggiormente importante il regolare monitoraggio. Tuttavia, un monitoraggio su larga scala è difficile a causa dell'enorme numero di alberi da ispezionare. *E. amylovora* è un patogeno geneticamente monomorfico, sebbene differenti ceppi mostrino differente patogenicità, severità e resistenza agli antibiotici. Studiare la diversità genetica di *E. amylovora* è quindi estremamente importante per tracciare i punti di origine dell'infezione, mappare la distribuzione di ceppi dannosi, studiare l'evoluzione e migliorare le varietà resistenti verso questo patogeno da quarantena.

Il primo obiettivo in questo studio è stato quello di discriminare piante di melo e pero fra sane e colpite da colpo di fuoco utilizzando la loro riflettanza iperspettrale, e sviluppare modelli di classificazione capaci di rilevare l'infezione anche in piante asintomatiche. L'analisi multivariata ha mostrato che i dati iperspettrali possono effettivamente discriminare tra meli e peri sani ed infetti (sintomatici ed asintomatici). Due modelli di classificazione, Partial Least Squares-Discriminant Analysis (PLS-DA) e Artificial Neural Networks (ANNs), sono stati sviluppati in questo studio. Questi modelli non parametrici hanno classificato con successo le piante studiate in sane e colpite da fuoco batterico con elevata accuratezza non soltanto in serra ma anche in pieno campo.

Il secondo obiettivo in questa tesi è stato di genotipizzare differenti ceppi di *E. amylovora*, isolate d melo e pero in diverse regioni principalmente Mediterranee, utilizzando un set aggiornato di loci Variable Number of Tandem Repeats (VNTR) polimorfiche. Tre nuovi loci polimorfici sono stati identificati nel presente studio ed aggiunti ad altri sei loci, precedentemente definiti, per stabilire un insieme di nove loci VNTR da usare in una Multiple-Locus VNTR Analysis (MLVA). Il migliorato approccio MLVA ha riconosciuto 38 aplotipi e sette complessi clonali per i ceppi studiati, ed ha dimostrato l'elevata diversità genetica tra gli isolati Mediterranei e quelli Italiani che in particolare suggeriscono eventi multipli di introduzione della malattia.

Acknowledgements

I wish to thank, first of all, the International Center for Advanced Mediterranean Agronomic Studies (CIHEAM) represented by the Mediterranean Agronomic Institute of Bari (MAIB) and Tuscia University represented by the Department of Agriculture, Forestry, Nature, and Energy Science and Technology (DAFNE) for giving me this precious opportunity and for their scientific efforts and contributions.

This dissertation would have remained a dream without the help and the support of my supervisors and advisers. Professor Leonardo VARVARO, the former director of the DAFNE department, for being my friend, guide, and philosopher. Anna Maria D'ONGHIA, the head of the Integrated Pest Management department in MAIB, for her continuous help and wise leadership. Franco SANTORO, Franco VALENTINI, and Alfredo FABI for their real help in developing the hyperspectral analysis part of the dissertation and for their encouragements. Angelo MAZZAGLIA, Giorgio M. BALESTRA, and their team for without their help in developing the molecular part I would have been lost.

My deepest gratitude to Zurich University for Applied Science (ZHAW) represented by Fabio REZZONICO, University of Belgrade represented by Professor Aleksa OBRADOVIC, and Emilia-Romagna Regional Phytosanitary Services represented by Alessandra CALZOLARI for providing us with the required DNA material and for the field support. Many thanks to Prof. Ines DELFINO from the Department of Ecological science and Biology (DEB) in Tuscia University and Davide BALLABIO from the department of Earth and Environmental Science (DISAT) in University of Milano-Bicocca for their statistical advice. Big thanks also to Mr. Angelo FABBRI and Mr. Michele MANTOVANI the owners of studied orchards for their help and support.

Sincere thanks to my friends Doriano VITTORI, Maria C. TARATUFOLO, Peter MOUBARAK, Giuseppe CAVALLO, Serena CIARRONI, Clelia TOSI, Rocco CARACCIOLO, Marsilio RENZI, Raied ABOU KUBAA, Farid AMMOURI, and Vincenzo TAGLIAVENTO for their precious help, continuous encouragement, and friendship.

Dedication

To the true angels in my life; my parents, Ibrahim and Ilham, brothers and sisters, and my dear wife Diana who have always stood by me throughout my PhD years;

To my lovely friends and colleagues everywhere;

To every researcher who works sincerely to contribute in the effort that makes our plant a better place to live in;

To everyone who struggles and sacrifices to live in freedom, dignity, and peace!

Table of content

Summary I
Riassunto II
List of figures VII
List of tables X
List of abbreviations XI
Introduction 1
1.1. Fire blight and its causal agent 2
1.2. Host range 8
1.3. Symptoms and signs 9
1.4. Fire blight monitoring 10
1.5. Fire blight detection and management 12
1.6. Epidemiology 15
1.7 Historical review of *Erwinia amylovora* 15
Chapter 2 18
Abstract 19
2.1. Introduction 20
2.1.1. Precision Agriculture 20
2.1.2. Spectral signature and Spectral Vegetation Indices (SVIs) 24
2.2. Materials and methods 28
2.2.1. Glasshouse trials 28
2.2.2. Field trials 32
2.2.3. Data analysis 37
2.3. Results and discussions 48
2.3.1. Artificial inoculation and laboratory tests 48
2.3.2. Hyperspectral discrimination 50
2.3.3. Classification models 60
2.4. Conclusions 81
Chapter 3 83
Abstract 84

3.1. Introduction 85
3.1.1. Homogeneity and molecular typing in *Erwinia amylovora* 85
3.1.2. Multiple-Locus Variable Number of Tandem Repeats Analysis (MLVA) 86
3.2. Materials and methods 88
3.2.1. Bacterial strains and DNA quantification 88
3.2.2. VNTR loci mining and primer designing 90
3.2.3. PCR amplification and capillary electrophoresis 90
3.2.4. MLVA analysis 91
3.3. Results and discussions 92
3.3.1. VNTR loci selection and MLVA optimization 92
3.3.2. VNTRs characteristics 98
3.3.3. MLVA genotyping 101
3.4. Conclusions 106
References 107
Annexes 121
Annex I 121
Annex II 124
Annex III 127
Annex IV 130

List of figures

Figure 1 Apple (upper) and pear (lower) production in Italy in thousand tonnes in the period between 2000 and 2012 (FAO, 2015). 1
Figure 2 The distribution of fire blight disease in Europe, the Middle East, and North Africa (EPPO, 2014). 2
Figure 3 Two pictures demonstrate the fast and the aggressive manner of the fire blight infection in a pear orchard in Barletta, Puglia in which the upper picture (A) was taken in 19 June 2013 and shows that severely infected orchard, and the lower picture (B) was taken 15-days after and shows the pruned dead trees before eradication. 3
Figure 4 A schematic diagram of *Erwinia amylovora* detailed life cycle adapted from Thomson (2000). The dashed arrows represent the movement of the bacteria inside the plants and the solid ones represent the movement outside the plants. 7
Figure 5 The six principal components of the Precision Crop Protection. 20
Figure 6 A schematic representations showing the main elements in satellite remote sensing process 22
Figure 7 A graphical representation of an electromagnetic wave showing the mutually perpendicular electric and magnetic fields, the orientation of the wave, and the characteristic wavelength λ. 23
Figure 8 (A) A representation of the three possible pathways (transmission, absorption, and reflection) of incident light energy on vegetation, and (B) a reprinted typical spectral reflectance curve (spectral signature) that is controlled by leaf pigments, cellular structure and water content of the vegetation (Gausman, 1977). 25
Figure 9 The location of the study area for the glasshouse trials on the Italian map (left), and an aerial view of the quarantine glasshouse at MAIB (right). 28
Figure 10 The artificial inoculation methods used in the glasshouse trials: (A) direct injection by the bacterial suspension using a hypodermic needle and (B) cutting wounding by a sterile blade dipped in the bacterial suspension. 30
Figure 11 The hyperspectral acquisition system used for leaf reflectance measurements. (A) ASD leaf-clip supplied with a halogen bulb, (B) optical fiber, (C) ASD FieldSpec®3 spectroradiometer, and (D) Instrument controller. 31
Figure 12 The location of the study area on the Italian map (left), and aerial view for the studied fields (right): (A) a pear orchard in Copparo, Emilia-Romagna, and (B) an apple orchard in Formignana, Emilia-Romagna. 32
Figure 13 The hyperspectral acquisition system with the supporting system used for canopy reflectance measurements. This system comprised: (A) ASD fore-optic pistol grip, (B) aluminum tripod, (C) optical fiber, (D) ASD FieldSpec® HandHeld2 spectroradiometer, and (E) instrument controller. 34
Figure 14 The special designed tripod that carry the pistol grip to mount the fore-optic above the trees during canopy reflectance measurements. 35

Figure 15 The anti-hail net above the studied apple orchard that hampered the acquisition of the reflectance data from the canopies. 36
Figure 16 The general workflow of the performed research which mainly consists of three parts: I) data acquisitions and preprocessing, II) test hyperspectral separability, and III) developing PLS-DA and ANNs classification models. 38
Figure 17 A box and whisker plot for 10 Spectral Vegetation Indices (SVIs) on original (a) and standardized data (b); the upper and the lower lines in the large boxes are the 75th and 25th percentiles of the samples, respectively. The small boxes inside are the medians of the samples, and the extending lines from the boxes are called whiskers, and the values beyond those whiskers are detected as potential outliers. 41
Figure 18 A Receiver Operating Characteristic (ROC) curve created by plotting true positive (sensitivity), and false positive rates (1 - specificity) of an example of a classification model at different thresholds. 45
Figure 19 Fire blight symptoms and signs resulted from the artificial inoculation in apple and pear seedlings in the first glasshouse trial. These symptoms and signs include: (A) shepherd's crook and wilted leaves, (B) tip wilting, (C) bacterial ooze, and (D) blackened areas on the leaves. 49
Figure 20 Average spectral curves of all compiled leaf reflectances, in the range 400–1800 nm, from healthy and *E. amylovora*-inoculated apple and pear in the first glasshouse trial. The rectangular window in the left side of the plot shows a scoop of different reflectances around the green peak. 51
Figure 21 The scatterplots of the first two Principal Components (factors) scores resulted from applying PCCA on healthy (green points) and fire blight-inoculated (red points) apple (left plot) and pear (right plot) in the first glasshouse trial. 53
Figure 22 The scatterplot of the first two Principal Components (factors) scores resulted from applying PCCA on healthy (green points) and fire blight-inoculated pear (red points) in the second glasshouse trial. 54
Figure 23 Average curves of all compiled leaf and canopy reflectances, in the range between 400 nm and 1070 nm, from healthy-looking and fire blight-symptomatic apple and pear in the open field trial. 56
Figure 24 The scatterplots of the first two Principal Components (factors) scores for healthy-looking (green points) and fire blight-symptomatic (red points) apple (left plot) and pear (right plot) in leaf field trial. 58
Figure 25 The scatterplot of the first two Principal Components (factors) scores for healthy-looking (green points) and fire blight-symptomatic pear trees (red points) in canopy field trial. 59
Figure 26 The optimal combination of the spectral intervals selected by synergy interval-PLS to be used for developing PLS-DA and ANNs models in apple glasshouse trial. The selected intervals (green ranges) are 450–500, 500–550, 1400–1450, 1600–1650, and 1750–1800 nm. .. 61
Figure 27 The error rates (red points), and the ratios of Not Assigned samples (black points) in the PLS-DA optimization process of apple glasshouse trial. The venetian blinds procedure,

integrated in the classification toolbox for MATLAB, was applied to 10 and five cross validation groups (left and right parts of the figure, respectively) as a function of the number of latent variables. 63
Figure 28 The calculated responses for the infected class with the healthy (green points) and the infected samples (red points) of the training dataset in apple glasshouse trial. The dotted redline indicates the infected class threshold (0.61). 64
Figure 29 The Receiver Operating Characteristic (ROC) curves (upper), and line plots of sensitivity and specificity values (lower), blue and red respectively, as a function of classification threshold changing in the PLS-DA model of apple glasshouse trial. 66
Figure 30 The scores of the healthy and the infected plants on the first two latent variables, which explained 100% of the total variance, in the PLS-DA model of apple glasshouse trial. ... 67
Figure 31 A bubble plot obtained by the Kohonen and CP-ANN toolbox for MATLAB (version 3.6), shows the optimized ANNs networks for RWs in the glasshouse trial of apple. The bubble in the upper right part of the plot, the red bubble of 10X10 neurons and 75 epochs was the best as it had high relative frequency and predictive performance. 69
Figure 32 The scores of the healthy and the infected plants (A) and the loading plots of the tested SVIs (B) of the first two latent variables, which explained 64.58% of the total variance, in the PLS-DA model of apple glasshouse trial. 73
Figure 33 Analysis the DNA of two *Erwinia amylovora* strains with the VNTR locus Ea41 by QIAxcel multicapillary electrophoresis system (QIAGEN, USA); the green bands in the gel images (the gray areas) represent the two bands (15 and 600 bp) of the alignment marker. 96
Figure 34 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains and organized by their origin. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci..... 102
Figure 35 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains and organized by their year of isolation. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci. 104
Figure 36 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains organized by their host plants. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci..... 105

List of tables

Table 1 Risk levels and categories of orchard infection history according to *CougarBlight*. 11
Table 2 The 30 Spectral Vegetation Indices (SVIs), reported to be correlated to plant biotic and abiotic stress, which were calculated, standardized, and used in this research. 40
Table 3 The divided datasets in each apple and pear trial into training (about 75% of total number of samples) and evaluation subsets (about 25% of the total number of samples) to be used for models calibration and performance assessment, respectively. 43
Table 4 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by RWs PLS-DA model in training and evaluation subsets (with five groups split in venetian blinds) of apple glasshouse trial. 68
Table 5 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by RWs ANNs model in training and evaluation subsets (with five groups split in venetian blinds) of apple glasshouse trial. 70
Table 6 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA and ANNs classification models for SVIs variables in apple glasshouse trial. 71
Table 7 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA and ANNs classification models for RWs and SVIs variables in apple field trial. 76
Table 8 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA, and ANNs classification models for RWs and SVIs variables in pear glasshouse trials. 78
Table 9 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA, and ANNs classification models for RWs and SVIs variables in pear field trials 80
Table 10 A list of the DNA from *Erwinia amylovora* strains which are collected used in this study. 89
Table 11 The selected VNTR loci in the *Erwinia amylovora* CFBP1430 genome sequence by the Tandem Repeats Finder program. 92
Table 12 List of designed primer pairs for the selected VNTR loci which were revealed from the European reference strain CFBP1430 by the TRF program. 94
Table 13 The amplicon sizes of the selected set of the VNTR loci in the studied list of *Erwinia amylovora* strains. 97
Table 14 The list of the studied *Erwinia amylovora* strains and their MLVA profile. 99
Table 15 Characteristics of the selected set of VNTR loci in which D: Simpson's diversity index. VNTRs Ea55 and Ea56 are on the plasmid pEA29 and the other VNTRs are on the chromosome. 100

List of abbreviations

%	Percentage
~	Approximately
λ	Wavelength
°C	Degree Celsius
a.m.	Ante Meridiem
ANNs	Artificial Neural Networks
ATCC	American Type Culture Collection
BGI2	Blue Green Index 2
bp	Base pairs
BRDF	Bidirectional Reflectance Distribution Function
CC	Clonal Complex
CFBP	Collection Francaise de Bacteries Phytopathogenes
CFU	Colony-Forming Unit
cm	Centimeter
CP-ANN	Counter Propagation-Artificial Neural Network
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeat
cv.	Cultivar
D.M.	Ministerial Decree
DA	Discriminant Analysis
DNA	Deoxyribonucleic Acid
DPI	Days Post-Inoculation
EMR	Electromagnetic Radiation
EPPO	European and Mediterranean Plant Protection Organization
ER	Error Rate
f	Frequency
FAO	Food and Agriculture Organization
FOV	Field of View
FPR	False Positive Rate
FR	Fluorescence Ration
GA	Genetic Algorithm
GI	Greenness Index
GIS	Geographical Information System
GPS	Global Positioning System
km	Kilometer
Kb	Kilobase
LAMP	Loop-Mediated Isothermal Amplification
LVs	Latent Variables
m	Meter
Mb	Mega base pairs
MCARI	Modified Chlorophyll (a and b) Absorption in Reflectance Index
min	Minute
mL	Milliliter
MLVA	Multiple-Locus VNTR Analysis
mm	Millimeter
MST	Minimum Spanning Tree
ms	Millisecond
NA	Not Assigned samples
NDVI	Normalized Difference Vegetation Index
NER	Non-Error Rate
ng	Nanogram
NGS	Next Generation Sequence
NIR	Near-Infrared
NIRs	Near-Infrared Spectroscopy
nm	Nanometer
NPPO	National Plant Protection Organization
NPQI	Normalized Plant
p.m	Post Meridiem
PA	Precision Agriculture
PBI	Plant Biochemical Index
PC	Personal Computer
PCA	Principal Components Analysis
PCCA	Principal Components and Classification Analysis
PCP	Precision Crop Protection
PCR	Polymerase Chain Reaction
PLS	Partial Least Squares
PLS-DA	Partial Least Squares-Discriminant Analysis
pmol	Picomole
PRI	Photochemical Reflectance Index
PSRI	Plant Senescence Reflectance Index
PWI	Plant Water Index
rDNA	ribosomal DNA
REP	Red Edge Position
RMSE	Root Mean Square Error
RMSE CV	Root Mean Square Error of Cross Validation
RNA	Ribonucleic Acid
ROC	Receiver Operating Characteristic
rRNA	ribosomal RNA
RVSI	Red Edge Vegetation Stress Index
RWs	Reflectance Wavebands
s	Second
SDW	Sterilized Distilled Water
Sn	Sensitivity

SNP	Single Nucleotide Polymorphism
Sp	Specificity
spp.	Species
SR	Simple Ratio
SSR	Simple Sequence Repeat
SVIs	Spectral Vegetation Indices
SWIR	Short Wave Infrared
TPR	True Positive Rate
TR	Tandem Repeat
TRF	Tandem Repeats Finder
VNTR	Variable Number of Tandem Repeat
WGS	Whole-Genome Sequence
YI	Yellowness Index

Introduction

Apples (*Malus* spp.) and pears (*Pyrus* spp.), which are classified in *Maloideae* subfamily (family *Rosaceae*), and several other Maloid taxa have long been cultivated and valued for their distinctive fruit. Although the pomes of many Maloid genera are account for the economic importance and worldwide cultivation, several plants of this group are highly prized as garden ornamentals and have become a significant component of the nursery and landscape industry. Yielding more than 76.3 million tons and 23.5 million tons, apples and pears were listed among the top important fruit crops worldwide in 2012 (FAO, 2015). Italy is ranked among the major producing countries for these crops and the largest producer in Europe. However, apple and pear production fluctuated heavily in Italy due to the infection by some epiphytotic diseases in the last decades (Figure 1).

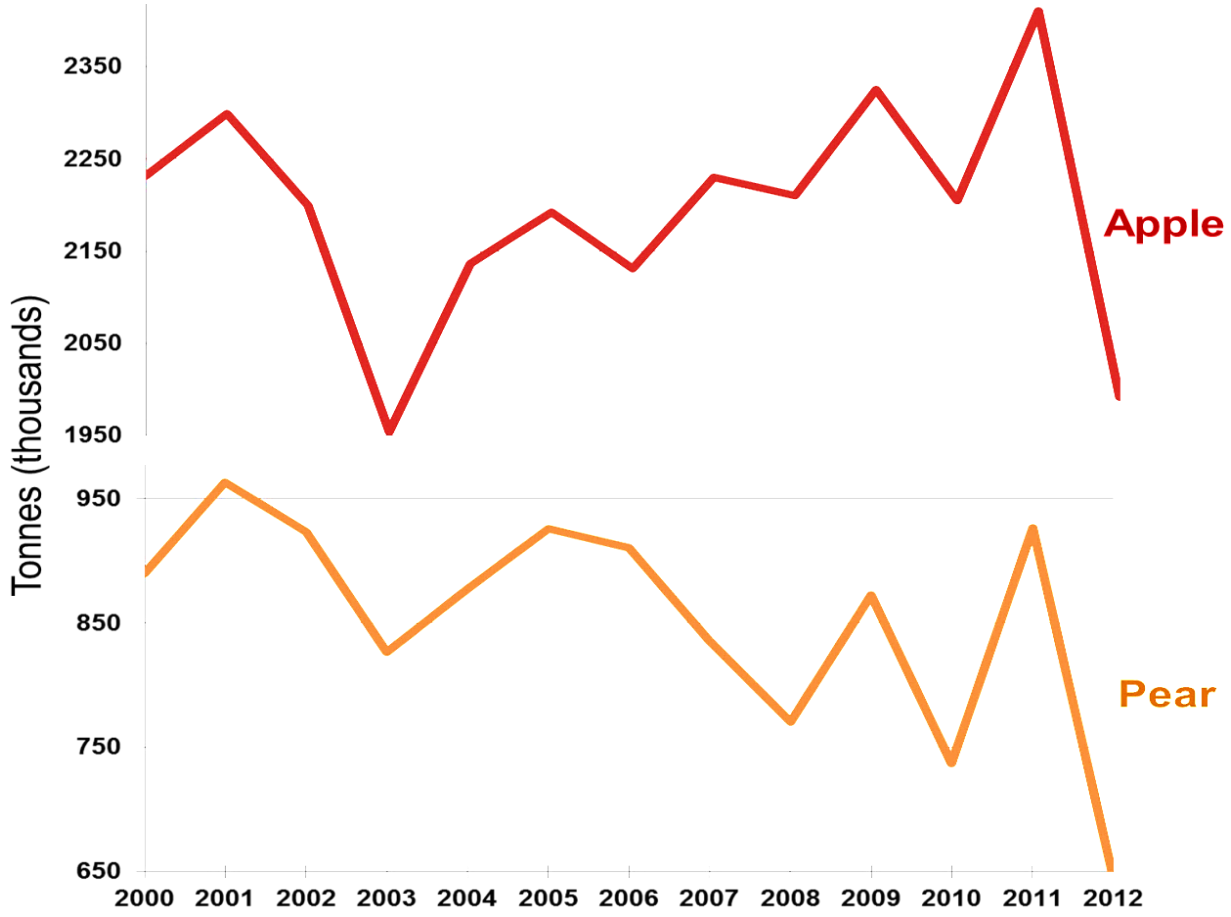


Figure 1 Apple (upper) and pear (lower) production in Italy in thousand tonnes in the period between 2000 and 2012 (FAO, 2015).

1.1. Fire blight and its causal agent

Fire blight is the most serious bacterial disease on apple and pear, its causal agent [*Erwinia amylovora* (Burrill) Winslow *et al*., 1920] is one of the most important 10 plant bacterial pathogens ever (Mansfield *et al.*, 2012). The economic impact of this disease is hard to determine, since low losses of a few flower clusters are negligible and a single outbreak can decimate orchards and subsequently stop the production for many years (Vanneste, 2000). Fire blight has caused several heavy losses around the world during the last decades. Between 1994 and 1999, a large number of outbreaks swept through Emilia-Romagna, Italy, and as a result more than 500000 fruit trees were rooted out (Calzolari *et al.*, 1999). The importance of fire blight is increasing steadily because of its continuous spread across the continents to new regions where no infection was reported before, like Australia and Finland. Moreover, the fast spread of fire blight in a destructive manner wherein course of weeks it is capable of destroying whole orchards (Figure 2). Furthermore, no pesticide can effectively control the disease.

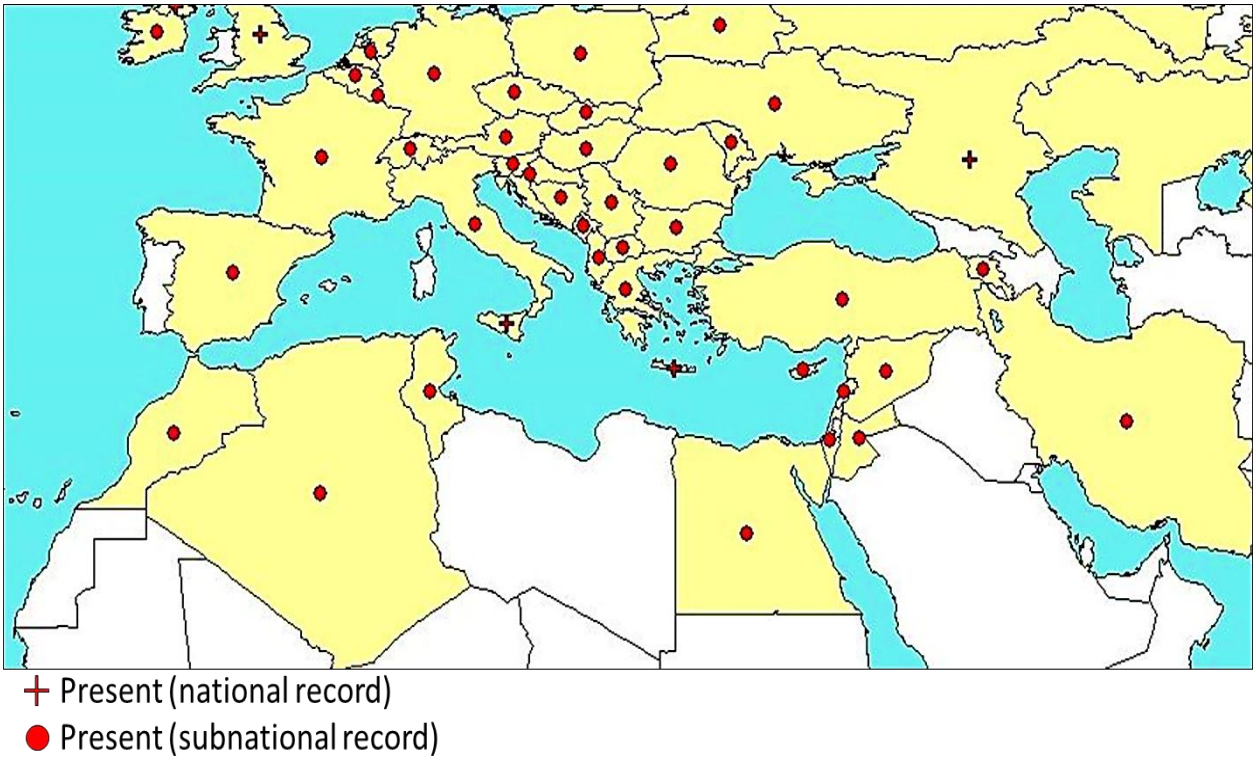


Figure 2 The distribution of fire blight disease in Europe, the Middle East, and North Africa (EPPO, 2014).

E. amylovora was recognized as the first bacterial plant pathogen by fulfilling Koch's postulates in 1884. Although *E. amylovora* has been reported on European and Mediterranean Plant Protection Organization (EPPO) A2 list (pests which are locally present in the EPPO regions) as a harmful quarantine pest since 1975, fire blight is present in all the European countries and almost in all the countries around the Mediterranean basin (Figure 3).

Figure 3 Two pictures demonstrate the fast and the aggressive manner of the fire blight infection in a pear orchard in Barletta, Puglia in which the upper picture (A) was taken in 19 June 2013 and shows that severely infected orchard, and the lower picture (B) was taken 15-days after and shows the pruned dead trees before eradication.

1.1.1 *Erwinia* species isolated from pome fruit trees

The genus *Erwinia*, named by the bacteriologist Erwin Frink Smith, has been reviewed several times. Many gram-negative, motile, aerobic-facultative, anaerobic, and non sporulated bacteria have been arbitrarily classified in the genus *Erwinia* on the basis of their ecological association with plants (Paulin, 2000). Recently, several studies have suggested the use of 16s rDNA sequences approach for identifying and resolving the taxonomic relationships among different species and groups of *Erwinia*. As a result, it was proposed that the genus *Erwinia* should be divided into four new genera: *Erwinia*, *Pectobacterium*, *Pantoea*, and *Brenneria* (Hauben *et al.*, 1998; Kim *et al.*, 1999). According to Hauben *et al.* (1998) the genus *Erwinia* comprises six species, *E. amylovora* (the type species of this genus), *E. mallotivora*, *E. persicina*, *E. psidii*, *E. rhapontici*, and *E. tracheiphila*; the genus *Pectobacterium* consists of four subspecies: *E. carotovora* (renamed as *Pectobacterium carotovorum*), *E. chrysanthemi* (renamed as *Dickeya dadantii*), *E. cacticida*, and *E. cypripedii* (renamed as *Pectobacterium cypripedii*); the genus *Brenneria* consists of six reclassified species: *E. alni*, *E. nigrifluens*, *E. paradisiaca*, *E. quercina*, *E. rubrifaciens*, and *E. salicis*. Finally, five species reclassified in the genus *Pantoea*: *E. ananas*, *E. herbicola*, *E. milletiae*, *E. stewartii*, and *E. uredovora*.

1.1.2 Pathogenic *Erwinia* species

1.1.2.1 *Erwinia pyrifoliae*

Kim *et al.* (1999) described this bacterium for the first time. It was also reported from Korea and Japan on pear trees (*P. pyrifoliae* and *P. communis*) with similar symptoms to fire blight. The pathogen was classified as a different species from *E. amylovora* by DNA hybridization assays (Kim *et al.*, 2001).

1.2.1.2 *Erwinia piriflorinigrans*

E. pirifloringrans has been reported only in Valencia, Spain until now (Roselló *et al.*, 2002; López *et al.*, 2011). It develops fire blight symptoms on blossoms solely. 16S rRNA sequence studies demonstrate high genetic similarity to *E. amylovora*, *E. pyrifoliae*, *E. tasmaniensis*, and *E. billingiae*. On the other hand, DNA-DNA hybridization resulted in classifying this bacterium as a new species (López *et al.*, 2011).

1.2.1.3 *Erwinia uzenensis*

This pathogen was reported only on European pear trees in Japan with similar symptoms to those of *E. amylovora*. G-C content of DNA, the fatty acid profile, and phenotypic characteristics *E. uzenensis* resembled those reported previously for members of the genus *Erwinia*. Nevertheless, DNA-DNA hybridization confirmed that the pathogen represents a novel species (Matsuura *et al.*, 2012).

1.2.2 Non-Pathogenic *Erwinia* species

1.2.2.1 *Erwinia billingiae*

E. billingiae was isolated, from different hosts, in different European regions by Billing and Baker (1963) and described later by Mergaert *et al*.,(1999). Conventional biochemical tests, BIOLOG metabolic fingerprinting system, fatty acid analysis, and 16S rRNA gene sequence studies showed that the bacterium is belong to the genus *Erwinia* but represents a novel species (Mergaert *et al.*, 1999).

1.2.2.2 *Erwinia tasmaniensis*

This species was isolated from apple and pear trees in three regions in Australia (Geider *et al.*, 2006). A polyphasic approach, which includes microbiological and API assays as well as fatty acid methyl ester analysis, DNA-DNA hybridization, and DNA sequencing, was later used to characterize the non-pathogenic *E. tasmaniensis*. Additionally, 16S rRNA gene sequence demonstrated *E. tasmaniensis* bacterium as a novel species (Geider *et al.*, 2006).

1.1.2 Disease cycle

1.3.1 Overwintering

E. amylovora overwinters in the holdover cankers which were formed previously in the last season (Figure 4). Once the temperature goes higher in spring, the bacteria in the margin of the cankers become active again. The released bacteria on the surface of the bark are the most probable origin of the inoculum to start the fire blight spring cycle (Sherman, 2000). Sometimes the released cells are visible as bacterial ooze onto cankers' surface which helps in the dissemination of the pathogen from cankers to open flowers. The bacterial ooze is a characteristic sign for *E. amylovora*, it is composed of viable bacterial cells embedded in a polysaccharide matrix (Van der Zwet and Keil, 1979). These bacterial cells can survive in low humidity conditions, in the dry exudates, for over a year (Sherman, 2000). Wind, rain splash, and passerby insects (like flies and ants) can easily transmit the bacteria from ooze, on cankers, to flowers. Although several insects can vector *E. amylovora*, there is no evidence in the literature about the role for honeybees (*Apis mellifera* L.) in the primary dispersal of inoculum, as honeybees have never been reported to visit the ooze. On the contrary, honeybees might be important vectors for biological agents (Nuclo *et al.*, 1996), so the role of these insects in the primary infection remains uncertain.

As the bacteria reach flowers, they colonize epiphytically the nutrient-rich stigmas. During the later phase, which is called the floral epiphytic, the population size of the pathogen is directly influenced by the temperature and the number of blossoms in which the pathogen become established. Temperature regulates the generation time of *E. amylovora*, and honeybees and other pollinating insects help increasing the number of contaminated blossoms.

1.3.2 Primary and secondary infections

Blossom blight, or the primary infection, starts when *E. amylovora* cells are washed externally from the stigma to the floral cup (hypanthium), and then they enter the plant through secretory cells called nectarthodes (Thomson, 2000). In plants that tend to produce nuisance (secondary blossom), like some pear species, the blossom blight importance expands long after blooming. On the other hand, secondary infections occur in shoot, fruit, and rootstock. They start from the primary inocula produced on diseased tissues as a result of blossom infection. Wounds are generally required to initiate shoot and fruit blight. Various insects, strong winds, rain, and hail can create numerous wounds in host tissues. Apple rootstock blight can result from shoot blight on water sprouts or from internal translocation of *E. amylovora* infections from the top of the tree.

1.3.3 Canker expansion

In summer, both infections (primary and secondary) can expand with ultimate severity based on the host species, cultivar, age, and the environment conditions. Young tender tissues and trees are more susceptible to the disease than older ones. Excessive nitrogen fertilization and irrigation can enhance the occurrence of the developing cankers.

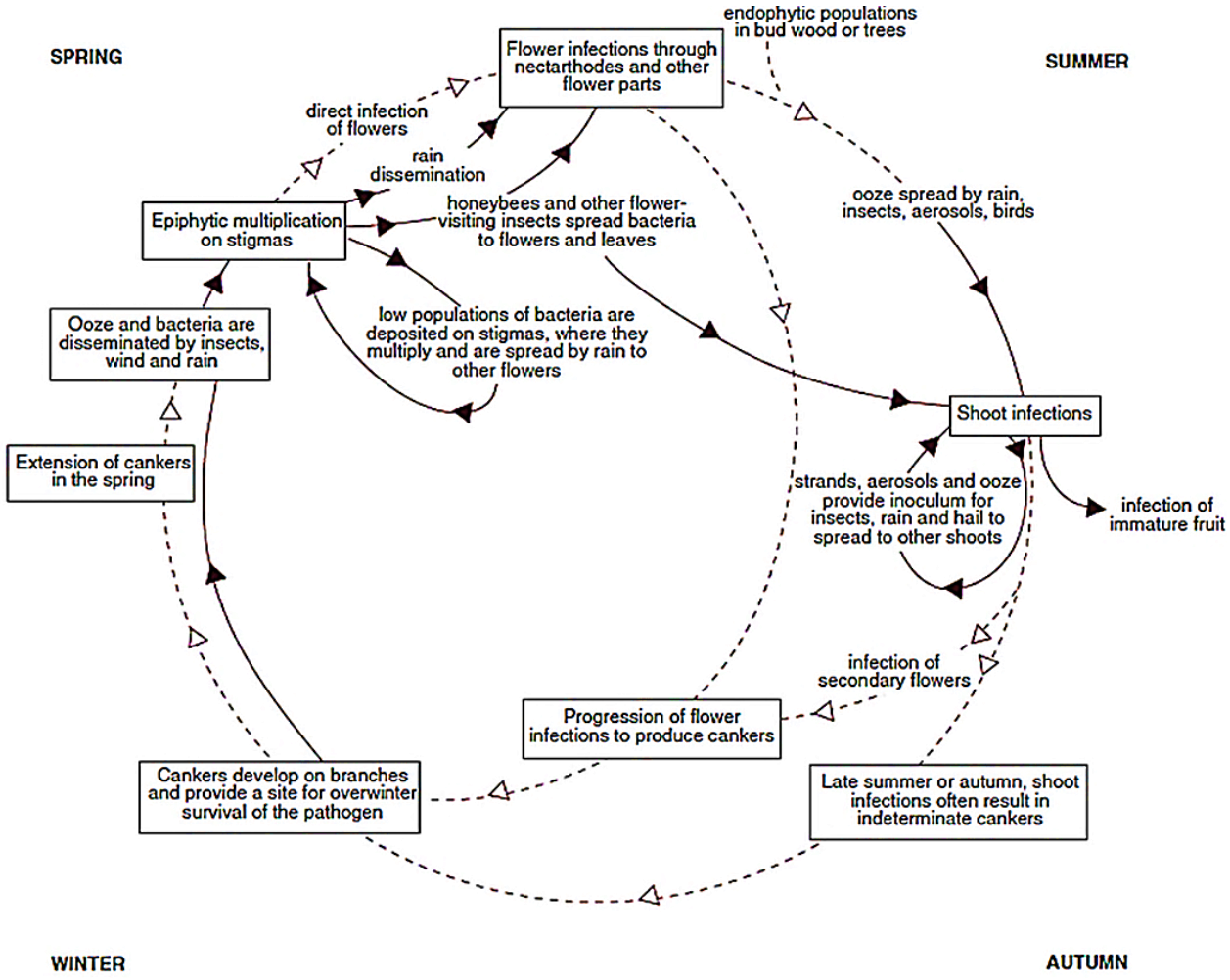


Figure 4 A schematic diagram of *Erwinia amylovora* detailed life cycle adapted from Thomson (2000). The dashed arrows represent the movement of the bacteria inside the plants and the solid ones represent the movement outside the plants.

1.2. Host range

E. amylovora is a unique pathogen and its strains are not host species-specific. For instance, most strains isolated from apple are pathogenic on pear, other *Maloideae* (family *Rosaceae*) subfamily species, and *Prunus* species. Strains isolated from *Rubus* species, on the other hand, are nonpathogenic on apple or pear (Timur Momol and Aldwinckle, 2000). Fire blight has broad host range within the plant family *Rosaceae*, in which the pathogen induces symptoms on 200 species in 40 Rosaceous genera (Van der Zwet and Keil, 1979). *E. amylovora* is well known as a pathogen of the four *Rosaceae* subfamilies members which are *Maloideae* (*Pomoideae*), *Rosoideae*, *Amygdaloideae*, and *Spiraeoideae*. Fire blight affects also other fruit crops and ornamental plants such as: *Amelanchier alnifolia* (serviceberry), *A. canadensis*, *Chaenomeles* spp. (Japanese quince), *Cotoneaster* spp. (cotoneaster), *Crataegus* spp. (hawthorns), *Cydonia* spp. (quince), *Eriobotrya japonica* (loquats), *Mespilus germanica* (medlars), *Pyracantha* spp. (firethorns), *Pyrus amygdaliformis* (almond-leaved pear), *Sorbus* spp. (mountain ash), and *Stranvaesia davidiana* (Chinese stranvaesia) (Timur Momol and Aldwinckle, 2000). Some researchers divided *E. amylovora* strains into two main groups according to their host specificity. The first group, *Spiraeoideae*-infecting strains, infects a large number of the subfamily *Spiraeoideae* such as apple, pear, hawthorn, and cotoneaster. The second group, which referred to as *Rubus*-infecting strains, infects host plants in the genus *Rubus* such as raspberry and blackberry (Braun and Hildebrand, 2005).

Fire blight was unexpectedly reported to infect plum trees naturally. The earliest report was on a Japanese plum tree (*Prunus salicina* L.) in the United States of America (USA; Rosen and Groves, 1928). During the summer of 1993, an extensive outbreak was reported on Japanese plum orchard in the USA, and recently it was reported on European plum trees (*Prunus domestica* L.) from Germany (Vanneste *et al.*, 2002) and Hungary (Végh *et al.*, 2012).

1.3. Symptoms and signs

- **Blossom blight:** the earliest symptom of the disease occurs on flowers. Blossoms appear discolored (water soaked, gray green progressing to black) before they turn necrotic. The dead blossoms may, or may not, remain attached to the tree, and under high humid conditions the infected blossoms exude drops of ooze (Steiner, 2000).

- **Shoot blight:** the infection of shoots is always rapid and first apparent in the actively growing young leaves on the tips of the shoots. Leaves on diseased shoots often show blackening along the midrib and veins before becoming fully necrotic. Occasional spots on older leaves may also appeared on leaves, but the typical wilting and ooze production do not develop on fully expanded leaves until after the tip has expressed symptoms. Sudden fire blight infection which is induced by severe weather are sometimes called trauma blight (Johnson, 2000).

- **Fruit blight:** symptoms on fruit varied according to the time of the infection. In the early infection (immature fruit), infected fruit keep small in size, become discolored, shrivel, blackened, and remain attached to the tree. In the late infection, the infected fruit remain fresh looking and do not become discolored. Infected fruit may also exude large amounts of bacterial ooze in humid conditions (Johnson, 2000; Steiner, 2000).

1.4. Fire blight monitoring

The most effective tool to prevent disease dissemination to new areas is the vigilant monitoring. Fire blight symptoms should be frequently inspected in pome fruit orchards and subsequently destroying the infected material.

1.4.1. Remote sensing detection

Remote and proximal sensing (Multispectral and Hyperspectral, respectively) have been used to detect, monitor, and quantify a range of diseases in different crops (Alnaasan, 2012). Remote and proximal sensing techniques allow also the replacement of time-consuming manual digitizing with more reliable automatic mapping procedures, thereby providing useful information on agricultural operations for updating regional databases as well as delivering a valuable aid to identify orchards and to position fruit trees individually (Santoro *et al.*, 2009; Santoro *et al.*, 2011). In 2006, Spinelli and others applied Near-Infrared spectroscopy (NIRs) to discriminate between fire blight-symptomatic and asymptomatic pear trees but no discrimination was obtained.

1.4.2. Fire blight risk assessment models

The fire blight disease cycle contains several phases: canker blight, blossom blight, shoot blight, and fruit blight. To stop the continuous dissemination later in the season, it is crucial to control blossom blight by chemical or biological intervention as early as possible. Until recently, predicting the start of blossom blight epidemics accurately and reliably, for timely spray applications, has been one of the most limiting factors in improving fire blight management. Several models to predict the disease on apples and pears have been developed and used, though MARYBLYT and *CougarBlight* are the best-know systems. These models were originally developed in the USA, and both of them use weather and apple, or pear, phenology to generate infection predictions, but they differ in how risk values are calculated (Dewdney *et al.*, 2007).

MARYBLYT, which was developed in Maryland and West Virginia by Steiner and Lightner (1990b;a), basically predicts both infection and symptoms development in the different fire blight phases. For instance in case of apple blossom blight, the infection is likely to be imminent, hypothetically, when the minimum values of the following four conditions are met in sequence: (I) flowers must open with healthy stigma and petals, (II) accumulation of at least 110 DH $>$ 18.3 °C [Degree-Hours (DH): number of hours above a given temperature threshold] within the last 44.4 DD $>$ 4.4 °C [Degree-Days (DD): a measure of how much (in degrees) and for how long (in days) the outside air temperature was above a given temperature threshold], (III) occurrence of precipitation or dew $\geq$ 0.25 mm on the current day, or $\geq$ 2.5 mm during the previous day, (IV) average daily temperature $\geq$ 15.6 °C (Steiner, 1990a).

CougarBlight was developed to predict fire blight (blossom blight) on apples and pears by Timothy J. Smith in Washington State (Smith, 1999). Unlike other fire bight models, the inoculum pressure (the estimate of overwintering inoculum) is considered in the *CougarBlight* prediction process. Although the inoculum pressure cannot be measured, the infection history of an orchard is used as an indicator of the inoculum pressure. The used categories of the infection history (potential for pathogen presence) in *CougarBlight* are given in Table 1. Within every inoculum pressure category four risk levels (low, caution, high, and extreme) are assigned based on a 4-day DH > 15.5 °C. Most growers consider management action when the risk level reaches high or extreme category (Dewdney *et al.*, 2007).

Table 1 Risk levels and categories of orchard infection history according to *CougarBlight*.

Potential for pathogen presence	Low	Caution	High	Extreme
No fire blight in area past two seasons	0–350	350–500	500–800	> 800
Fire blight in local area past two seasons	0–300	300–500	500–750	> 750
Fire blight in local area last year	0–250	250–450	450–700	> 700
Fire blight in your orchard or your neighbor's orchard last year	0–200	200–350	350–500	> 500
Active cankers present nearby	0–100	100–200	200–350	> 350

1.5. Fire blight detection and management

E. amylovora can be detected and identified by several methods and techniques, though the most known and commonly used tests are:

- Conventional and biochemical tests which basically include streaking or plating the bacteria onto non-selective or semi-selective substrates. Other conventional identification methods are made by the physical examination of the morphological characteristics of the colony. A well-known group of biochemical key tests are commonly used to identify *E. amylovora*. These tests include Levan production, tobacco hypersensitivity, fluorescence on King's B substrate, and agglutination tests.
- Serological tests are less used in the regular detection of *E. amylovora*, though the most common tests are: Enzyme-Linked Immunosorbent Assay (ELISA), lateral flow Immunoassay, slide agglutination, and Immuno-Fluorescence (IF).
- Molecular tests are widely used in the detection and the identification of *E. amylovora*. PCR-based methods such as repetitive element palindromic PCR (rep-PCR), Amplified Fragment Length Polymorphism (AFLP), fluorescent-AFLP (Yaich *et al.*, 2011), Random Amplified Polymorphic DNA Fragment (RAPD) analysis, and Short Sequence Repeats (SSR) are intensively used for fire blight detection and for the molecular investigations on genetic variability among strains. Recently, Loop-Mediated Isothermal Amplification (LAMP) has been successfully developed and used as a rapid detection method (Notomi *et al.*, 2000). Besides the mentioned molecular diagnostic methods, new discriminatory markers to subdivide *E. amylovora* have been developed successfully. Multiple-Locus Variable Number of Tandem Repeats Analysis (MLVA) and Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) analysis (Rezzonico *et al.*, 2011; Bühlmann *et al.*, 2014) are the best example of these discriminatory markers which are needed to: (I) test whether outbreaks have a common source, (II) track the epidemic and the pandemic spread of particular strains, and (III) reconstruct the disease evolutionary history (Achtman, 2008).

1.5.1. Quarantine measurements

Fire blight is difficult to control and nearly impossible to eradicate when the infections are established in a given area. In countries that are considered a fire blight-free, *E. amylovora* spreading must be controlled by strict quarantine legislations on both national and regional level, and nursery stock and fruit importing should also be restricted from countries with fire blight risks and outbreaks.

According to the Italian Ministerial Decree 356/99, fire blight control is mandatory in all Italian territories to prevent new introductions and spread. Establishing several strict procedures, the decree was issued in 1999. These procedures include: (I) performing fire blight official systematic investigation on defined plant genera with particular attention to nurseries, (II) setting up monitoring networks in both public and private areas, (III) reporting the contaminated areas with immediate destruction of symptomatic plants, and in case of an outbreak all symptomatic and asymptomatic plants must be destroyed on a radius of 10 m, (IV) creating a safety zone of at least 3.5 km^2 (1 km radius around the outbreak point), (V) prohibiting plant materials trade or beehives movement in the reported areas for at least 12 months, and lastly (VI) the phytosanitary services frequently monitor and investigate to decide the retention or the displacement of the protected zones (M.D., 1999).

1.5.2. Cultivar selection

Selection of a resistant cultivar is the most effective method to control fire blight. In apple, for example, some cultivars are moderately resistant to the disease (e.g. Red Delicious and Golden Delicious). For pears, cultivar choices are limited because superior horticultural traits (e.g. taste, storage, and marketing qualities) have been difficult to combine with higher levels of disease resistance (Johnson, 2000).

1.5.3. Orchard sanitation

Removing of expanding and overwinter cankers is essential for fighting against fire blight in apple and pear. Elimination of holdover cankers is accomplished by inspecting and pruning trees during winter. In spring and summer, established infections are principally controlled by bloom thinning and pruning. Effective control through pruning requires the cutting at approximately 20–25 cm below the visible end of the expanding canker (Johnson, 2000). Generally, pruning should be done by cutting, in the healthy-looking area, about 30–70 cm below the infection site using disinfected tools (by alcohol or bleach solution). Recent study showed that *E. amylovora* can survive in apple dead leaves for five months, so special attention should be paid to eliminate the dead leaves in orchards (Sobiczewski *et al.*, 2013).

The rate of fire blight canker development can be decelerated by slowing the tree growth rate, which includes plant growth regulator (e.g. prohexadione calcium), withholding irrigation water, nitrogen fertilizer, and cultivation (Deckers and Daemen, 1993). Similarly, practices that reduce tree wounding and bacterial movement can reduce secondary infection. These practices include controlling insects, setting windbreaker and hail preventing net, limiting use of limb spreaders (which are woody or plastic rods used to spread tree's limbs and extend the canopy volume) in young orchards, and avoiding the use of overhead sprinklers.

1.5.4. Dormant and blossom spray

In apple and pear orchards with a history of severe fire blight, it is advisable to carefully prune out overwintering cankers and spray with a copper and oil mixture at the delayed dormant stage. Copper compounds can be phytotoxic if they are sprayed after the bud burst stage. This spray is thought to reduce the levels of primary inoculum in the orchard, but may also be effective in reducing insect vectors (Steiner, 2000).

Concerning blossom spray, all sprayed products are ineffective in controlling fire blight unless they are applied within a very narrow timing period. *E. amylovora* colonies establish daily on newly opened untreated flowers. The effective control product must be applied into each flower within a day or two of its opening to protect it adequately from infection. Spraying during blossom is difficult, as many other important sprays are being applied at this time, and sprayers may be risky (Steiner, 2000). Antibiotics, which are currently restricted in too many countries, and fixed copper (copper oxychloride sulfate) sprays are effective in reducing fire blight once they are properly applied at correct times (Deckers *et al.*, 1990). They are only preventive sprays and their application must be repeated every four to five days as long as new flowers are opening.

1.5.5. Biological control

The biological agents, for fire blight control, are considered the only exception to the narrow-timing spray particularly during blooming. The sprays of living organisms are applied to newly opening flowers during the two–four days of warming weather leading up to an actual infection event. Technically, the potential infection periods must be anticipated to use these biological products effectively. Several biological control compounds (antagonistic bacterial competitor e.g. *Pantoea agglomerans* and *Pseudomonas fluorescens*) have recently been licensed for fire blight management (Vanneste *et al.*, 2004). Biopro® and Serenade Max® are newly registered products, which are produced based on *Bacillus subtilis* strains, for fire blight control in Europe. More important is the use of *Aureobasidium pullulans*, which is a live formation yeast commercialized in Germany few years ago (Kunz *et al.*, 2011), as an alternative to streptomycin to protect blossoms for fire blight infection.

1.6. Epidemiology

E. amylovora spreads in several ways and the main routes are: importing infected propagating materials, migratory birds may harbor the pathogen, deposition of aerosols, and rain splash carried by air current (Mazzucchi *et al.*, 1993). However, it has not been confirmed that migratory birds are involved in the long-distance dissemination of *E. amylovora*, but the movement of nursery plants and grafting materials from asymptomatic infected plants (symptomless carriers) into regions considered free of fire blight contribute effectively in this type of dissemination.

The primary fire blight foci resulted from the later mentioned routes have a random distribution, whereas for the secondary foci they are distributed according to a disease tendency in the main directions (Burdon, 1987; Mazzucchi *et al.*, 1993).

1.7 Historical review of *Erwinia amylovora*

Fire blight bacterial disease has been reported from 55 countries around the world (EPPO, 2014). The story of this dangerous disease began when the disease has spread from its original site the Hudson Valley in New York, USA through the North American countries and the Pacific coast, Europe, and the Middle East during the past years (Bonn and van der Zwet, 2000).

1.7.1. World

- **1780**: Denning (1794), the earliest recorded observation of fire blight disease in Hudson valley of New York, USA.
- **1890**: McMichael (1890) reported, for the first time, pear blight disease on pear in Waterford, Canada. The disease was recorded as personal observations of fire blight by a pear grower during the previous 25 years.
- **1938**: Waterston (1938), the first report of *E. amylovora*, on loquat branches at blooming, in the Bermuda islands a territory of the UK in the North Atlantic Ocean.
- **1903**: Bonn and van der Zwet (2000) mentioned the first report of fire blight, on both apple and pear, outside the North America in one Japanese island.
- **1958**: Crosse *et al*. (1958), the first authenticated record of fire blight disease in Europe, in which it was reported on pear trees near Maidstone, England**.**
- **1960** till now, fire blight was reported in Eastern Europe, from Greece, Armenia, Bulgaria, Romania, and former Yugoslavia (Serbia, Bosnia, and Croatia). Also it was confirmed to be present in other countries like Albania, Spain, and Australia (Vanneste, 2000). Recently, the disease was recorded in Algeria and Tunisia (Laala *et al.*, 2012; Rhouma *et al.*, 2014).

1.7.2. Italy

- **1980**: Calzolari *et al*. (1982), *E. amylovora* live cells were found in a lot of asymptomatic apple trees imported from the Netherlands.
- **1981**–**1983**: Calzolari *et al*. (1984), the first monitoring for fire blight in the region without any finding 1981–1983.
- **1990**: Cariddi and Piglionica (1990; 1990), first record of fire blight in Italy. The bacteria were found on pear in two isolated areas of 500 km^2 in Apulia region, particularly on cultivars Guyot, Bella di Giugno, Williams, and conference.
- **1991**: Mazzucchi *et al*. (1993) and Mazzucchi (1994), a fire blight national monitoring network was set up for early detection of primarily foci after spread. This network was divided into three interregional sub-networks (Northern, Adriatic, and Tyrrhenian) and two islands (Sicilia and Sardegna). In 1992, the monitoring network was completed and consisted of 3601 points, each point had at least one *E. amylovora*-susceptible host at about five km interval. Fire blight symptoms were therefore inspected twice a year (May-June and September-October).
- **1991**: Cariddi (1993), *E. amylovora* was found in one orchard of local pear in Sicily and that was the first reported outbreak of the disease in Sicily, the single detected focus was officially directly eradicated.
- **1994**: Griffo *et al*. (1998), two fire blight limited foci were identified as the first report of the disease in Caserta, Campania. The two foci were eradicated to stop the disease spreading.

- **1994**–**1997**: Calzolari *et al*. (1999), two *E. amylovora* foci were found in Emilia-Romagna region, Bologna in 1994. One focus in commercial pear orchard in Casale and the other in four nurseries in Minerbio area, where the distance between the two areas was nearly 10 km. later after this detection, six cases of fire blight were detected in 1995 and 30 cases in 1996 in the same area. In 1997 a fire blight outbreak occurred as a total of 721 foci were officially identified on both pear (90%) and ornamental plants (10%), as a result of this outbreak approximately 500000 fruit trees were destroyed including 370000 pear trees in nurseries and orchards.
- **1999**: The Italian Ministerial Decree (D.M. 356/99) on the mandatory control of fire blight was issued after the epidemics of 1994. This national regulation lay down the measures for occurrence of new foci in new areas, infected plants (in orchard, nursery, and wild environment), marketing of propagating materials, and even the movement of beehives.
- **2002**: Clabassi *et al*. (2002), the first report of fire blight in Friuli-Venezia Giulia region on pear orchard.
- **2003**: Gianetti *et al*. (2004), fire blight reported in Piemonte on cotoneaster.
- **2005**: NPPO of Italy (2005), *E. amylovora* was found on *Malus* in the 'Valle di Non' province of Trento (Trentino-Alto Adige region).
- **2009**: NPPO of Italy (2009), fire blight was detected in a few plants of hawthorn and pear in Lombardia.
- **2012**: Pucci *et al*. (2013), fire blight was reported for the first time in Latium on a quince plant.

Chapter 2

Hyperspectral discrimination of fire blight infection in apple and pear

Abstract

Fire blight is a capricious bacterial disease that can destroy apple and pear orchards in a single season. Until 2015, no known cure for this disease was found, and all successful control strategies rely on prevention, eradication, and containment measurements. Although early detection and frequent monitoring are fundamental tools in any effective management strategy of fire blight, asymptomatic infections and the enormous number of trees reduce the efficiency of these tools dramatically. Moreover, large-scale monitoring needs endless time, huge budget, and a large number of workers. The only available data source that provides access to systematic observations at scales ranging from local to global is remote and proximal sensing. The role of remote sensing is therefore indispensable in all the efforts to improve new tools that help in detecting stressed plants through their reflectances and ensure vigilant monitoring for countless number of trees. This study aims at exploring the potential of hyperspectral sensor systems for detecting fire blight infections, symptomatic and asymptomatic, on apple and pear, and subsequently developing of predictive models that can recognize the fire blight-infected plants. Hyperspectral reflectance data were capable of discriminating between healthy and fire blight-infected (symptomatic or asymptomatic) apple and pear plants in glasshouse and field trials. Furthermore, Partial Least Squares-Discriminant Analysis (PLS-DA) and Artificial Neural Networks (ANNs) classification models were successfully built, based on Reflectance Wavebands (RWs) and Spectral Vegetation Indices (SVIs), and the performance of each model was evaluated in terms of non-error rate, sensitivity, and specificity. PLS-DA and ANNs models resulted in high sensitivity to fire blight-infected plants, though PLS-DA outperformed ANNs models specially the models developed from the SVIs. The current research shows the hyperspectral reflectance with classification models can be used as an effective tool in fire blight detection and surveillance on large-scales.

2.1. Introduction

2.1.1. Precision Agriculture

Precision Agriculture (PA), or site-specific crop management, is a relatively new concept that can be defined by innumerable series of definitions. However, one of the first and most significant definitions was proposed by Blackmore (1994): the use of information technologies to obtain knowledge about important crop parameters, and consequently optimize the management of the production system on an appropriate scales. PA tries particularly to increase crops productivity by the efficient use of natural resources (such as land, seeds, and water) and agriculture practices such as plant protection (Precision Crop Protection; PCP), fertilization, cultivation, and other practices. Today, PA is totally motivated by many economic issues including the cost reduction of large-scale monitoring of plant diseases, and the reduction of fuel consumption. Moreover, PA is motivated by environmental concerns such as the excessive use of chemicals and other agricultural inputs (Suzuki *et al.*, 2002).

PCP requires spatial and temporal information on the within-field heterogeneity of crop growth conditions. These data are collected, managed, analyzed, and implemented by six main technologies and tools that are illustrated in Figure 5.

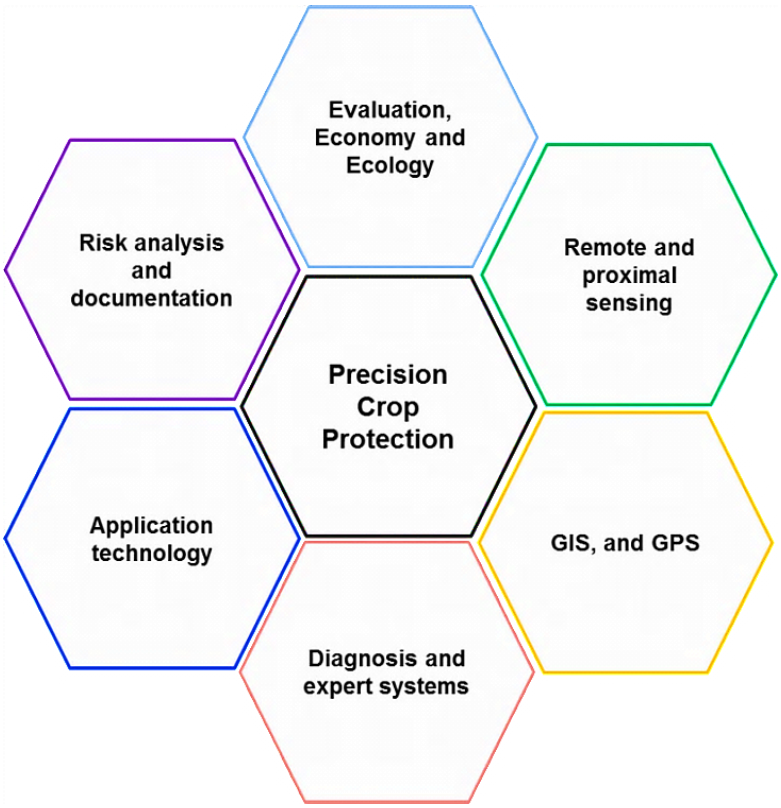


Figure 5 The six principal components of the Precision Crop Protection.

Commonly, remote and proximal sensing technologies are extremely important to collect spatial and spectral data about plant vitality, stress, pest monitoring, and other aspects. Geographical Information System (GIS) is useful for the management and the handling of spatial data. These collected data must be georeferenced by using Global Positioning System (GPS), and then they are organized and optimized in expert systems (informatics systems) for the previously developed models, and control techniques and strategies (Decision Support System).

2.1.1.1. Remote and proximal sensing in plant protection

Remote and proximal sensing can be broadly defined as the process of acquiring information about an object, area, or a phenomenon through the analysis of data acquired by sensors that are not in a direct contact with that object, area, or phenomenon (Lillesand *et al.*, 2008). In proximal sensing, the distance between the sensor and the object of interest is usually less than, or comparable to, the dimensions of the sensor, whereas this distance is vastly higher in case of remote sensing. The process of remote sensing based on six elements, illustrated in Figure 6, including:

- **Energy source**: a fundamental component for remote sensing to provide the target with the required energy such as the sun radiation or any artificial light source.
- **Sensors**: usually mounted on satellite, drone, or aircraft, sensors are basically classified according to their energy source into active sensors (emit their own electromagnetic energy and record the reflected energy) and passive sensors (record the reflected or radiated energy from naturally illuminated targets).
- **Target**: the radiation interacts differently with the target, depending on the nature of the target and the radiation properties.
- **Reception and transmission**: the reflected energy from the target is received and recorded by sensors, and then the recorded data are transmitted to receiving stations to be preprocessed and stored in digital form.
- **Analysis and interpretation**: the data analyzed and interpreted using special instruments and pieces of software to extract the information from the target. Specialist and experts use these pieces of information to solve particular problems and to apply the relevant solutions and applications.
- **Application**: the remotely sensed data are widely used in several applications and the main three levels for application are: (I) mapping and monitoring of natural resources, (II) determining the bio-geo-physical parameters, and (III) Decision Support System.

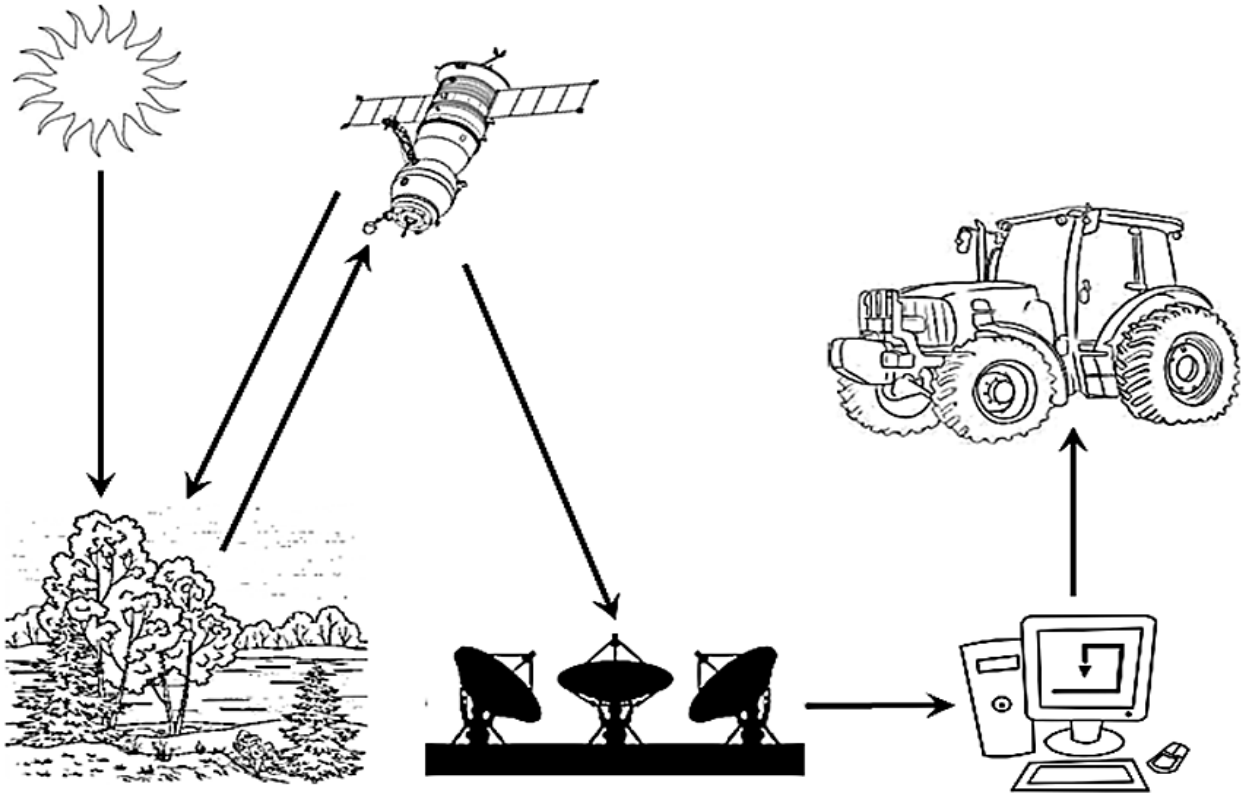


Figure 6 A schematic representations showing the main elements in satellite remote sensing process.

Remotely and proximally sensed data are collected in different forms and sizes, such as force-field, acoustic energy, and Electromagnetic Radiation (EMR). However, EMR waves are the most used type, as depicted in Figure 7 its waves comprise both electric and magnetic fields perpendicular to each other and perpendicular to the direction of energy propagation. The range of all electromagnetic waves, which are arranged according to their frequency (*f*) and wavelength (λ), is called Electromagnetic Spectrum. This spectrum extends from short wavelengths (extremely high frequency) gamma radiation that measure less than 10^{-12} m to radio waves (low frequency) which can measure thousands of kilometers. Although the visible portion of electromagnetic wavelengths was the original form of remote and proximal sensing, the technology development has enabled the obtaining of information from other wavelength including Near-Infrared (NIR), thermal infrared, and microwave. The data are commonly collected either over a small number of wavelength bands (multispectral data) or over a large number of bands (hyperspectral data).

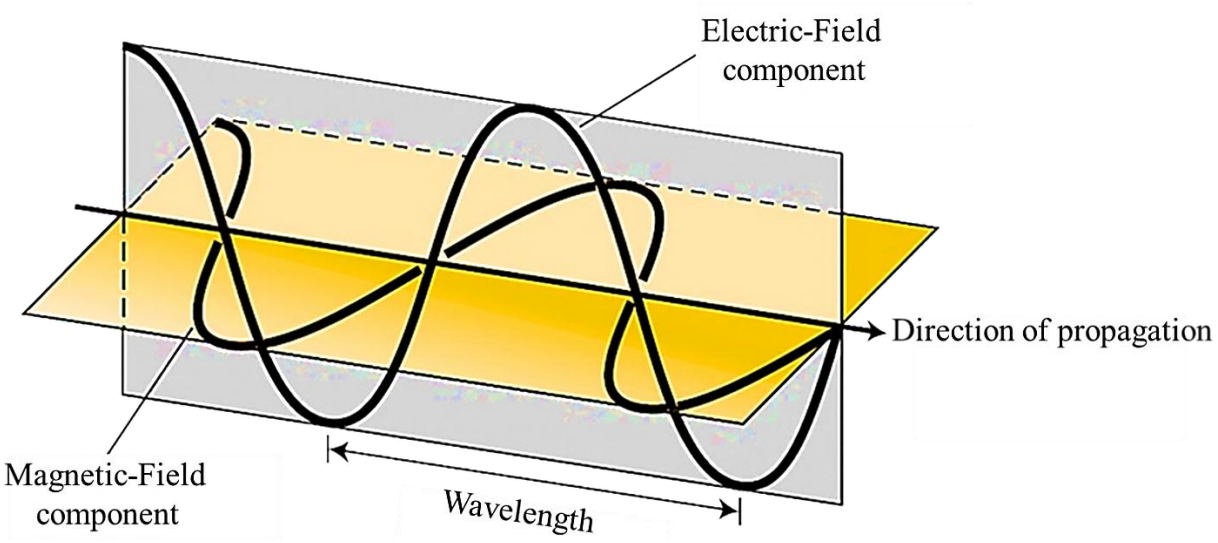


Figure 7 A graphical representation of an electromagnetic wave showing the mutually perpendicular electric and magnetic fields, the orientation of the wave, and the characteristic wavelength λ.

Today, several remote and proximal sensing models can be successfully applied to estimate crops acreage and production, crop nutrient deficiency detection, crop condition assessment, crop yield modeling, soil mapping, and most importantly in plant disease detection. Norman and Fritz (1965) used the infrared photography successfully to detect diseased and declined citrus trees in Florida, USA. Same technology was also used for locating nematode infestation (*Rotylenchulus reniformis* Linford and Oliveira) in cotton fields in Texsas, USA (Heald *et al.*, 1972). Later, in 1989, Lorenzen and Jensen noticed an increase in the reflectance, in the visible wavelength, of barley leaves infected by powdery mildew due to the degradation of chlorophylls.

The global potential loss, in some crops, due to pests ranged from 50% in wheat to more than 80% in cotton production (Oerke, 2006). Luckily plant protection measures can prevent or reduce crop loss. However, the excessive use of pesticides for protecting crops will increase not only the cost of production, but also raises the danger of toxic residue in nature and in agricultural products. Early detection of plant pests is critical to prevent plant diseases and stress and therefore to minimize acute and chronic loss of productivity. Furthermore, the control success and effectiveness depends directly on fast identification and local treatment of the plant disease which requires collecting the disease information in the field as early as possible. Although manual field survey is the most common method for disease identification, it is time consuming, inefficient, and requires high labor cost. Thus, manual inspection for plant diseases is definitely impractical in large areas. Remote and proximal sensing, fortunately, can provide us with the necessary information about the spatial distribution of crop disease over large areas.

As a result, remote and proximal sensing technology provides a rapid, accurate, harmless and cost-effective tool to assess plant stress resulted from differences in the vegetation spectral characteristics due to biotic and abiotic factors. Furthermore, this technology can be effectively employed in the Integrated Pest Management (IPM) strategies to detect and scout the plant diseases in early manner.

2.1.2. Spectral signature and Spectral Vegetation Indices (SVIs)

When light strikes a surface such as vegetation (e.g. plant leaf) the biggest part of the light is transmitted, a smaller part is absorbed, and the smallest proportion is reflected (Figure 8 A). The relative amounts of reflected, transmitted, and absorbed light depend on the surface, and they vary with the wavelength of the light. The response characteristics can be recorded and plot against wavelength using a spectroradiometer, and the resulted curve is called the spectral signature of that vegetation type (Figure 8 B). The spectral reflectance is the ratio of reflected energy [E_R] to incident energy [E_I] as a function of wavelength. It is given by the equation:

$$R = \frac{E_R}{E_I} \times 100$$

Where,
R is the spectral reflectance, E_R is the reflected energy from the vegetation, and E_I is the incident energy on the vegetation.

The spectral absorption features are mainly attributed to plant pigments (chlorophylls, xanthophyll, and carotenoids) and water, whereas secondary absorption features are attributable to other chemical components such as cellulose, lignin, proteins, starches, and sugars (Gausman, 1977; Fourty *et al.*, 1996; Asner, 1998). Generally, leaf pigments, cell structure, and water content are the main determinants that control the spectral responses in the visible, Near-Infrared, and Short- Infrared wavelengths, respectively as shown in Figure 8 B. Most of the light in the NIR wavelength is transmitted and reflected with little absorption, in contrast to the visible wavelength where the light absorption is predominant with some reflection and little transmission.

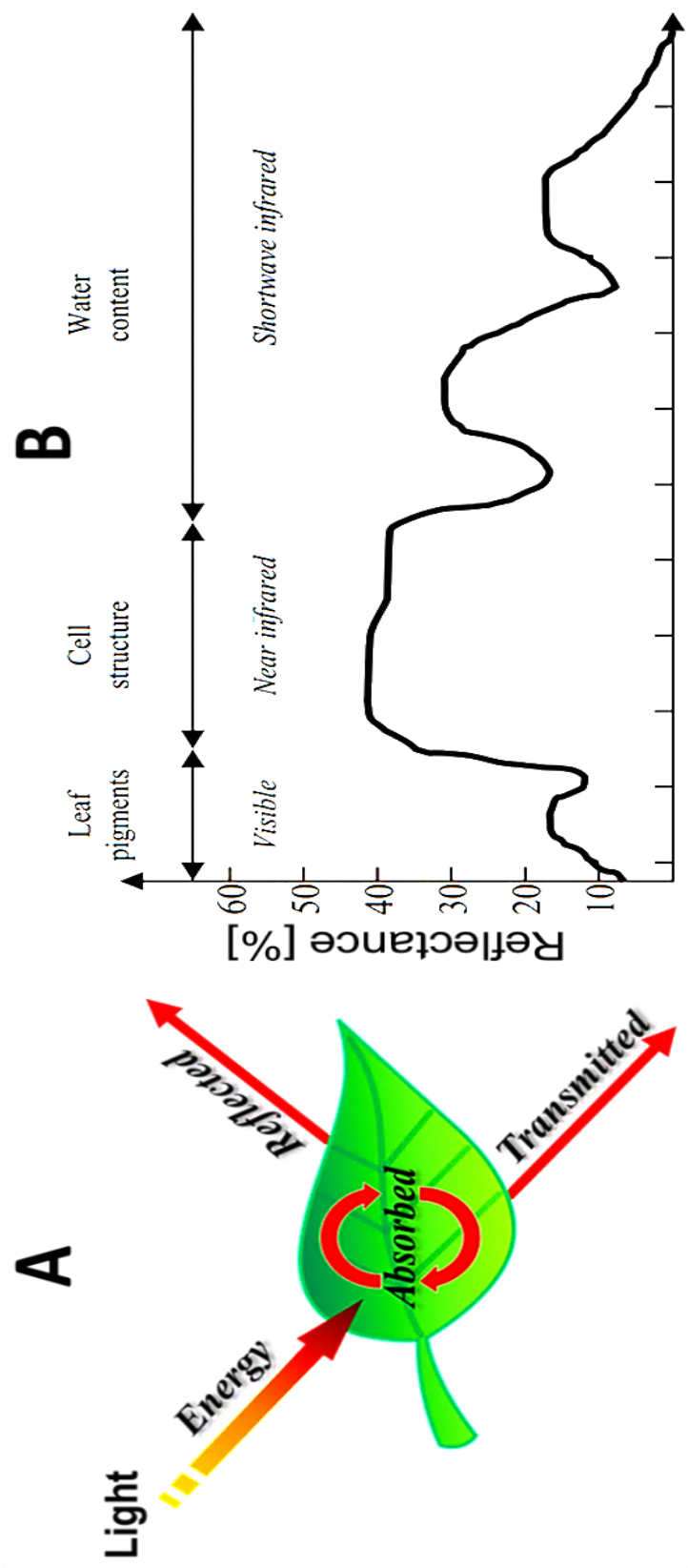


Figure 8 (A) A representation of the three possible pathways (transmission, absorption, and reflection) of incident light energy on vegetation, and (B) a reprinted typical spectral reflectance curve (spectral signature) that is controlled by leaf pigments, cellular structure and water content of the vegetation (Gausman, 1977).

In remote and proximal sensing, the solar irradiance varies with time and atmospheric conditions, thus a simple measure of light reflected from a surface is insufficient to characterize the surface in a repeatable manner. This problem can be circumvented somewhat by combining data from two or more spectral bands (wavelengths) to form what is commonly known as Spectral Vegetation Index (SVI). SVIs have been therefore proposed to estimate vegetation biophysical and biochemical characteristics by reduce the effects caused by external factors such as the atmosphere and the soil background.

SVIs are classified in two general classes, ratios and linear combinations, and their applications ranged from the plant leaf to the global level. The first reported SVIs was the Simple Ratio (SR), it takes advantages from the inverse relationship between the absorption in the Red region and the increased reflectance in the NIR region for healthy plants (Jordan, 1969). The most used SVI, however, is the Normalized Difference Vegetation Index (NDVI). NDVI is used to estimate agriculture land productivity, identify ecoregions, monitor the phonological patterns of global vegetation, and to assess the length of the growing and drying down periods (Huete and Liu, 1994; Ramsey *et al.*, 1995; Lenney *et al.*, 1996).

Objectives

Due to the high economic importance of apple and pear worldwide, particularly in Italy, and to the serious damage of fire blight disease, that can develop asymptomatic infections, there is a substantial need to develop new detection methods capable of performing early, accurate, and non-destructive monitoring on large scales. In this study two main objectives were sought:

- Check the capability of hyperspectral data to separate between healthy and fire blight-infected (symptomatic and asymptomatic) apple and pear in glasshouse and open fields.

- Calculate, calibrate, and validate non-parametric classification models for fire blight detection in apple and pear.

2.2. Materials and methods

2.2.1. Glasshouse trials

2.2.1.1. Area of study

Two main trials were carried out in the quarantine glasshouse of the CIHEAM-Mediterranean Agronomic Institute of Bari (MAIB; Figure 9), in which the ambient conditions are ~ 25 °C during the day and ~ 20 °C during the night, 60 ± 10% relative humidity, and a photo-period of 16 hours light and eight hours darkness.

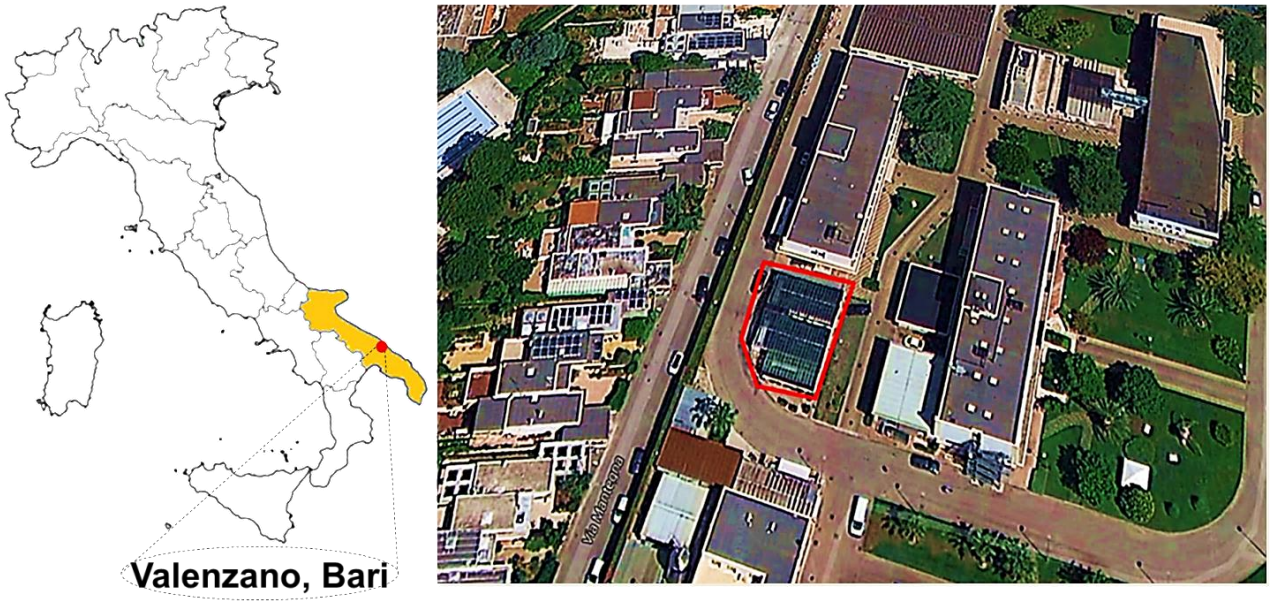


Figure 9 The location of the study area for the glasshouse trials on the Italian map (left), and an aerial view of the quarantine glasshouse at MAIB (right).

2.2.1.2. Plant materials, bacterial strains and artificial inoculation

Two important pome fruit species, apple and pear, were used in the glasshouse trials in this study. The studied plants were certified, though they were tested biochemically and molecularly to confirm being free from *E. amylovora* and *P. syringae* pv. *syringae*. These plants were also grown into regular soil mixture of 1/3 peat, 1/3 compost, and 1/3 sand in black plastic pots (10cm X 10cm X 25cm deep), and watered and sprayed with pesticides as necessary. Finally, pear plants were grafted onto common pear rootstocks and apple plants were grafted onto common apple rootstocks to make these tested plants similar to those commonly grown in farmers' fields.

Twenty–thirty days after grafting, the studied plants were artificially inoculated with the studied bacteria. One *E. amylovora* strain (OMPBO392,1), kindly provided by Dr. Alessandra CALZOLARI (Regional Phytosanitary Services, Emilia-Romagna), was used to prepare the inoculum in this research, in which a loopful of 48-hour-old bacteria, grown on King's B substrate at 28 °C, was emulsified in 3 mL of sterilized distilled water, and the suspension adjusted to 10^8–10^9 CFU/mL. The concentration of the inoculum was estimated by measuring the absorbance at 620 nm (Momol *et al.*, 1998).

Apple and pear plants, planned to be inoculated by *E. amylovora*, were divided in two groups (three plants in each apple group and two plants in each pear group) and two inoculation methods were evaluated on these groups separately. The first inoculation method, the cutting wounding, was carried out by a sterile blade dipped in the prepared bacterial suspension and then several superficial cuts were made in the stem of the experimental (Figure 10 A). The second method was the direct injection of the inoculum. Here plant shoots were inoculated by inserting a hypodermic needle in the stem, leaf midrib, and the attach point between the petiole and the stem. The emergence of inoculation drops from the needle wound was considered as a sufficient inoculum was introduced to the plant (Norelli *et al.*, 1987; Van der Zwet and Bell, 1995) as shown in Figure 10 B. All inoculated plants in the two trials were retested one month after the artificial inoculation. Consequently, *E. amylovora* was isolated directly from the *E. amylovora*-inoculated plants by plating fresh plant extracts onto two different substrates: Nutrient Sucrose Agar and King's B. Then, the pathogen was identified properly by the reported nutritional, enzymatic, and molecular tests.

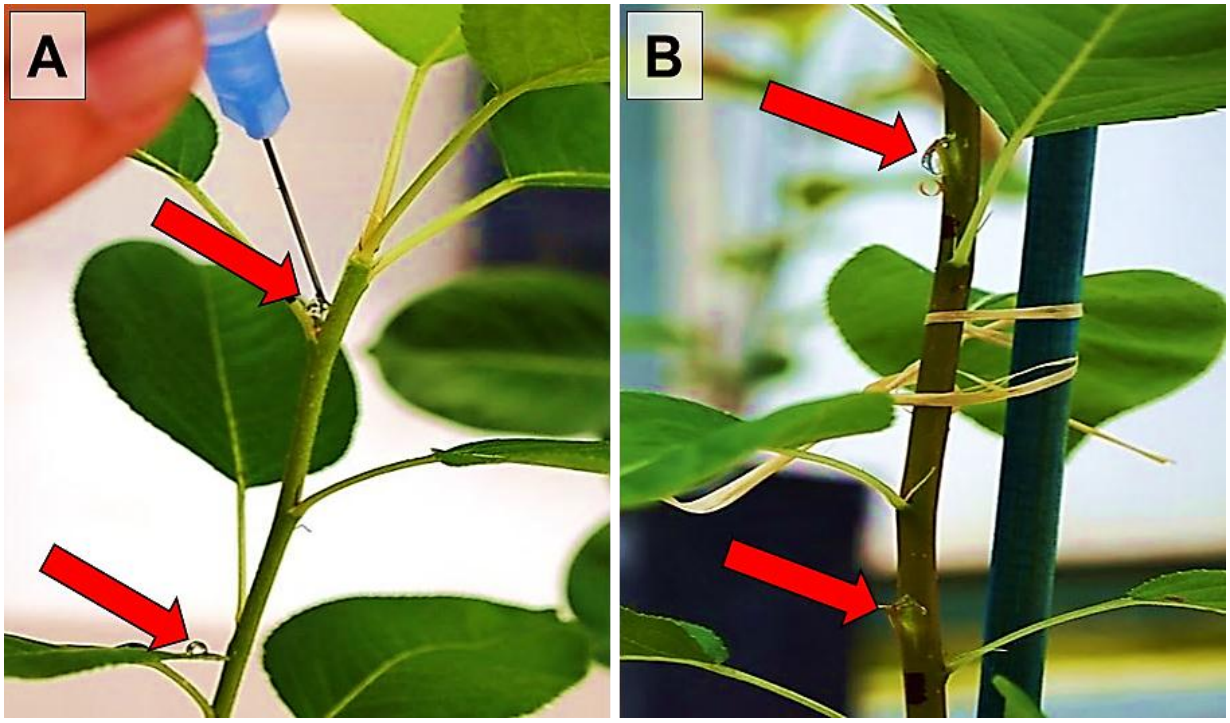


Figure 10 The artificial inoculation methods used in the glasshouse trials: (A) direct injection by the bacterial suspension using a hypodermic needle and (B) cutting wounding by a sterile blade dipped in the bacterial suspension.

2.2.1.3. The first glasshouse trial

Ten apple (cv. Golden Delicious) and seven pear (cv. Beurre Hardy) plants were grafted onto *P. communis* and *M. communis*, respectively in late July 2014. Then, six grafted apple and four grafted pear plants were artificially inoculated by *E. amylovora* (OMPBO329,1 severe strain) using two inoculation methods, cutting wounding and direct injection, and four apple and three pear plants were injected by Sterilized Distilled Water (SDW) in September 2014.

2.2.1.4. The second glasshouse trial

A second glasshouse trial was carried out only on pear potted trees in September 2014. Eight pear plants (3-year-old; cv. Bella di Giugno) were used in the second glasshouse trial. Four of the tested plants were artificially inoculated by *E. amylovora* (OMPBO329,1 severe strain) and the other four were injected by SDW. The cutting wounding inoculation was used only in this trial because it was more handy and faster in performance than the direct injection.

2.2.1.5. Collection of leaf hyperspectral reflectance data

Hyperspectral data, on leaf level, were obtained using a hyperspectral acquisition system (Figure 11). This system includes: (I) a FieldSpec®3 spectroradiometer [Analytical Spectral Device (ASD), Boulder, USA] measures the spectral range 350–1830 nm with spectral sampling interval of 1.4 nm at 350–1050 nm and 2 nm at 1000–1830 nm. (II) A fore-optic contact probe and a leaf-clip holder (ASD, USA) supplied with an integrated light source (Halogen bulb), and linked to the spectroradiometer by (III) an optical fiber cable. (IV) An instrument controller (portable PC) to display and save the data. In practice, the acquisition system was warmed up for 45 minutes prior to the measurement to increase the quality of the acquired spectra. Then, the reflectance was standardized by calibrating the measurement of dark current to the measurement of a white reference (Spectralon®, Labsphere, USA). Next, the leaf-clip of 10 mm spot size was placed in the middle of the leaf beside the midrib to collect uniform measurements, and the integration time was set at 68 ms during measurements. Finally, the collected reflectance spectra were displayed, saved, and automatically interpolated from 1.6 nm to 1 nm by RS$^{3™}$ Spectral Acquisition Software (ASD, Boulder, USA).

Eight to 10 leaves from each studied plant were used for hyperspectral data acquisition, and in each observed leaf three to five reflectance measurements were properly taken and averaged. In the first glasshouse trial, the reflectance data were collected in three times at 24, 31, and 38 Days Post-Inoculation (DPI), and in the second trial the reflectance data were also collected for three times at 12, 14, and 22 DPI.

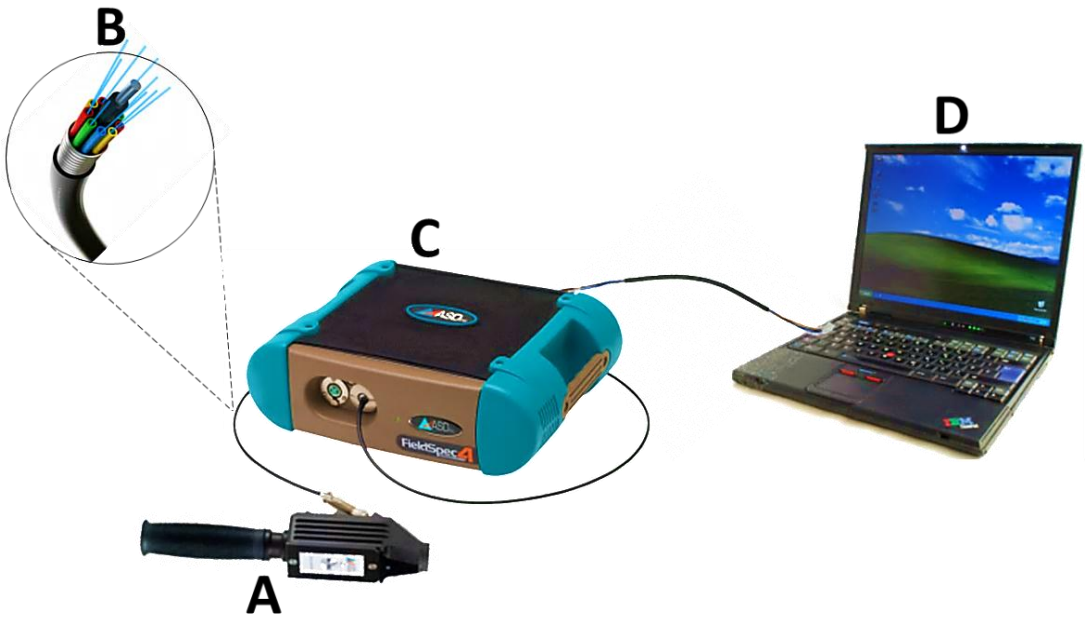


Figure 11 The hyperspectral acquisition system used for leaf reflectance measurements. (A) ASD leaf-clip supplied with a halogen bulb, (B) optical fiber, (C) ASD FieldSpec®3 spectroradiometer, and (D) Instrument controller.

2.2.2. Field trials

2.2.2.1. Area of study

Field hyperspectral data were obtained from apple and pear trees in the fertile Po valley, Ferrara (Emilia-Romagna). The pear orchard, located in Copparo town (Figure 12 A), contained 7-year-old homogeneous trees cv. Conference, whereas the apple orchard which is located in Formignana town contained 4-year-old trees cv. Pink Lady (Figure 12 B). The studied apple and pear trees were grown on vertical axis planting system, and only the apple orchard was covered by anti-hail net.

Based on the visual inspection and the help of well-trained inspectors of Emilia-Romagna phytosanitary services, vigilant scouts on healthy-looking and fire blight-symptomatic trees were carried out in the studied orchards in late June 2013. Next, several healthy-looking and fire blight-symptomatic trees were labeled with unique identification codes, and then geographical coordinates were obtained for each selected tree by a handheld GPS receiver (MobileMapper™CX Magellan). Thirty apple trees (15 healthy-looking, and 15 symptomatic) in Copparo, and 21 pear trees (10 healthy-looking, and 11 symptomatic) in Formignana were tested in these trials.

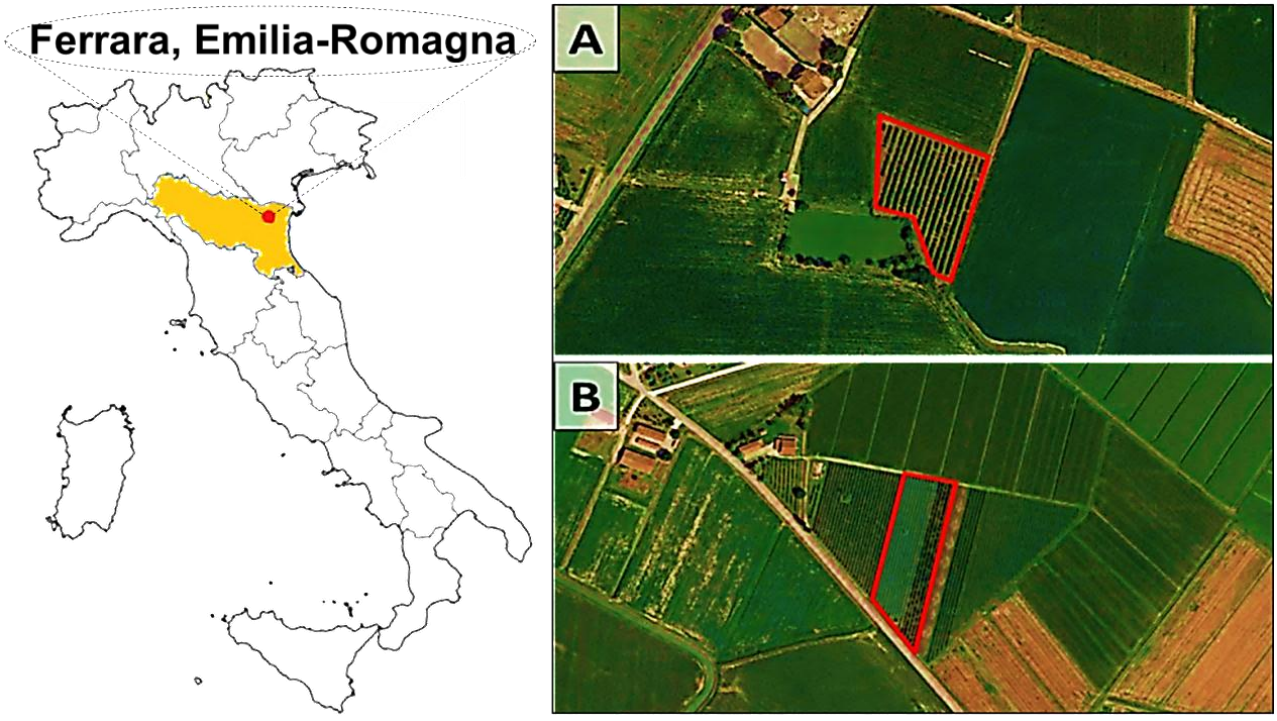


Figure 12 The location of the study area on the Italian map (left), and aerial view for the studied fields (right): (A) a pear orchard in Copparo, Emilia-Romagna, and (B) an apple orchard in Formignana, Emilia-Romagna.

2.2.2.2. Collection of spectral reflectance data:

Spectral reflectances in the field were obtained in late June 2013, on both leaf and canopy levels, by a FieldSpec® HandHeld2 portable spectroradiometer (ASD, Boulder, USA). This instrument has a 512-channel photodiode, and collects the spectral in the range 325–1075 nm with spectral-sampling intervals of 1.6 nm.

2.2.2.2.1. Leaf reflectance measurements

Leaf hyperspectral data were acquired from 30 apple trees (15 healthy-looking and 15 symptomatic) and 21 pear trees (10 healthy-looking and 11 symptomatic) by following the same methodology mentioned in glasshouse trials (section 2.2.1.5.). Five shoots from each tested tree were marked and collected in particular plastic bags, and then they were transferred to shade to obtain the measurements, and finally the plant residuals were discarded in the proper way. The reflectance data were collected at the same day in each field, in which four to five reflectance measurements were taken and averaged from each leaf in each collected shoot.

2.2.2.2.2. Canopy reflectance measurements

Canopy hyperspectral data were practically obtained by the acquisition system shown in Figure 13. This system is supplied by a fore-optic pistol grip of 25° Field of View (FOV). The reflectance measurement was optimized by calibrating the measurement of dark current to a white reference (Spectralon®, Labsphere, USA), then four reflectance measurements were consistently obtained and averaged from each tested tree at 544 ms integration time. The collected reflectance spectra were displayed, saved, and automatically interpolated from 1.6 to 1 nm by RS3™ Spectral Acquisition Software (ASD, Boulder, USA).

Figure 13 The hyperspectral acquisition system with the supporting system used for canopy reflectance measurements. This system comprised: (A) ASD fore-optic pistol grip, (B) aluminum tripod, (C) optical fiber, (D) ASD FieldSpec® HandHeld2 spectroradiometer, and (E) instrument controller.

Canopy reflectance generally depends on the properties of leaves, their amount and spatial distribution, direction of the light, the orientation of the sensor, properties of the soil underneath the canopy (background), and the atmospheric conditions (Cierniewski and Verbrugghe, 1997). The effects of all mentioned factors were significantly reduced by taking the measurements in cloudless days between 10 a.m. and 14 p.m., and by setting the pistol grip on nadir view at approximately 50 cm above the center of the tested canopy (Figure 14).

Figure 14 The specially designed tripod that carried the pistol grip to mount the fore-optic above the trees during canopy reflectance measurements.

Hyperspectral data, on the canopy level, were obtained only from the marked pear trees in Copparo (10 healthy-looking and 11 symptomatic). Unfortunately, no hyperspectral reflectances were obtained from the studied apple trees because the whole orchard was covered by anti-hail net (Figure 15). The studied pear plants were the same ones that used for leaf field trial.

Figure 15 The anti-hail net above the studied apple orchard that hampered the acquisition of the reflectance data from the canopies.

2.2.3. Data analysis

Discriminant Analysis (DA) is a broad term which is basically used to determine which continuous variable (or combination of variables) discriminates between two or more naturally occurring groups (Klecka, 1980). The determination of the optimal variables or predictors, which are the spectral regions in our case, contributes to the reduction of the dimensionality of the hyperspectral space which is extremely important to develop robust classification models using these predictors. DA is usually performed primarily by testing of discriminant functions sets, and then the classification. The real goal of DA is the classification which can be achieved by parametric or non-parametric methods. Parametric methods are commonly used when the studied dataset represents a sample from a multivariate normally-distributed population, and non-parametric methods are used when the multivariate distribution is different from the normal (Hand, 1997).

In the current study, the hyperspectral data were analyzed in three main steps: data collection and preprocessing, hyperspectral discrimination, and classification models (Figure 16). These steps will be illustrated in details through the next paragraphs. However, the first step was the datasets preparation by collecting the hyperspectral reflectance data, gathering all information about the studied plants (species, age, cultivars, inoculation, infection, and others), preprocessing the data, and Spectral Vegetation Indices (SVIs) calculation. The second step was the evaluation of the hyperspectral separability between the studied plants by applying a multivariate data reduction tool (Principal Components and Classification Analysis; PCCA) on the preprocessed datasets. The final steps were the development of two classification models using the Partial Least Square-Discriminant Analysis (PLS-DA) and the Artificial Neural Networks (ANNs) methods on two levels, the hyperspectral Reflectance Wavebands (RWs) and the SVIs, and the evaluation of the models performance.

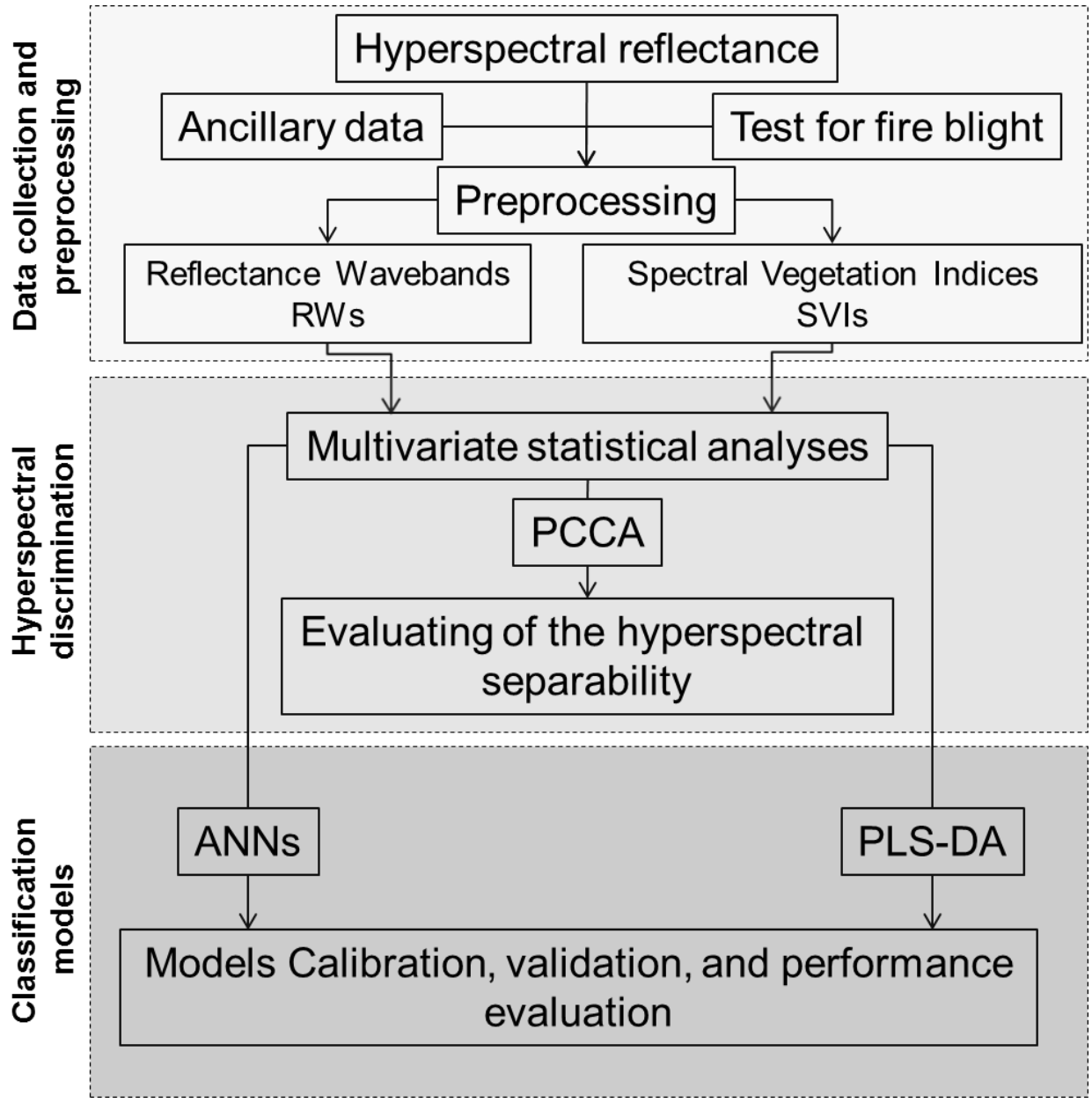


Figure 16 The general workflow of the performed research which mainly consists of three parts: I) data acquisitions and preprocessing, II) test hyperspectral separability, and III) developing PLS-DA and ANNs classification models.

2.2.3.1. Data preprocessing and calculation of Spectral Vegetation Indices (SVIs)

Hyperspectral data were preprocessed through ViewSpecPro® software ASD to be manageable. The collected data were preprocessed by cleaning irregular and distorted spectral signatures, converting the original format (.asd) to reflectance format (.ref), and converting them to text format (.txt). Then, by using a special algorithm (Post_Processing_Spectroradiometer), developed in MATLAB R2011b (the MathWorks, Inc., USA) environment (Santoro, 2011), all the converted reflectance spectra were smoothed and cleaned from noise by the Savitzky-Golay filter, and then they were organized separately in Microsoft Excel files. Finally, these Excel files were imported to STATISTICA 8 software (StatSoft, Inc., USA) to complete the preprocessing and start the statistical analysis. The hyperspectral reflectances between 400 nm and 1800 nm in glasshouse trials and between 400 nm and 1070 nm in field trials were solely considered and analyzed. These data were selected because the obtained hyperspectral reflectances, which extended beyond the mentioned ranges, were noisy and entangled due to the different instrumental errors and other environmental factor.

The processed hyperspectral data were also used to calculate 30 SVIs which are listed in Table 2. These SVIs were reported to have a strong correlation with plant biotic stress, biomass, water content, chlorophyll absorption, and other characteristics. The use of SVIs in spectral analysis is extremely important practice not only to enhance the hyperspectral features, but also to reduce the noise and the effects of atmospheric properties, background reflectance, and view geometry. Next, the computed SVIs values were standardized (Z-score normalization) to have a comparable ranges and distribution as illustrated in Figure 17. Then, statistical exploratory analyses were performed on the SVIs to check their skewness and kurtosis. Box and whisker plot tool and Grubbs test were then used to detect and eliminate the real outliers from the obtained data. Finally, data multivariate normality was tested by using the Henze-Zirkler's algorithm (1990) in MATLAB environment.

Table 2 The 30 Spectral Vegetation Indices (SVIs), reported to be correlated to plant biotic and abiotic stress, which were calculated, standardized, and used in this research.

Vegetation index	Reference	Algorithm
ARI: Anthocyanin Reflectance Index	(Gitelson et al., 2001)	=(1/R550 - 1/R700)
BGI2: Blue Green Pigment Index 2	(Zarco-Tejada et al., 2005)	= R450 / R550
ChlNDI: Chlorophyll Normalized Difference Index	(Gitelson and Merzlyak, 1994)	=(R750 - R705) / (R750 + R705)
DD: Double Difference Index	(le Maire et al., 2004)	=(R749 - R720) - (R701 - R672)
DPI: Double Peak Index	(Zarco-Tejada et al., 2003)	=(R688 × R710) / (R697)^2
FR: Fluorescence Ratio	(Lichtenthaler and Rinderle, 1988)	= R690 / R735
GI: Greenness Index	(Smith et al., 1995)	= R554 / R677
Lic2: Lichtenthaler indices2	(Lichtenthaler, 1996)	= R440 / R690
MCAI: Modified Chlorophyll Absorption Integral	(Laudien et al., 2003)	=(R545 + R750) / 2 × (750 - 545) - (∑R545: R750) × 1.423)
MCARI: Modified Chlorophyll (a + b) Absorption in Reflectance Index	(Daughtry et al., 2000)	=((R701 - R670) - 0.29×(R701-R550)) × (R701 / R670)
MSR670: Modified Simple Ratio 670,800	(Chen, 1996)	=(R800 / R670 - 1) / Sqrt(R800 / R670 + 1)
NBNDVI: Narrow-band normalized Difference vegetation index	(Thenkabail et al, 2000)	=(R850 - R680) / (R850 + R680)
NDVI: Normalized Differences Vegetation Index	(Rouse et al., 1973)	=(R800 - R670) / (R800 + R670)
NPQI: Normalized Phaeophytinization Index	(Barnes et al., 1992)	=(R415 - R435) / (R415 + R435)
PBI: Plant Biochemical Index	(Rama Rao et al., 2008)	= R810 / R560
PRI: Photochemical Reflectance Index	(Gamon et al., 1992)	=(R531 - R570) / (R531 + R570)
$PSND_a$: Pigment Specific Normalized Difference	(Blackburn, 1998)	=(R800 - R680) / (R800 + R680)
$PSND_b$: Pigment Specific Normalized Difference		=(R800 - R635) / (R800 + R635)
$PSND_c$: Pigment Specific Normalized Difference		=(R800 - R470) / (R800 + R470)
PSRI: Plant Senescence Reflectance Index	(Merzlyak et al., 1999)	=(R680 - R500) / R750
PWI: Plant Water Index	(Penuelas et al., 1993)	= R900 / R970
RDVI: Renormalized Differences Vegetation Index	(Roujean and Breon, 1995)	=(R800 - R670) / Sqrt(R800 + R670)
REP: Red Edge Position	(Guyot et al., 1988)	= 700 + 40 × (((R670 - R780) / 2) - R700) / (R740 - R700)
RVSI: Red-Edge Vegetation Stress Index	(Merton, 1993)	=(R714 + R752) / 2- R733
SIPI: Structure Insensitive Pigment Index	(Penuelas et al., 1995)	=(R800 - R445) / (R800 + R680)
SPAD	(Santoro, 2012)	=log10(R940 / R650)
SPRI: Simple Ratio Pigment Index	(Penuelas et al., 1995)	= R430 / R680
VI_{opt}: Optimal Vegetation Index	(Reyniers et al., 2006)	= (1 + 0.45) × ((R800)^2 + 1) / (R670 + 0.45)
Vog1: Simple Ratio 740/720 Voglemann Indices 1	(Vogelmann et al., 1993)	= R740 / R720
YI: Yellowness Index	(Richardson et al., 2002)	= -(R580 - 2 × R624 + R668) / (0.044^2)

R: the reflectance value at the reffered wavelength.

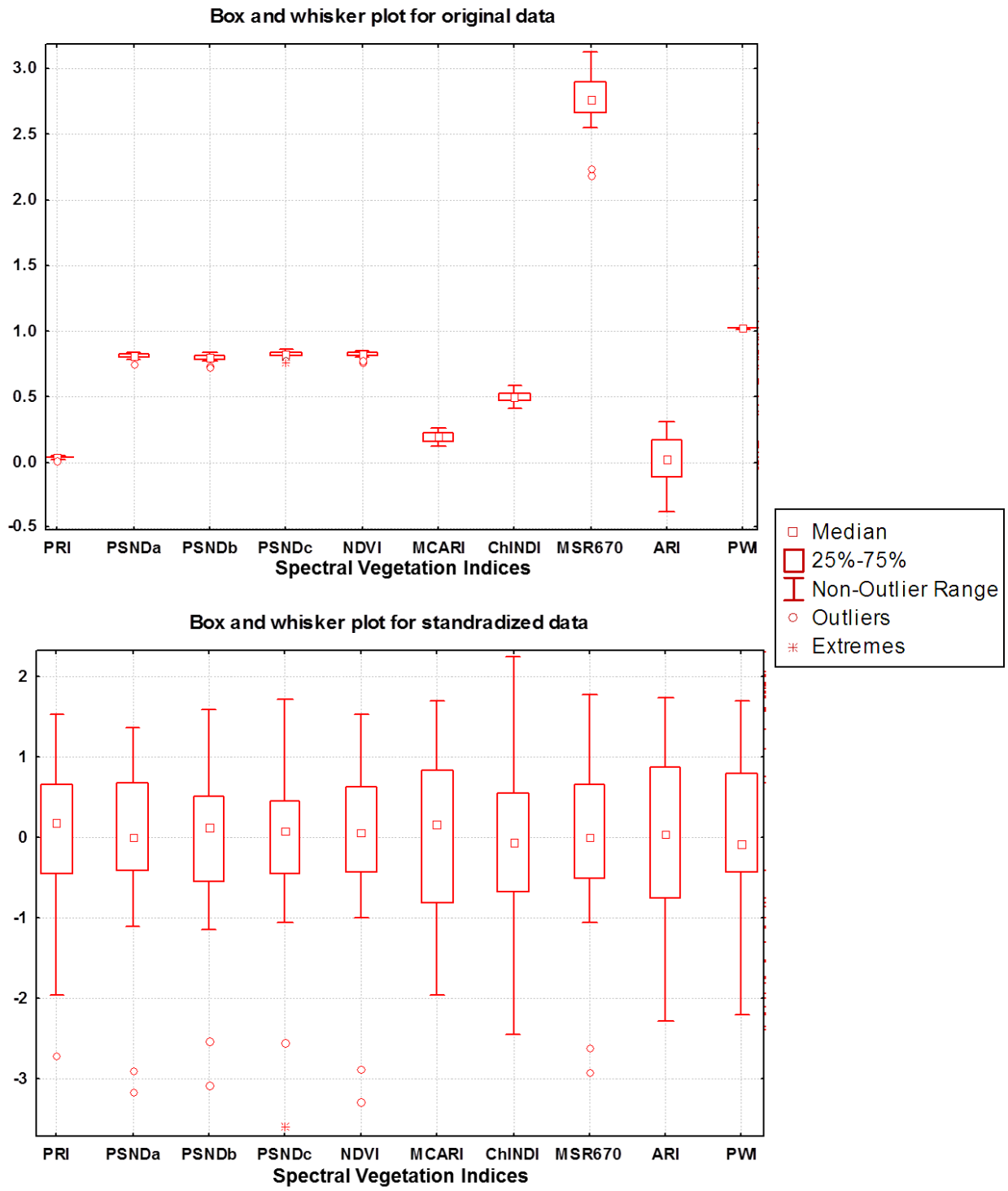


Figure 17 A box and whisker plot for 10 Spectral Vegetation Indices (SVIs) on original (a) and standardized data (b); the upper and the lower lines in the large boxes are the 75th and 25th percentiles of the samples, respectively. The small boxes inside are the medians of the samples, and the extending lines from the boxes are called whiskers, and the values beyond those whiskers are detected as potential outliers.

2.2.3.2. Multivariate classification analysis

This type of multivariate analysis is increasingly used to develop mathematical recognition models capable of assigning the studied cases (samples) to predefined classes (groups). Furthermore, the developed classification models can predict the class of unknown samples based on a set of measurement (Ballabio and Todeschini, 2008). Multivariate classification analyses are used in several fields such as food science, pharmaceutical chemistry, and remote and proximal sensing. Classification methods are divided in two main categories. The first one, linear or non-linear, is based on the form of the decision boundaries among classes. The second category, parametric or non-parametric, is the multivariate probability distribution. If the studied dataset is multivariate normally distributed, parametric methods like Discriminant Analysis and Nearest Mean Classifier (NMC) should be used for the classification, but if the dataset is not normally distributed, non-parametric methods such as PLS-DA, ANNs, and Classification and Regression Tree (CART) should be used for classification.

2.2.3.2.1. Evaluating the hyperspectral separability of healthy and infected plants using Principal Components and Classification Analysis (PCCA)

Principal Components Analysis (PCA) has been widely used as a data reduction tool. This technique maintains the total variance and minimizes the mean square errors in the same time. Technically, PCA transfers the original dataset into a new smaller set contains uncorrelated variables that reflect most of the information in the original dataset. The new dataset is simpler to manage and easier to interpret (Jolliffe, 2002). In our study the obtained hyperspectral data had an astronomical number of values, thus PCA was implemented on the whole acquired spectral data (400 – 1070 nm) using the Principal Components and Classification Analysis (PCCA) module in STATISTICA 8 software (StatSoft, Inc., USA); this module can also be used successfully for classification variables or cases so any relations among them can be underlined. PCCA has another graphical advantage of providing visual aid for the classification of variables and cases, so one can have an idea about the capability of the collected hyperspectral data to separate the studied plants in different classes.

2.2.3.2.2. Classification models

In this research the statistical classification models were developed on two levels the Reflectance Wavebands (RWs) and the Spectral Vegetation Indices (SVIs).

On the level of the SVIs, the classification models were developed using the standardized values, whereas on the RWs level the hyperspectral data undergone a data reduction process before calculating the models. This reduction was because the collected hyperspectral data provided a vast amount of information but most adjacent wavebands were redundant and highly correlated. Moreover, the risk of overfitting in the classification models increases significantly when elaborating such huge number of multicollinear variables (Hand, 1997). In our study, Partial Least Squares (PLS) approach was used to reduce the information redundancy, remove interferences from spectral regions, and provide a clear idea about the relevant information in

different spectral subdivisions. Practically, a special MATLAB algorithm called synergy interval-PLS (si-PLS), in the iToolbox for MATLAB (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004) was used in which the collected hyperspectral datasets were split into few spectral intervals, 50 nm each, and all possible combinations of two, three, or four intervals were calculated. Eventually the combination intervals that produce the least Root Mean Square Error of Cross Validation (RMSECV) were selected and used for building PLS-DA and ANNs classification models.

The collected datasets in each trial (apple and pear) were divided on the basis of fire blight infection into two classes (healthy and infected). All considered samples (cases) were also divided randomly into training and evaluation subsets (Table 3). Each training subset contained ~ 75% of the total number of the samples, whereas each evaluation subset contained ~ 25% of the samples. The training subsets were used to calibrate the classification models, and the evaluation subsets were then used to evaluate the predictive capability of the calibrated models.

Table 3 The divided datasets in each apple and pear trial into training (about 75% of total number of samples) and evaluation subsets (about 25% of the total number of samples) to be used for models calibration and performance assessment, respectively.

	Trials	Subsets	Healthy	Infected	Total
Apple	Glasshouse	Training	33	95	128
		Evaluation	9	24	33
		Total	42	119	161
	Open field	Training	15	20	35
		Evaluation	4	5	9
		Total	19	25	44
Pear	First Glasshouse	Training	45	65	110
		Evaluation	11	16	27
		Total	56	81	137
	Second Glasshouse	Training	75	104	179
		Evaluation	19	26	45
		Total	94	130	224
	Field leaf/canopy	Training	9	12	21
		Evaluation	3	4	7
		Total	12	16	28

The developed classification models were evaluated based on Non-Error Rate (NER), Error Rate (ER), Sensitivity (Sn), Specificity (Sp), and percentage of Not Assigned (NA) samples which are the most useful classification parameters. In this research, the samples were classified into only two categories (binary classification) healthy and fire blight-infected which can be labeled as Negative (N) and Positive (P) samples, respectively. Therefore, any classification model will generate four outcomes:

1. True Positive (TP) when the model classifies the sample as positive and the real class is positive;
2. True Negative (TN) when the model classifies the sample as negative and the real class is negative;
3. False Positive (FP) when the model classifies the sample as positive and the real class is negative;
4. False Negative (FN) when the model classifies the sample as negative and the real class is positive.

Consequently, the Classification performance parameters in our case (binary classification) are defined as follows:

- $\text{Non Error Rate (NER)} = \frac{\text{TP} + \text{TN}}{\text{n}}$
- $\text{Error Rate (ER)} = \frac{\text{FP} + \text{FN}}{\text{n}} = 1 - \text{NER}$
- $\text{Sensitivity(Sn)or True Positive Rate (TPR)} = \frac{\text{TP}}{\text{TP} + \text{FN}}$
- $\text{Specificity (Sp)} = \frac{\text{TN}}{\text{TN}+\text{FP}}$
- $\text{False Positive Rate (FPR)} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{Specificity}$

Where,

n: the number of the classified samples.

NER is the percentage of the correctly assigned samples, whereas the percentage of the wrongly assigned samples is expressed by ER. These two parameters describe the performance of the calculated classification models. However, the classification quality of the models at the level of each class should also be evaluated. Models sensitivity and specificity can be used for that purpose, in which the sensitivity describes the capability of a model to recognize the samples of each class correctly, and the specificity describes the capability of a model to reject the samples from other classes.

Another effective tool can be used to analyze the classification results and select the optimal classification model. The Receiver Operating Characteristic (ROC) curve is a two-dimensional graphical plot of FPR (1 - specificity) and TPR (sensitivity) parameters, on *x* and *y* axes respectively as shown in Figure 18. Each classification model has a particular ROC curve, in which each point represents sensitivity-specificity values corresponding to a specific threshold. A random classification results in points along the diagonal line from the left bottom up to the upper right corner. The perfect model has a ROC curve that passes through the upper left corner (maximum sensitivity and specificity). Therefore, the closer the ROC curve to the upper left corner, the higher the accuracy of classification model (Zweig and Campbell, 1993).

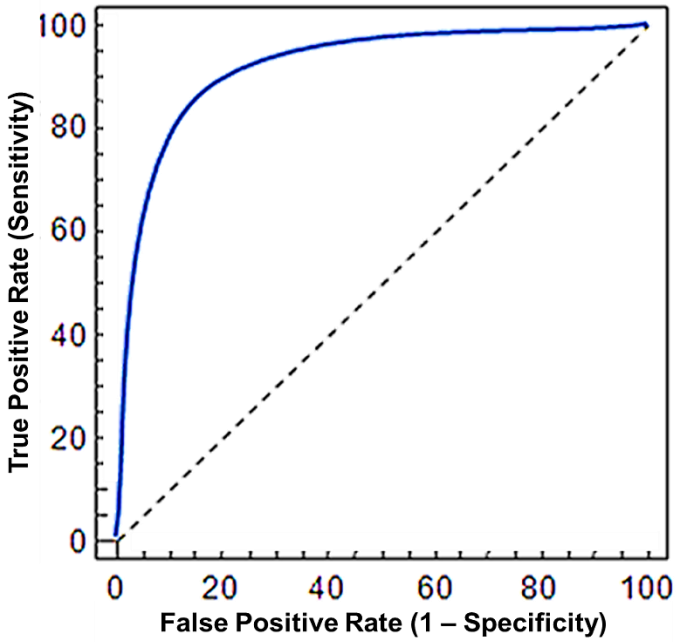


Figure 18 A Receiver Operating Characteristic (ROC) curve created by plotting true positive (sensitivity), and false positive rates (1 - specificity) of an example of a classification model at different thresholds.

2.2.3.2.2.1. Partial Least Squares Analysis-Discriminant Analysis (PLS-DA)

Partial least Square (PLS) is a bilinear calibration method that works like PCA. PLS was originally designed to study the regression, in which the multidimensional space is transformed from a large number of collinear variables into few non-correlated factor variables, and the background effects and the noise are explained by the less important factors (Geladi and Kowalski, 1986). PLS algorithm was lately modified to be used for classification purposes in several fields. Recently, scientists showed that PLS-DA corresponds with the inverse least squares approach to Linear Discriminant Analysis, and produces the same results with the advantages of noise reduction and latent variables selection (Barker and Rayens, 2003).

The classification toolbox for MATLAB Version 3.1 (Ballabio and Consonni, 2013) was used for the optimization, the calculations, and the performance evaluation of all PLS-DA models. The first step in the optimization was to select the number of the Latent Variables (LVs) due to the great impact of their number on the calculation and the calibration of PLS-DA models. Therefore, to enable the assessment of the optimal complexity model, the number of the LVs must be optimized before the calculations. Practically, venetian blinds cross validation is the appropriate procedure because of the distribution nature of the classes (healthy and infected) in our samples. This method was used for the selection by dividing samples in cross validation groups, and then each group was removed from the dataset, one at a time. Each time, the model was recalibrated by using the remaining samples, and consequently used to predict samples of the cross validation group.

Additionally, a reliable number of cross validation groups needs to be selected before the calculation. The selection of many cross validation groups overestimates the prediction power of the model, whereas selecting only few groups underestimates that power. It is advisable therefore to select a high number of cross validation groups when the sample dataset is small, and to select a smaller number of groups when the dataset is relatively large. Since the datasets in our research were fairly moderate sized, two cross validation runs were performed for the model with 10, and five groups respectively. After that the resulted ER and NA percentage in each run were compared. Then, the number of LVs corresponding to the minimum ER and the minimum number NA samples was selected to be analyzed. Finally, PLS-DA models were calculated from the training datasets, of each trial, on the basis of the optimal number of the LVs and cross validation groups.

2.2.3.2.2.2. Artificial Neural Networks (ANNs)

ANNs are learning strategies inspired by biological neuron networks. The artificial neurons (nodes), which are supposed to mimic biological ones, in ANNs are connected with each other by weighted links (synapses). Since the late 1980s, researchers have developed many types of ANNs to solve supervised and unsupervised problems in pattern optimization, classification, and prediction (Reby et al., 1997). However, counter propagation or back-propagation algorithm is the most used ANNs (Rumelhart and McClelland, 1986). The artificial neurons in Counter Propagation-Artificial Neural Network (CP-ANN) are organized in three layers (input, hidden, and output) and they send their signals forward, and then the errors are propagated backwards. CP-ANNs are learned by training not programing, and perform perfectly when the relationship between the input and the output is non-linear.

The Kohonen and CP-ANN toolbox for MATLAB Version 3.6 (Ballabio and Vasighi, 2012) was used in this research to optimize, calculate, and evaluate the classification performance of the developed ANNs models. Like PLS-DA models, and as a supervised classification tool, ANNs require an optimization step before the model calculation. The best strategy for that is the use of the integrated Genetic Algorithm (GA) inside the CP-ANN toolbox (Ballabio et al., 2011) to select the optimal number of neuron and epochs (iterations). Venetian blinds cross validation procedure and two cross validation runs were used for the optimization with 10 and five groups, and then the optimization results were visualized in bubble plots. These plots contain different bubbles (network architectures) with their relative frequency in the GA runs and the mean of the optimization criteria (fitness function). The size and the color of the bubbles express respectively the network size (number of neurons) and the number of epochs. The stronger color and the bigger size the bubble, the larger the number of neurons and epochs of the corresponding network. The network in the upper right part of the bubble plot was selected and considered the best ones because they had both high relative frequency, and high predictive performance. Finally, ANNs models were calculated from the training datasets, of each trial, on the basis of the optimal numbers of neurons, training epochs, and cross validation groups.

2.3. Results and discussions

2.3.1. Artificial inoculation and laboratory tests

The most problematic issues for fire blight researchers in the past were the plant materials to be used for the artificial inoculation and how to inoculate the plants. Years later, it was proved that mature trees are considered as the most representative plant materials, and young seedlings are certainly the most relevant materials especially for laboratory experiments (Vanneste and Eden-Green, 2000). Artificial inoculation can be done without wounding by spraying the bacterial suspension directly on the flowers, but in the absence of flowers it needs wounding or low infection will result. On the other hand, artificial inoculation by wounding results in high levels of infection, but pushing the bacteria inside the plant is more difficult and sometimes reflects misleading results.

In the current study young apple and pear seedlings and two artificial inoculation methods (cutting wounding and direct injection) were used. Basically, no real differences were found between the two inoculation methods. However, in the first glasshouse trial fire blight symptoms and signs (Figure 19; blackened areas on leaves near the midrib, wilting, necrotic leaves, bacterial ooze on the petioles and the stems, and lately the shepherd's crook symptom) were visible and clear on few inoculated pear at five DPI and the rest plants were asymptomatic, while on the inoculated apple, some mild fire blight symptoms started to develop on one inoculated apple at 10 DPI, and these symptoms became evident at 39 DPI. This difference in showing the symptoms can be explained by the higher susceptibility of pear species to fire blight compared with apple species. In the second glasshouse trial, which was performed on pear only, fire blight symptoms appeared faster, more sever, and evident than in the first trial because the studied pear cultivars in the second trial were older and vigor, and the atmospheric conditions inside the glasshouse were relatively more conducive compared with the first trial. The conducted identification tests in laboratory confirmed that the isolated bacteria from all inoculated plants (apple and pear) in all glasshouse trials were *E. amylovora*.

No differences between the two artificial inoculation methods were revealed in terms of symptoms, though the cutting wounding method was faster and easier to perform than the direct injection. So, only the cutting wounding inoculation method was used in the second glasshouse trial.

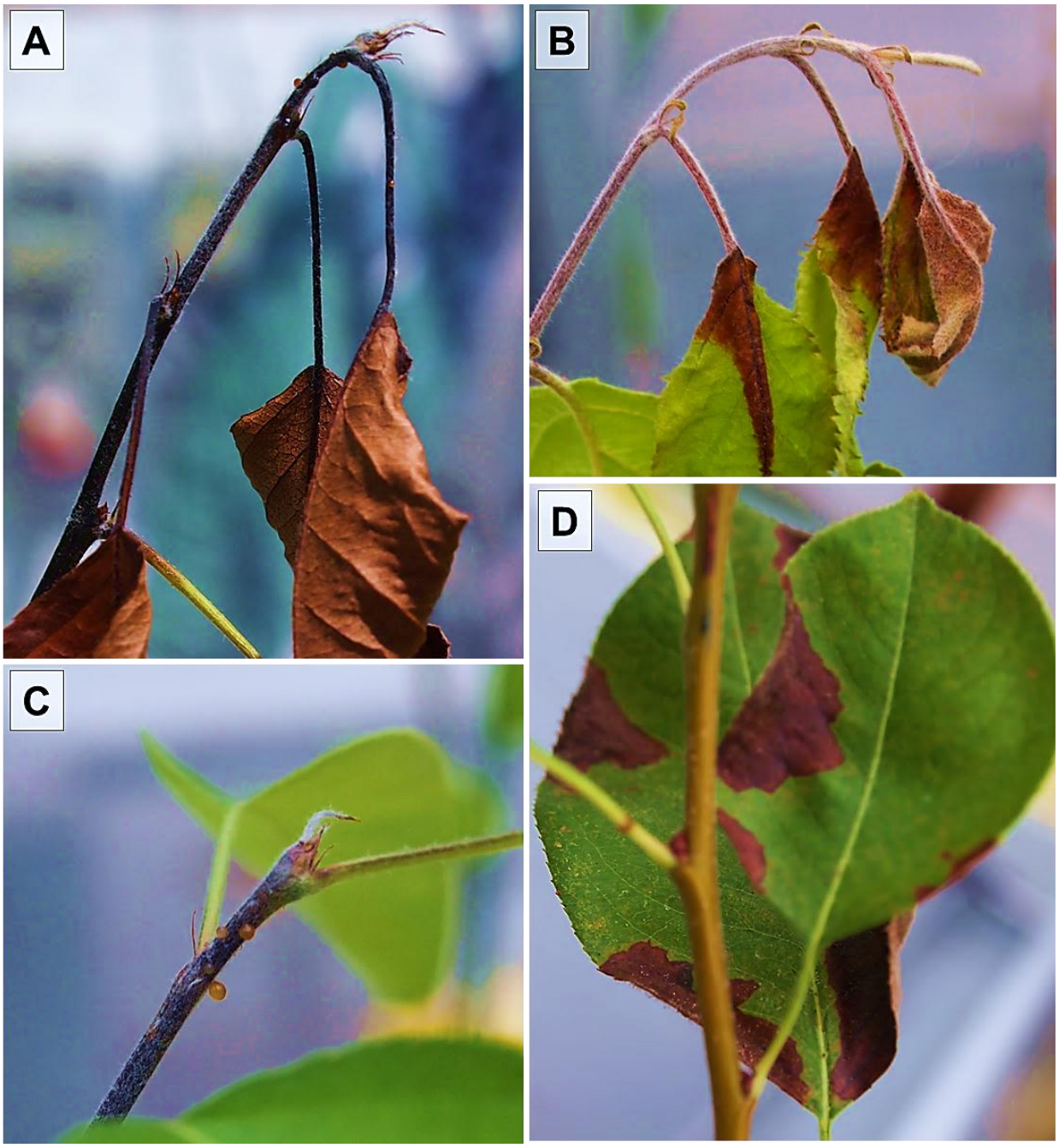


Figure 19 Fire blight symptoms and signs resulted from the artificial inoculation in apple and pear seedlings in the first glasshouse trial. These symptoms and signs include: (A) shepherd's crook and wilted leaves, (B) tip wilting, (C) bacterial ooze, and (D) blackened areas on the leaves.

2.3.2. Hyperspectral discrimination

2.3.2.1. Glasshouse trials

In the current research, hyperspectral trials in controlled conditions are indispensably important to reduce the complexity which is resulted from atmospheric conditions, the Bidirectional Reflectance Distribution Function (BRDF), and other factors. Otherwise, the high heterogeneity and the interference factors in open field trials may lead to misleading results.

The spectral curves acquired on the leaf level were assessed under constant conditions (light and temperature), so the number of excluded spectral signatures from these experiments was quite small. However, the preprocessing to smooth the spectrum and to reduce noise was necessary.

The average spectral signatures of the healthy and fire blight-infected apple and pear in the first glasshouse trial are depicted in Figure 20. The used hyperspectral acquisition system covered three main regions of the electromagnetic spectrum (400–1800 nm), the visible (400–700 nm), the NIR (700–1350 nm), and a part of the Shortwave Infrared (SWIR; 1350–1800 nm). Even though the obtained curves had the same shape which was the typical vegetation reflectance pattern, several differences of varying magnitudes can be noticed between the studied plants.

Firstly, in the visible range of the spectrum, apple plants reflected more amount of energy than pear and that difference was clear in the Blue region (400–500 nm). In addition, infected pear showed higher values than their control plants particularly around the Green peak (550 nm; Figure 20); the situation in that region for apple was inversed where the healthy plants had higher reflectance values than the infected ones. Although apple and pear reflected different amount of energy around the green peak due to their species and stress status, the high reflectance of the infected pear is unnecessarily linked to fire blight infection. Unless the plants are in late severe fire blight infection or mixed infection, they commonly do not show evident yellow symptoms.

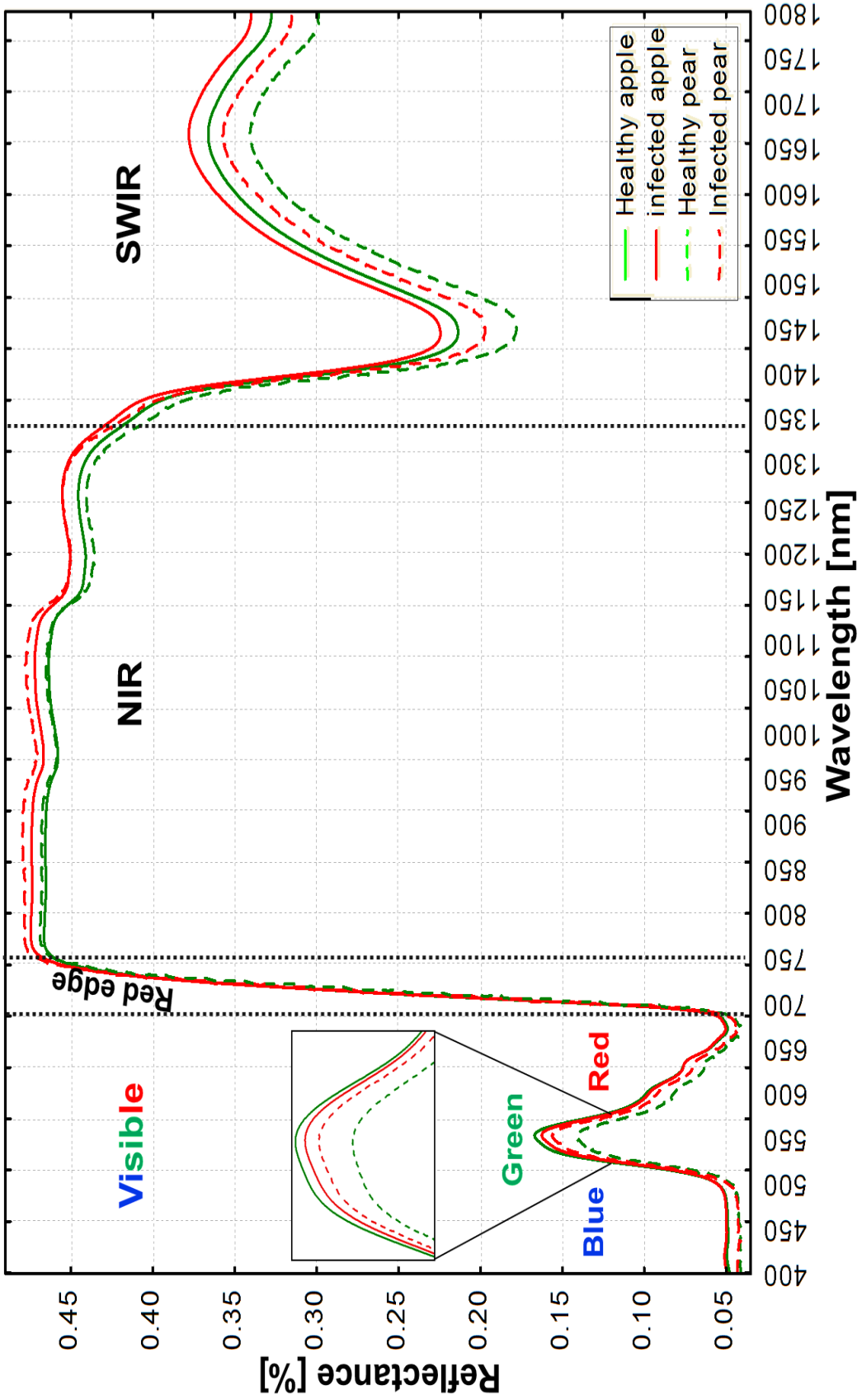


Figure 20 Average spectral curves of all compiled leaf reflectances, in the range 400–1800 nm, from healthy and *E. amylovora*-inoculated apple and pear in the first glasshouse trial. The rectangular window in the left side of the plot shows a scoop of different reflectances around the green peak.

Next, in the Red Edge region (700–750 nm), where the reflectance sharply increased until reaching NIR and SWIR regions, spectral reflectance curves from apple and pear fluctuated (healthy and infected) and did not follow any specific pattern. Lastly, the mean reflectance curve of inoculated apple and pear had higher reflectance in the NIR and the SWIR than in the controls (Figure 20). The higher reflectance in these regions, which contain four atmospheric water absorption bands (at 975, 1175, 1450, and 1750 nm), confirmed that the infected plants had undergone a strong water stress. Any natural vegetation usually gives low reflectance in the visible region compared with the NIR and the SWIR regions due to: (I) the high absorbance of leaf pigments and chlorophyll in the visible region, (II) the low absorbance of subcellular particles in the NIR and the SWIR, and (III) the light scattering in the mesophyll (Gausman, 1977).

The differences between the healthy and the infected spectral signatures are important, but they are not the conclusive evidence of the hyperspectral discrimination possibility between these plants. The application of multivariate discriminant analysis was therefore necessary to check the capacity of the hyperspectral data to separate between the healthy and the infected plants.

PCA is unsupervised dimensional reduction method, though Principal Components and Classification Analysis (PCCA) module was applied to the collected hyperspectral reflectance for mainly two goals. The first one was to reduce the huge number of variables, or what is called the data compression, and the second goal was to extract the main features for classification purposes. Practically, PCCA was used to investigate the separability between the healthy and the fire blight-inoculated apple and pear plants. Practically, only the collected reflectance in the range 400–1399 nm was analyzed because the used PCCA module is incapable of analyzing more than 999 variables.

In the first glasshouse trial, the first two Principal Components accounted ~ 90% of the total sample variation in apple, while in pear they accumulated more than 82% of the total variance (Figure 21). In both species, apple and pear, the majority of the infected samples had negative values on the first vector (PC1), whereas the majority of the healthy samples had positive values on the same vector. On the other hand, the infected and the healthy samples in apple and pear had approximately similar negative and positive values on the second vector (PC2). In the second glasshouse pear trial, the first two vectors explained more than 94% of the variation (Figure 22). Like the first trial, the majority of the infected plants had negative values and the majority of the healthy ones had positive values on PC1. Most importantly, healthy and infected samples gathered in relatively separate groups in each trial as shown in Figure 21 and 22.

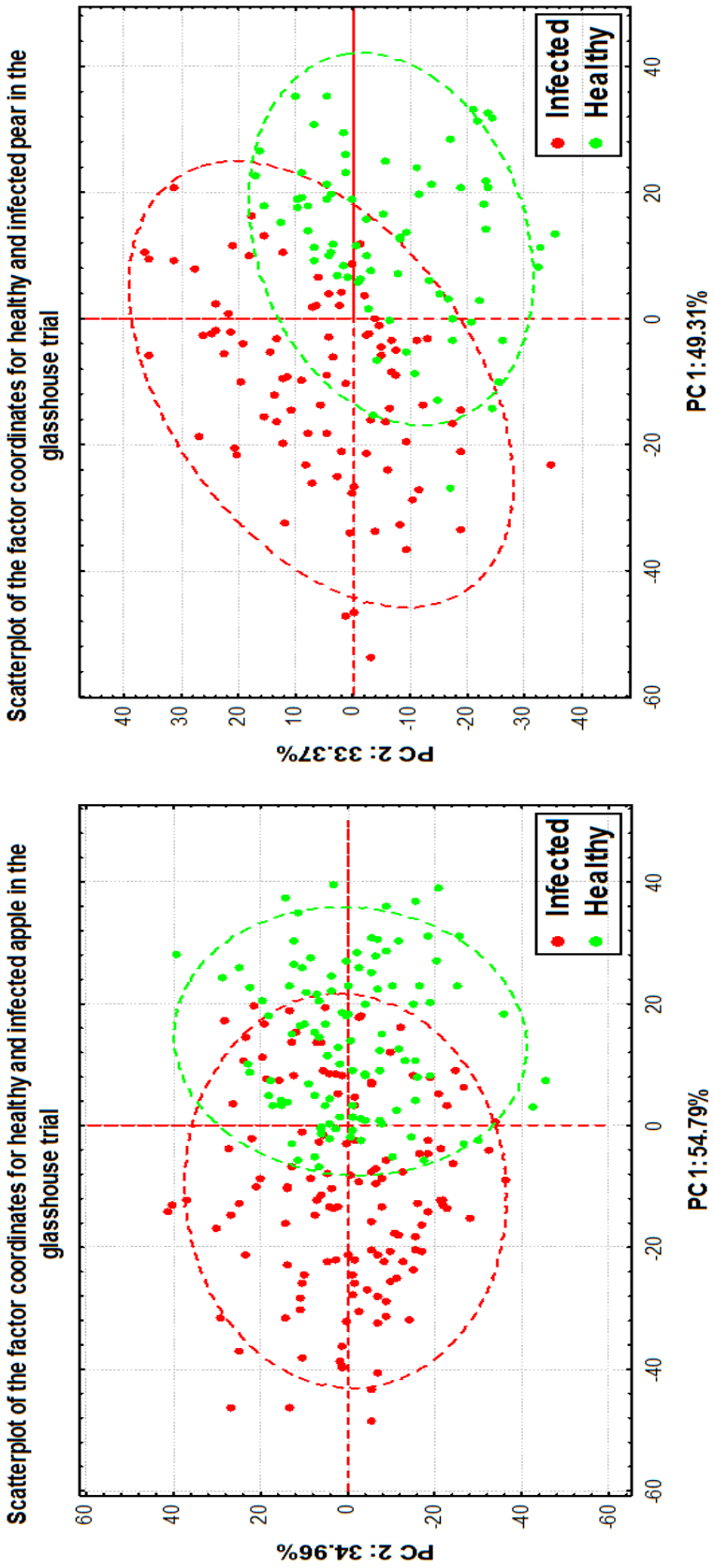


Figure 21 The scatterplots of the first two Principal Components (factors) scores resulted from applying PCCA on healthy (green points) and fire blight-inoculated (red points) apple (left plot) and pear (right plot) in the first glasshouse trial.

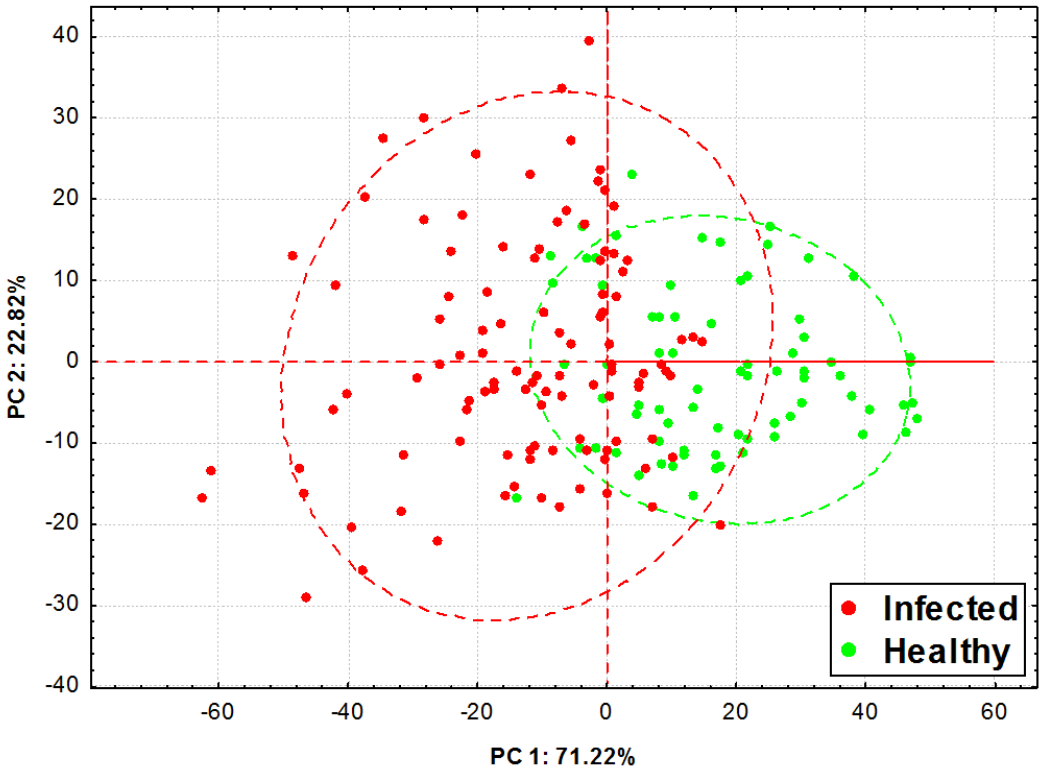


Figure 22 The scatterplot of the first two Principal Components (factors) scores resulted from applying PCCA on healthy (green points) and fire blight-inoculated pear (red points) in the second glasshouse trial.

2.3.2.2. Field trials

Hyperspectral data were collected from healthy-looking and fire blight-symptomatic apple and pear in open field trials. Reflectance data were obtained in the range 400–1070 nm on two levels. The first one was on the leaf level from both apple and pear, and the second level was the canopy but only from pear. The average reflectance spectra of leaves (apple and pear) and canopies (only pear) are shown in Figure 23.

The canopy spectral curves were quite different from those produced by leaves. As depicted in Figure 23, the canopy reflectance values were generally lower than those of leaves in all recorded spectrum regions. The used hyperspectral acquisition system during data collection can explain that difference in reflectance. For leaf measurements, the data were collected by a leaf-clip tool supplied by a stable light source and placed on a tiny distance (~ 3 mm) from the surface of the leaf, whereas a pistol grip placed on large distance (~ 2 m) from canopy was used for the measurements. Additionally, the canopy measurements are more sensitive to several interference factors like the atmosphere and the soil color.

Other significant differences were observed in the visible and the NIR regions of spectrum (Figure 23). In the visible region, leaf spectral reflectance of the symptomatic apple was higher than that of the healthy-looking trees, whereas the leaf reflectance of the healthy-looking pear trees was lower than the symptomatic ones in the Blue portion of spectrum, lower in the Green region, and nearly identical in the Red region. As explained previously in the glasshouse trials, the reflectance around the Green peak is unnecessarily related to fire blight infection. However, chlorophyll pigments absorb the light most strongly in the Blue region so the high reflectance of the symptomatic plants was a strong evidence for plant stress. Besides, the stress on symptomatic pear trees can be noticed clearly in the slope of the Red Edge region, which is normally seen between the visible and the NIR regions. As shown in Figure 23, the sudden transition in symptomatic pear was shifted towards longer wavelengths, and in one severely infected tree the Red Edge slope entirely disappeared (data not shown). In the NIR region, the average canopy reflectance the symptomatic pear was lower than the healthy-looking ones, and that difference was even greater than in the visible region. Furthermore, apple leaf reflectance was higher than leaf and canopy reflectances of pear in this region.

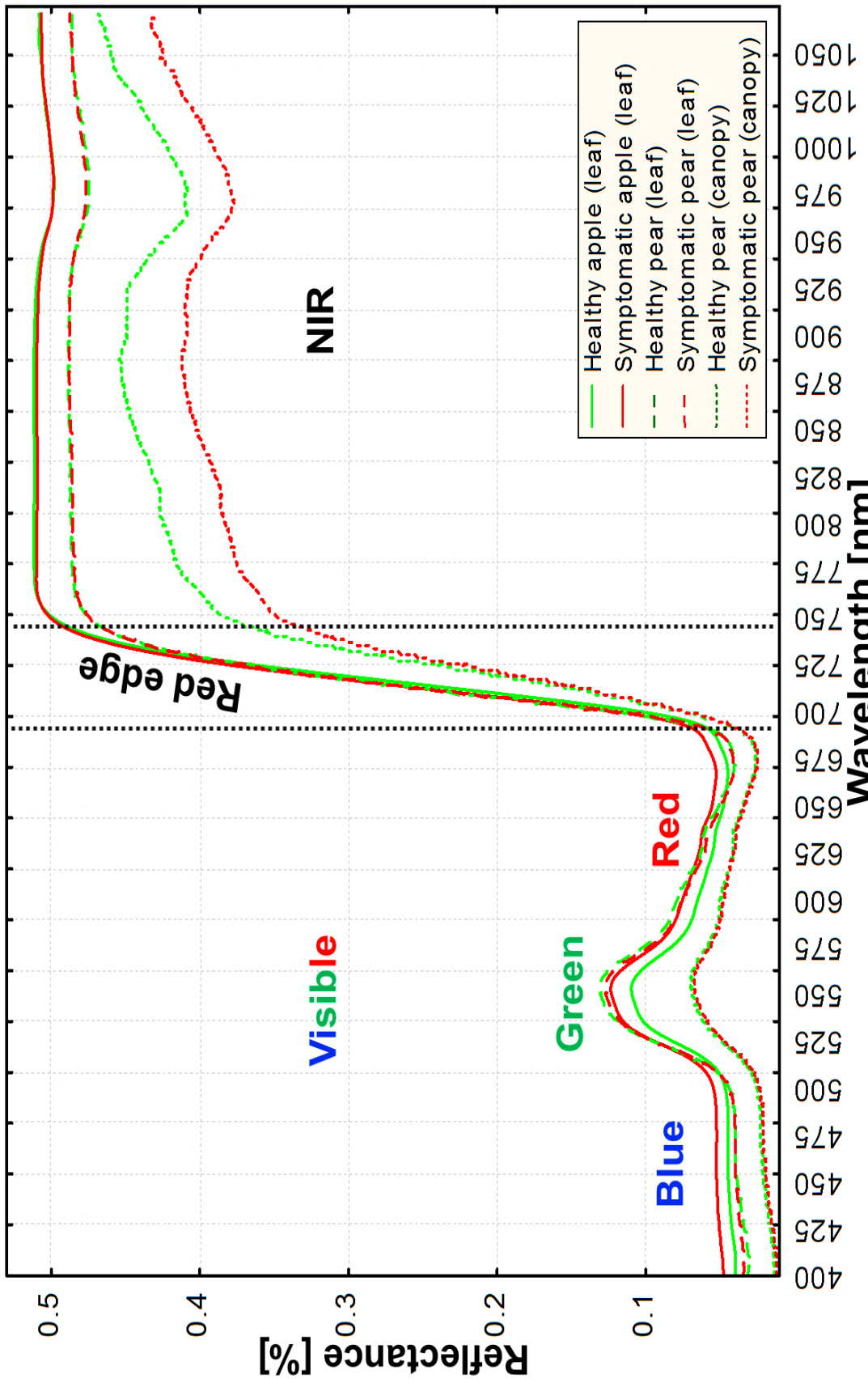


Figure 23 Average curves of all compiled leaf and canopy reflectances, in the range between 400 nm and 1070 nm, from healthy-looking and fire blight-symptomatic apple and pear in the open field trial.

The spectral signatures collected in the field from canopies are, as expected, difficult to be acquired and appear dissimilar from those of single leaves collected under controlled conditions. The interference of weather conditions and the reflectance from soil and other vegetation change the canopy spectra and even complicate the subsequent analysis. However, PCCA method was used to investigate the capacity of the collected hyperspectral data (400–1070 nm) in separate the healthy-looking plants from the fire blight-symptomatic ones.

The first couple of Principal Components accounted more than 92% of the total variation in apple leaf trial, and nearly 84% in pear leaf trial (Figure 24), while in pear canopy trial they accumulated more than 99% (Figure 25). Despite the small number of the tested samples and the relative overlapping, the separation between healthy and symptomatic apple and pear was still evident.

As a results, the separation between the inoculated and the healthy apple and pear in all glasshouse trials was obvious in spite of the relatively small overlapping between the tested samples. At this point, it was statistically proved that the collected hyperspectral reflectance data were capable of discriminate between healthy and fire blight-infected apple and pear, and most importantly was the possibility of proceeding in this study to develop classification models by using the non-parametric methods.

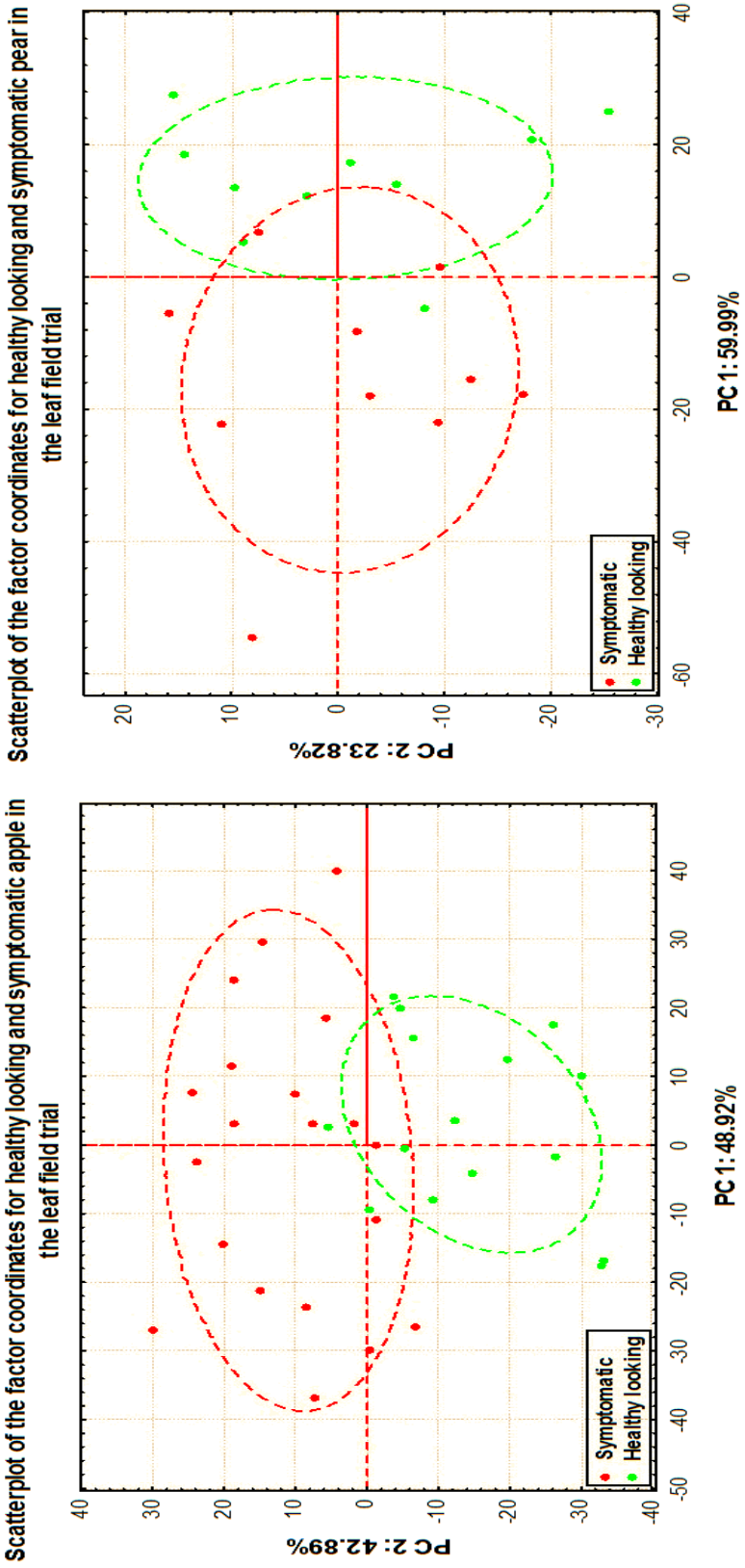


Figure 24 The scatterplots of the first two Principal Components (factors) scores for healthy-looking (green points) and fire blight-symptomatic (red points) apple (left plot) and pear (right plot) in leaf field trial.

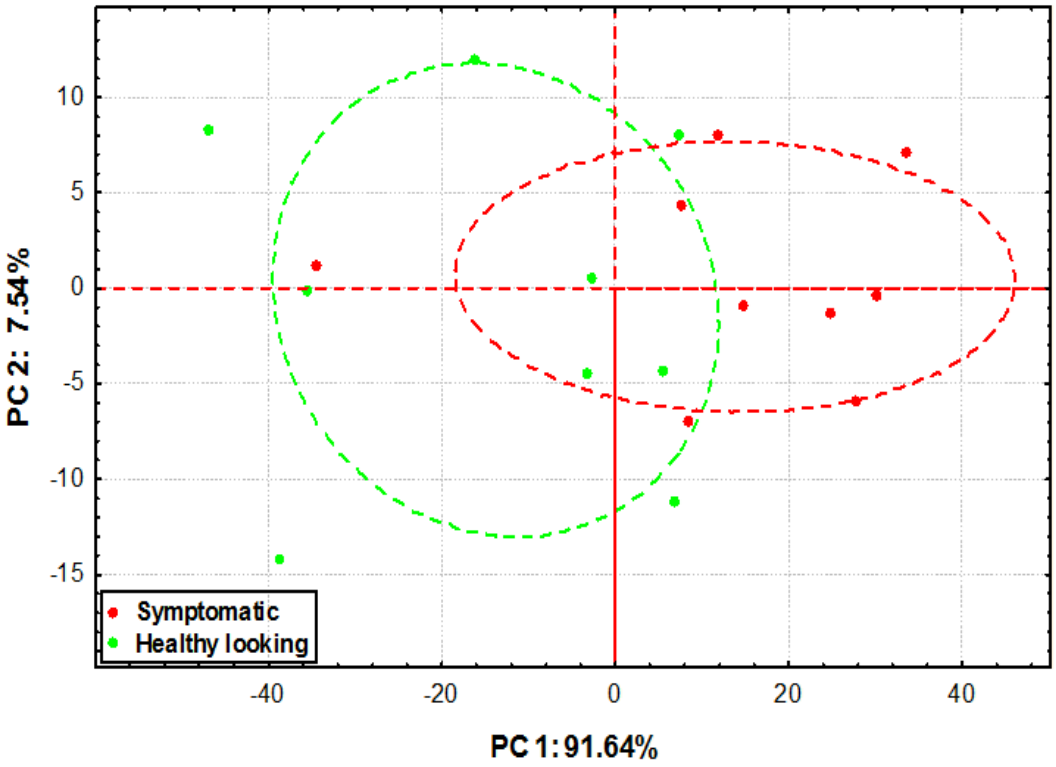


Figure 25 The scatterplot of the first two Principal Components (factors) scores for healthy-looking (green points) and fire blight-symptomatic pear trees (red points) in canopy field trial.

2.3.3. Classification models

Although several wavelengths showed a univariate normal distribution (Shapiro-Wilk's test), the Henze-Zirkler's test confirmed that the studied datasets were not multivariate normally distributed. Therefore, non-parametric statistical approaches are the best techniques to differentiate between the hyperspectral data of the healthy and the diseased plants.

Before starting with the optimization step, of PLS-DA and ANNs models, the number of the compiled hyperspectral variables (RWs 400–1800 nm in glasshouse trials, and between 400 nm and 1070 nm in field trials) was reduced to avoid the overfitting problem. This step was performed by using the synergy interval-PLS (si-PLS) algorithm integrated in the iToolbox for MATLAB (Nørgaard *et al.*, 2000; Leardi and Nørgaard, 2004). Figure 26 illustrates an example of the selected combination of intervals, which produced the least RMSECV, in apple glasshouse trial. In this glasshouse trial si-PLS was applied to full spectral range (400–1800 nm) to calibrate the global model, and the noisy spectral information was removed to reduce the number of variables and improve the prediction accuracy. The selected and used intervals in the rest of the glasshouse and field trials in apple and pear are show in Annex I.

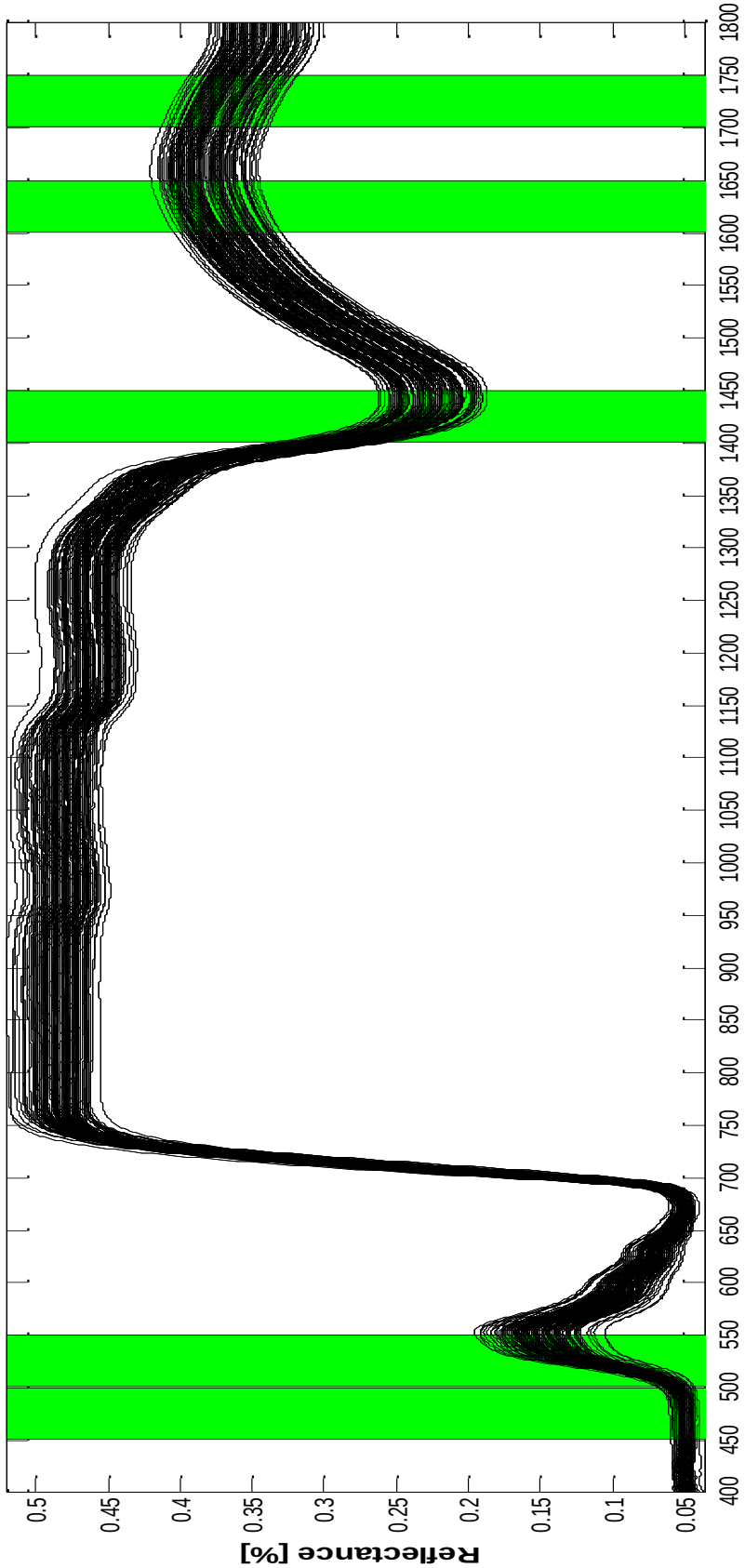


Figure 26 The optimal combination of the spectral intervals selected by synergy interval-PLS to be used for developing PLS-DA and ANNs models in apple glasshouse trial. The selected intervals (green ranges) are 450–500, 500–550, 1400–1450, 1600–1650, and 1750–1800 nm.

2.3.3.1. Apple

2.3.3.1.1. Glasshouse trial

The results from classification models will be shown and illustrated in great details only for the calculated PLS-DA and ANNs models in apple glasshouse trial. However, tables of classification performance parameters, which summarize the classification results of the calculated models, will be used to discuss and evaluate the performance of the PLS-DA and ANNs models in all other trials.

Starting with PLS-DA model from the RWs, PLS-DA components were optimized before the model calculation, in which the optimal number of LVs and the best number of cross validation groups were selected. Two validation runs with two cross validation groups (10 and five, respectively) were separately carried out by using the classification toolbox for MATLAB. To illustrate the selection procedure, the results of the two cross validation runs in apple glasshouse trial at the RWs level is shown in Figure 27. In the two validation runs, similar results of ER, the most used parameter to assess classification model quality, and NA ratios were obtained. This similarity means that the cross validation results were reliable, and either of the two runs can be used to calculate valid classification models. Theoretically, 11 is the optimal number of LVs as it was associated with both minimum ER, and minimum number of NA samples (Figure 27). Nevertheless, the differences between the calculated values with 11 and two LVs were small on the ER (0% with 11 LVs, and 0.11% with two LVs, respectively) and the NA samples (0% with 11 LVs, and 0.1% with two LVs, respectively). Thus, two LVs were selected because the use of many latent variables to calculate complex classification model do not improve the performance of the model on one hand, and simple classification models are preferable for their stability and simple interpretation on the other hand (Ballabio and Consonni, 2013). The optimal number of LVs and cross validation groups in the rest of apple and pear trials were selected following the same way, the resulted figures are shown in Annex II.

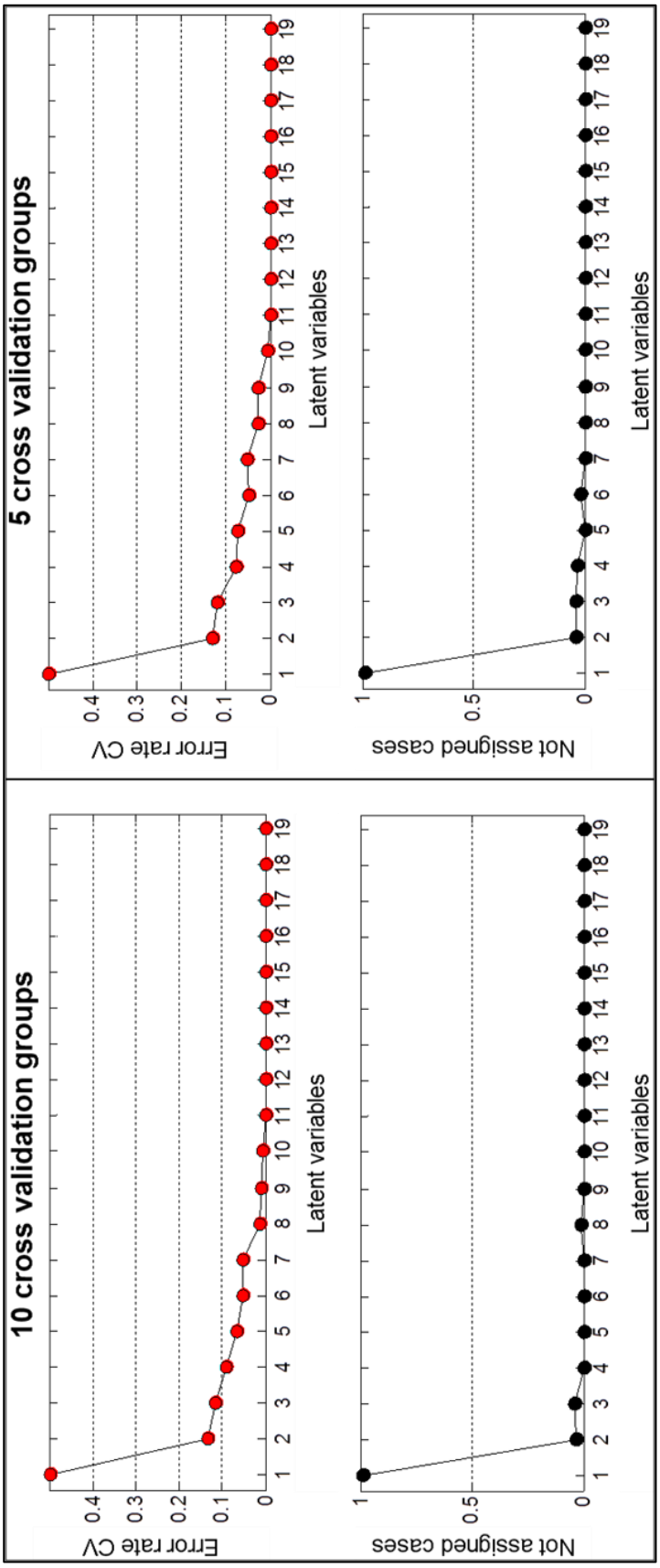


Figure 27 The error rates (red points), and the ratios of Not Assigned samples (black points) in the PLS-DA optimization process of apple glasshouse trial. The venetian blinds procedure, integrated in the classification toolbox for MATLAB, was applied to 10 and five cross validation groups (left and right parts of the figure, respectively) as a function of the number of latent variables.

Hence, two LVs and five cross validation groups were selected in apple glasshouse trial, and then PLS-DA model was calculated and calibrated from the training dataset (fitting subset) by using the classification toolbox for MATLAB. The developed classification model works as follows: (I) it uses several variables from every sample in the training dataset; next, (II) computes a specific value for each class, the healthy and the infected, called the calculated response for that class; then, (III) based on the sample calculated response, the model assigns it to one of the defined classes; eventually, (IV) the developed model is used for internal cross validation through classifying the evaluation dataset (cross validation subset). Thus, the calculated model produced two calculated responses for each analyzed sample, one for the healthy class and the other for the infected class. Figure 28 displays only the calculated responses of the infected class in all training samples of apple glasshouse trial. The dotted redline in the figure indicates to the value of the infected class threshold (0.61). The calculated model assigned all samples with calculated response higher than 0.61 as infected samples. This model misclassified three samples, as shown in Figure 28, in which two infected samples were excluded from the infected class (false negative) and one healthy sample was classified incorrectly infected (false positive).

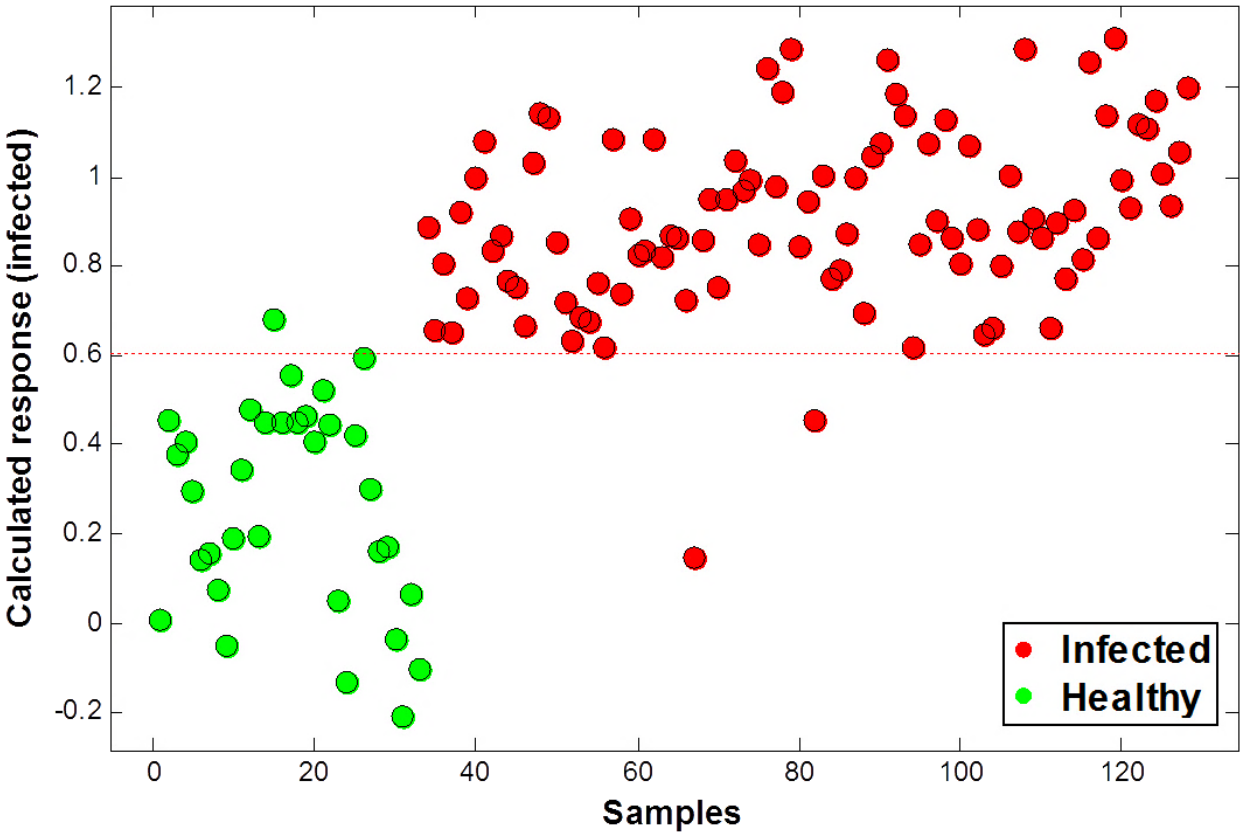


Figure 28 The calculated responses for the infected class with the healthy (green points) and the infected samples (red points) of the training dataset in apple glasshouse trial. The dotted redline indicates the infected class threshold (0.61).

Thereafter, the produced Receiver Operating Characteristic (ROC) curve for each class in the trial (healthy and fire blight-infected) was used to evaluate the capability of the classification model. Generally, perfect classification models produce ROC curves cover the plot space, which represent the maximum sensitivity and specificity of the model, whereas random or bad classification models yield points along the diagonal line of the ROC plot. In this glasshouse trial, the ROC curves resulted by this calculated model were approximately perfect in healthy and infected classes (Figure 29). The line plot in the bottom of the same figure represents the class threshold, which is a point where the number of false positives and false negatives is minimized. The class threshold value corresponds to the point where the specificity line crosses the sensitivity line. In this trial, the class threshold for the infected class was 0.61, as mentioned before, which means that the classification model calculates a value for each trial case, and if this value is higher than 0.61 it will be classified as infected. The ROC curves of the infected and the healthy classes for the rest of apple and pear trials are shown in Annex III.

Before evaluating the model performance and accuracy, it is useful to understand the studied data patterns and the model separability by plotting the scores of the selected latent variables. Figure 30 shows the score plot of the first two LVs which were selected to calculate the PLS-DA classification model in apple glasshouse trial. The first latent variable explained 99.99% of the total variance, whereas the second one explained 0.01% of the variance. The two selected LVs explained 100% of the whole variance of all analyzed variables in this trial (Figure 30). This percentage means that this PLS-DA model was built using the relevant information for class separation, and the experimental noises were successfully avoided by ignoring the non-significant latent variables. In reality, the model quality is not evaluated by the percentage of the explained variance by the LVs, and It is still possible to obtain valid and sufficient classification models even if the explained variance is low, particularly when dealing with complex classification datasets of thousands of variables (Ballabio and Todeschini, 2008).

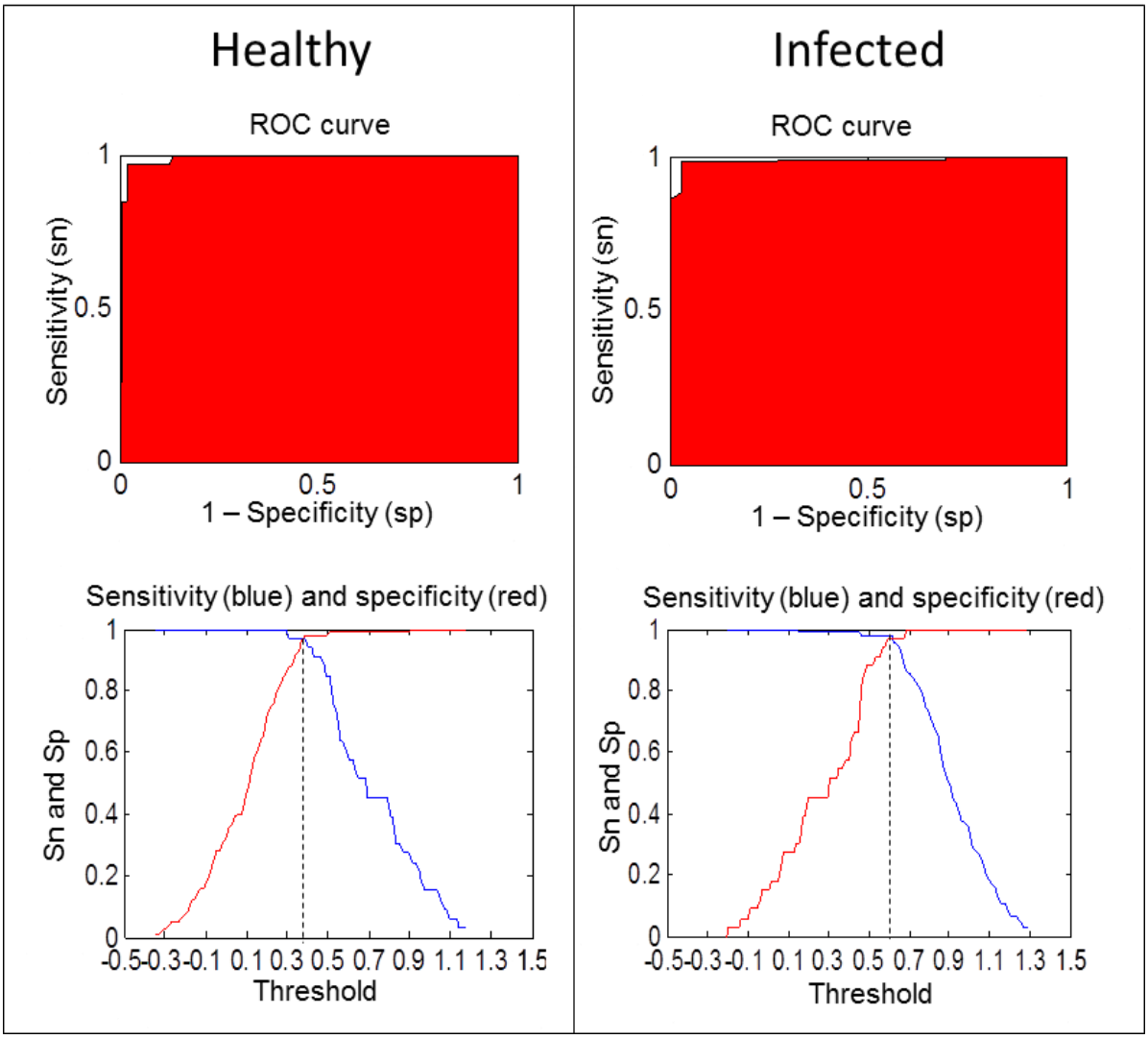


Figure 29 The Receiver Operating Characteristic (ROC) curves (upper), and line plots of sensitivity and specificity values (lower), blue and red respectively, as a function of classification threshold changing in the PLS-DA model of apple glasshouse trial.

Many healthy and fire blight-infected samples overlapped to some extent as shown in Figure 30, though the separation between samples was still clear. Infected samples had proportionally greater scores on the first LV than healthy ones. However, all samples had positive scores on the first LV. On the other hand, healthy and infected samples were distributed on the second LV in the range between -0.15 and +0.15, but the majority of the healthy samples had positive scores.

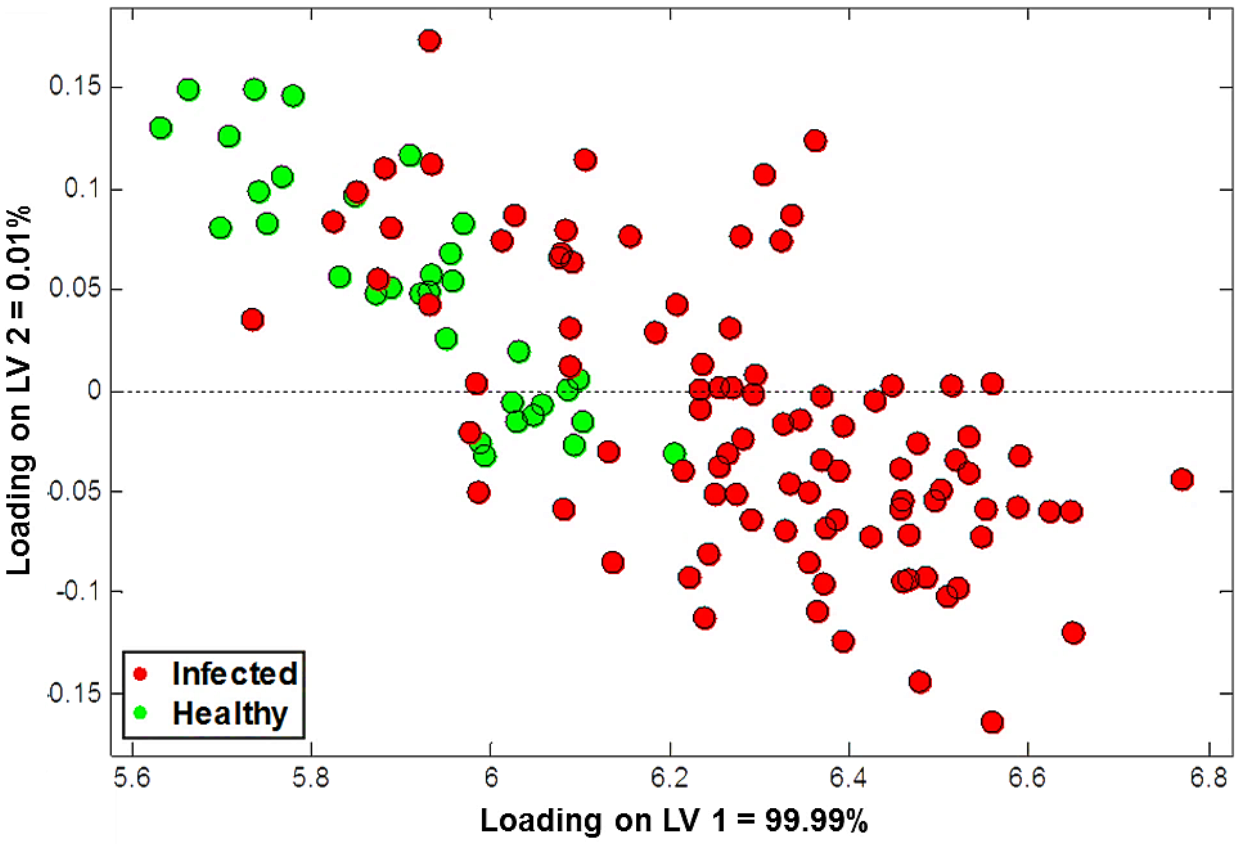


Figure 30 The scores of the healthy and the infected plants on the first two latent variables, which explained 100% of the total variance, in the PLS-DA model of apple glasshouse trial.

Detailed evaluation of the calculated model can be made by using the tables of the performance classification parameters. These tables contain the models' NER, specificity and sensitivity of each class (healthy and infected), and the ratio of NA samples. The calculated PLS-DA model from the training subset had an overall NER of 0.97 and the NA samples was 0 (Table 4), which means that 97% of the studied samples were correctly classified in their classes. Specificity and sensitivity values of the healthy class were reversed in the infected class, and that is reasonable in binary classification because only two classes are defined. In the infected class, specificity and sensitivity were equal to 0.93 and 1 respectively, while in the healthy one they were equal to 1 and 0.93, respectively. The developed PLS-DA classification model therefore recognized 100% of the infected samples, and 93% of the healthy ones. The differences in the obtained results from the training (fitting) and the evaluation subsets (cross-validation) were small, and this similarity confirms that overfitting did not occur and also assured the robustness and the reliability of the developed model. The tables of classification performance parameters are the real summery of the classification performance, so only these tables will be reported as results in all next calculated PLS-DA and ANNs models for all apple and pear trials.

Table 4 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by RWs PLS-DA model in training and evaluation subsets (with five groups split in venetian blinds) of apple glasshouse trial.

PLS-DA*	Subsets	NER	Healthy		Infected		NA
			Sp	Sn	Sp	Sn	
Glasshouse trial	Training	0.97	1	0.93	0.93	1	0
	Evaluation	0.92	0.93	0.9	0.9	0.93	0

*PLS-DA: the selected number of LVs is 2.

Concerning the ANNs model from RWs data in this apple glasshouse trial, the Kohonen and CP-ANN toolbox for MATLAB was used to optimize the network architecture of the ANNs classification model. The result of the optimization step is shown in the bubble plot (Figure 31). Each bubble in the figure represents different network architecture, which means different number of neurons and training epochs. The network that had 10X10 neurons and 75 epochs (the red bubble in Figure 31) was selected to calculate this ANNs model because it had both high relative frequency and optimization criteria values (predictive performance). Using the same optimization procedure, all ANNs classification models in this study were optimized and their plots are shown in Annex IV.

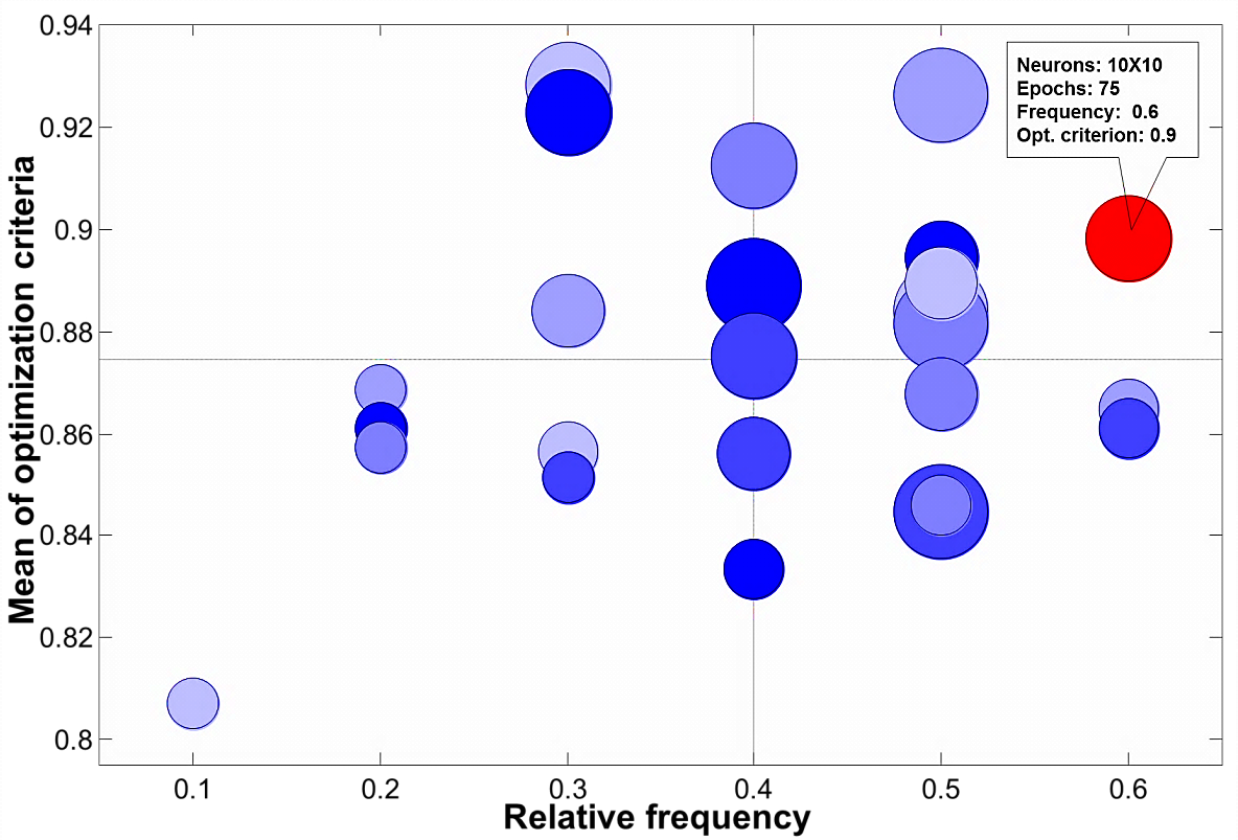


Figure 31 A bubble plot obtained by the Kohonen and CP-ANN toolbox for MATLAB (version 3.6), shows the optimized ANNs networks for RWs in the glasshouse trial of apple. The bubble in the upper right part of the plot, the red bubble of 10X10 neurons and 75 epochs was the best as it had high relative frequency and predictive performance.

Consequently, the selected parameters were used to develop an ANNs classification model using the training subset, and then the developed model was validated by the evaluation subset. The results obtained by the developed classification model, for both the training and the evaluation subsets, are summarized in Table 5. Giving comparable results in the training and the evaluation subsets with a maximum difference of 0.09 in the corresponding parameters, the developed ANNs model was highly reliable and no overfitting was occurred. However, the performance of the ANNs model was slightly worse than the PLS-DA model that was developed for the same glasshouse trial (Table 4).

Table 5 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by RWs ANNs model in training and evaluation subsets (with five groups split in venetian blinds) of apple glasshouse trial.

ANNs*	Subsets	NER	Healthy		Infected		NA
			Sp	Sn	Sp	Sn	
Glasshouse trial	Training	0.8	0.96	1	1	0.96	0
	Evaluation	0.91	0.92	0.91	0.91	0.92	0

*ANNs: the optimal network architecture is 10×10 neurons and 75 epochs.

On the SVIs level, the classification models were developed using the standardized values of the 30 SVIs calculated in this study. Using the same procedure, which was used for calculating the RWs classification models, SVIs classification models (PLS-DA and ANNs) were optimized and calculated in this apple glasshouse trial. Eight LVs were selected and used to calculate the PLS-DA model, and the optimal network architecture contained 6X6 neurons and 100 epochs. The results of the classification performance obtained from the training and the evaluation subsets are reported in Table 6. Although both models were efficient and robust due to their comparable performance in the training and the evaluation subsets, the calculated PLS-DA model outperformed the ANNs model in this glasshouse trial. The PLS-DA model correctly classified 98% of the studied samples in the training subset, whereas the ANNs model correctly identified 88% of the same samples. Moreover, the high sensitivity of the PLS-DA model for the healthy and the infected plants (96%-98%) confirmed the high accuracy of this model comparing with notably low sensitivity (70%-82%) in the ANNs model. However, the PLS-DA model was unable to assign 7% of the training samples and 12% of the evaluation samples.

The performance of the PLS-DA and the ANNs models in both RWs and SVIs data was similar to some extent (Tables 4, 5 and 6), though 30 variables were used in the PLS-DA classification model of the SVIs data. Thus, excellent classification results can be obtained using only 30 SVIs instead of using a large number of spectral reflectance ranges.

Table 6 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA and ANNs classification models for SVIs variables in apple glasshouse trial.

		Subsets	PLS-DA*						ANNs*					
			NER	Healthy		Infected		NA	NER	Healthy		Infected		NA
				Sp	Sn	Sp	Sn			Sp	Sn	Sp	Sn	
Apple glasshouse trial	SVIs	Training	0.98	0.98	0.98	0.98	0.98	0.07	0.88	0.95	0.82	0.82	0.95	0
		Evaluation	0.97	0.98	0.96	0.96	0.98	0.12	0.8	0.89	0.7	0.7	0.89	0

*PLS-DA: the selected number of LVs is 8.

*ANNs: the optimum network architecture is 6X6 neurons and 100 epochs.

It is also useful to study the importance, or the behavior, of the studied variables (the calculated SVIs) in the developed classification models for predicting fire blight infection in apple. Unlike PLS-DA classification algorithm, which were capable of giving a comprehensive insight into the relationships between variables and samples, ANNs algorithm plots the weights of only one variable at a time for all the neurons. The scores and loadings of the first two LVs for the calculated PLS-DA model are displayed in Figure 32. The first LV contained the largest part of information because it explained 48.09% of variance, while the second LV explained just 16.49%. However, there is no relation between the explained variance and the LV capability of discriminating classes, as discussed previously. Looking at the score plot (Figure 32 A), healthy and infected samples overlapped in a certain degree. Fire blight-infected samples had greater scores than healthy samples on the two LVs. Moreover, on the two LVs the majority of the infected samples had positive scores, whereas the majority of the healthy samples had negative scores. This particular pattern of the sample distribution can be better interpreted looking at the corresponding loading plot (Figure 32 B).

The negative loading of some SVIs, on either of the two LVs, is generally not related to the real values of these variables in the original dataset because all SVIs were standardized. However, the negative loading of Greenness Index (GI), Modified Chlorophyll (a and b) Absorption in Reflectance Index (MCARI), Plant Water Index (PWI), and Red Edge Vegetation Stress Index (RVSI) on both LVs can be explained by their strong effect on the classified samples as healthy, and the weak effect on the assigned samples as infected. The different loading also means that GI, MCARI, PWI, and RVSI contributed effectively in this developed PLS-DA model to discriminate the healthy apple plants. Despite their greater values in the healthy plant, every SVI is correlated with different plant aspects. GI and MCARI are correlated with the chlorophyll content (Smith *et al.*, 1995; Daughtry *et al.*, 2000), PWI is obviously related to water content (Penuelas *et al.*, 1993), and RVSI is a vegetation stress indicator (Merton, 1993).

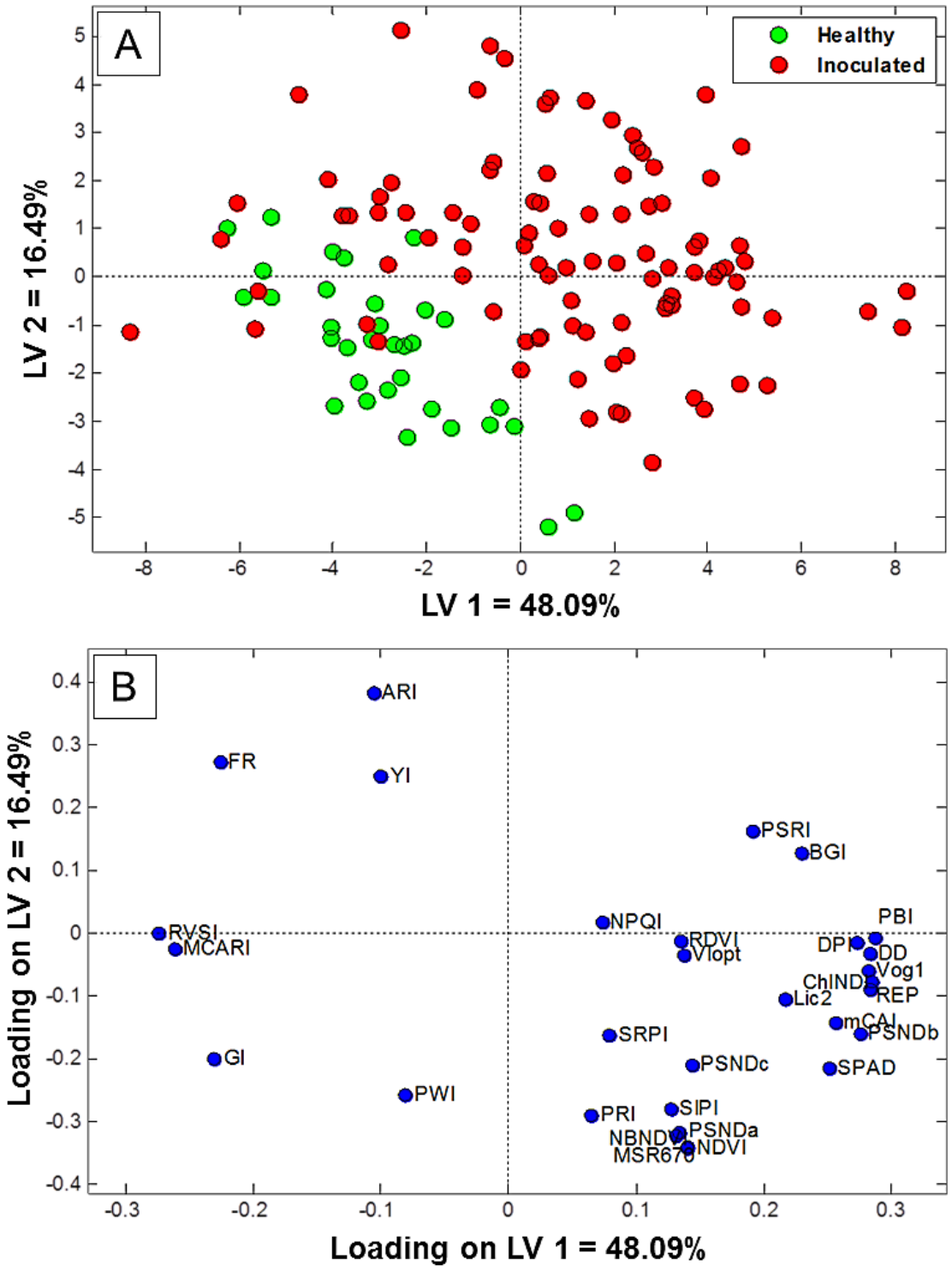


Figure 32 The scores of the healthy and the infected plants (A) and the loading plots of the tested SVIs (B) of the first two latent variables, which explained 64.58% of the total variance, in the PLS-DA model of apple glasshouse trial.

Likewise, Blue Green Index 2 (BGI2), Normalized Plant (NPQI), and Plant Senescence Reflectance Index (PSRI), which had positive loadings on the two LVs as shown in Figure 32 B, contributed the most in recognizing the infected samples by this developed PLS-DA model. Having greater values in the infected samples, BGI2, NPQI, and PSRI are reported to be correlated respectively with chlorophyll content C_{ab}, chlorophyll degradation, and plant senescence (Barnes *et al.*, 1992; Merzlyak *et al.*, 1999; Zarco-Tejada *et al.*, 2005). Other traditional indices related to vegetation structure and condition, such as Normalized Difference Vegetation Index (NDVI) or Simple Ratio (SR), showed low relationships with the infected samples. Additionally, some SVIs such as Fluorescence Ration (FR), Yellowness Index (YI), and Photochemical Reflectance Index (PRI), which are located in the corresponding area of the overlap between the healthy and the inoculated samples (Figure 32 B), had relatively similar effects on this PLS-DA classification model.

Despite the demonstrated success of these indices in the developed models, they cannot classify the healthy and the infected samples singlehandedly. A considerable contribution for other calculated indices was observed in these developed classification model. For example, some calculated indices which had positive loading only the first LV, such as Red Edge Position (REP) and Plant Biochemical Index (PBI), contributed partially in the characterization of the infected samples, but not as strong as the contribution of PSRI, BGI2, and NPQI.

2.3.3.1.2. Field trial

Field experiments are indispensable in our study as our ultimate goal was to recognize fire blight infections in open orchards. However, several technical obstacles and regulatory difficulties can hamper and even stop any fire blight outdoor experiments. The main encountered challenges in the current research were: (I) the strict prohibition of fire blight artificial inoculation outdoor to avoid further disease introductions or disseminations, (II) the difficulty in obtaining a prior permission to carry out any activities in any fire blight-infected orchards, (III) the lack of fast and reliable *in-situ* detection test for the disease, and (IV) the limitation of time to implement the experiments because they depend on weather conditions and visual symptoms. Nevertheless, meeting all the mentioned requirements, apple and pear field trials were successfully conducted in Emilia-Romagna.

In apple field trial, in which hyperspectral data were obtained only from apple leaves, the results of the developed PLS-DA and ANNs classification methods calculated from RWs variables are summarized the Table 7. Two and three LVs were selected respectively for RWs and SVIs to calculate their PLS-DA models, and for ANNs models the optimal network architecture was 10×10 neurons and 150 epochs for the RWs, whereas 6×6 neurons and 100 epochs for SVIs. From the performance Table 7, it was clear that the calculated PLS-DA classification models were more stable and reliable than the ANNs ones due to their comparable results (sensitivity and specificity) in the used training and evaluation subsets. So, the disparate results from the training and the evaluation subsets by ANNs models evidently reduced their stability and reliability. Although the performance of the PLS-DA models was better than that of the ANNs models on RWs and SVIs levels, ANNs model performance was still efficient with more than 80% correct classification for all samples.

On the level of the RWs, the calculated PLS-DA and ANNs models had similar NER (100%) in the training subset, but in the evaluation subset the ANNs model classified correctly only 88% of the samples compared with 100% in PLS-DA model. Moreover, sensitivity and specificity of healthy and symptomatic classes were perfect in the PLS-DA model (100%) and lower performance parameters (95% sensitivity and 80% specificity for symptomatic class) were revealed by the ANNs model. Dissimilarly, on the level of SVIs, the PLS-DA model correctly classified 97% of the training and the evaluation samples; whereas the ANNs model resulted in larger NER (100%) from the training samples and lower NER (82%) from the evaluation samples. However, the inconsistency in the performance in ANNs model was a big disadvantage and means that this models were unstable.

Table 7 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA and ANNs classification models for RWs and SVIs variables in apple field trial.

			PLS-DA*						ANNs*					
		subsets	NER	Healthy		Symptomatic		NA	NER	Healthy		Symptomatic		NA
				Sp	Sn	Sp	Sn			Sp	Sn	Sp	Sn	
Apple field trial	RWs	Training	1	1	1	1	1	0	1	1	1	1	1	0
		Evaluation	1	1	1	1	1	0	0.88	0.95	0.8	0.8	0.95	0
	SVIs	Training	0.97	0.95	1	1	0.95	0.03	1	1	1	1	1	0
		Evaluation	0.97	0.94	1	1	0.94	0.11	0.82	0.85	0.8	0.8	0.85	0

*PLS-DA: the optimum number of selected LVs for RWs: 3, whereas for SVIs: 2.

*ANNs: the optimum network architecture is 10×10 neurons and 150 epochs for RWs, and 6×6 neurons and 100 epochs for SVIs.

2.3.3.2. Pear

2.3.3.2.1. Glasshouse trials

Two pear glasshouse trials for 1-year-old and 3-year-old seedlings were carried out in MAIB quarantine glasshouse. The classification performance parameters obtained by PLS-DA and ANNs models, from RWs and SVIs data in the first and the second pear glasshouse trials, are summarized in Table 8. In the first glasshouse trial, three and five latent variables were selected respectively for RWs and SVIs to calculate their PLS-DA models, and for ANNs models the optimal network architecture was 8×8 neurons and 75 epochs for the RWs, whereas 14×14 neurons and 300 epochs for SVIs. As stated before, the comparable results of the performance parameters (NER, Sp, Sn, and NA) of the training and the evaluation samples provided a clear idea about the stability and the reliability of the developed classification model. The similarity in these parameters results in the calculated PLS-DA models on one hand, and the disparity of these results in the calculated ANNs models on the other hand proposed that PLS-DA models were more stable and reliable. Moreover, the calculated PLS-DA models from the SVIs (NER = 98%) outperformed the one calculated from the RWs (NER = 95%) in this glasshouse trial as shown in Table 8.

In the second glasshouse trial, seven and six latent variables were selected respectively for RWs and SVIs to calculate their PLS-DA models, and for ANNs models the optimal network architecture was 4×4 neurons and 150 epochs for the RWs, whereas 4×4 neurons and 300 epochs for SVIs. Also in this trial the calculated PLS-DA models (for both RWs and SVIs) performed better than the ANNs models due to their greater values of NER, specificity, and sensitivity (Table 8). Moreover, these PLS-DA models were more stable due to their results similarity for both the training and the evaluation subsets. The greater NER in the PLS-DA model from the SVIs (87%), compared with the PLS-DA model from RWs (84%), made the most accurate classification model in this trial. However, the calculated PLS-DA model from SVIs in the first pear glasshouse trial was even more accurate than the corresponding model in the second trial.

Table 8 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA, and ANNs classification models for RWs and SVIs variables in pear glasshouse trials.

	Pear trials	Subsets	PLS-DA*						ANNs*					
			NER	Healthy		Infected		NA	NER	Healthy		Infected		NA
				Sp	Sn	Sp	Sn			Sp	Sn	Sp	Sn	
RWs	1st Glasshouse	Training	0.95	1	0.9	0.9	1	0	0.98	0.97	1	1	0.97	0
		Evaluation	0.95	1	0.9	0.9	1	0	0.89	0.88	0.9	0.9	0.88	0
	2nd Glasshouse	Training	0.84	0.83	0.85	0.85	0.83	0.02	0.78	0.88	0.69	0.69	0.88	0
		Evaluation	0.84	0.85	0.84	0.84	0.85	0.03	0.69	0.79	0.6	0.6	0.79	0
SVIs	1st Glasshouse	Training	0.98	0.96	1	1	0.96	0.17	1	1	1	1	1	0
		Evaluation	0.98	0.97	1	1	0.97	0.15	0.89	0.88	0.9	0.9	0.88	0
	2nd Glasshouse	Training	0.87	0.89	0.86	0.86	0.89	0.09	0.76	0.83	0.69	0.69	0.83	0
		Evaluation	0.87	0.88	0.87	0.87	0.88	0.15	0.7	0.81	0.59	0.59	0.81	0

*PLS-DA: the optimum number of selected LVs for RWs in 1st glasshouse trial: 3, and in 2nd glasshouse trial: 7, for SVIs in 1st glasshouse trial: 5, and in 2nd glasshouse trial: 6.

*ANNs: the optimum network architecture for RWs in 1st glasshouse trial 8×8 neurons and 75 epochs, and in 2nd glasshouse trial 4×4 neurons and 150 epochs, while for SVIs in 1st glasshouse trial 14×14 neurons and 300 epochs, and in 2nd glasshouse trial 4×4 neurons and 300 epochs.

2.3.3.2.2. Field trials

The pear open field trials were carried out on two levels, canopy and leaf. Table 9 summarizes the results of the developed PLS-DA and ANNs classification methods which were calculated from RWs and SVIs variables.

Firstly, on the leaf field trial, six and four LVs were selected respectively for RWs and SVIs to calculate their PLS-DA models, and for ANNs models the optimal network architecture was 12×12 neurons and 350 epochs for the RWs, whereas 4×4 neurons and 150 epochs for SVIs. Like the previous glasshouse trials, the performance of the PLS-DA models exceeded the ANNs ones because of their greater values of NER, specificity, and sensitivity. Namely the calculated PLS-DA model from SVIs of the leaves was the most accurate model for classifying the healthy and the fire blight-symptomatic pear trees in field, though some samples were not classified at all by this model. On the other hand, the significant differences in the performance of the ANNs model between the training and the evaluation subsets lessened its classification accuracy (Table 9). These differences reduced also the stability and the reliability of the calculated ANNs classification models.

Secondly, on the canopy field trial, seven and six LVs were selected respectively for RWs and SVIs to calculate their PLS-DA models, and for ANNs models the optimal network architecture was 4×4 neurons and 250 epochs for the RWs, whereas 6×6 neurons and 350 epochs for SVIs. The overall performance of the PLS-DA models, calculated from both RWs and SVIs, was notably higher than the ANNs models as reported in Table 9. Unlike the PLS-DA model calculated from SVIs, the PLS-DA model calculated from RWs provided incomparable results for the training and the evaluation samples. Thus, the PLS-DA classification model calculated from SVIs outperformed both the ANNs models and the PLS-DA model calculated from RWs.

The loading plots of the successful indices in leaf and canopy trials showed that the successful indices developed at the leaf level did not perform similarly at the canopy level due to atmosphere conditions and soil background. This result agrees with Zarco-Tejada *et al*., 2001 and 2004 who developed assessment models for vineyards' condition.

Table 9 Classification performance parameters [Non-Error Rate (NER), class Specificity (Sp) and Sensitivity (Sn), and ratio of Not Assigned samples (NA)] obtained by PLS-DA, and ANNs classification models for RWs and SVIs variables in pear field trials

	Pear trials	Subsets	PLS-DA*						ANNs*					
			NER	Healthy		Symptomatic		NA	NER	Healthy		Symptomatic		NA
				Sp	Sn	Sp	Sn			Sp	Sn	Sp	Sn	
RWs	Field (leaf)	Training	1	1	1	1	1	0	1	1	1	1	1	0
		Evaluation	0.9	0.92	0.89	0.89	0.92	0	0.94	0.5	0.78	0.78	0.5	0
	Field (canopy)	Training	1	1	1	1	1	0.06	0.78	0.78	0.78	0.78	0.78	0
		Evaluation	0.65	0.75	0.56	0.56	0.75	0.06	0.56	0.44	0.67	0.67	0.44	0
SVIs	Field (leaf)	Training	1	1	1	1	1	0.05	0.88	0.75	1	1	0.75	0
		Evaluation	1	1	1	1	1	0.43	0.71	0.75	0.67	0.67	0.75	0
	Field (canopy)	Training	1	1	1	1	1	0.17	0.89	0.78	1	1	0.78	0
		Evaluation	1	1	1	1	1	0.67	0.56	0.56	0.56	0.56	0.56	0

*PLS-DA: the optimum number of selected LVs for RWs in field (leaf) trial: 6, and in field (canopy) trial: 7, for SVIs in f field (leaf) trial: 4, and in field (canopy) trial: 6.

*ANNs: the optimum network architecture for RWs in field (leaf) trial 12×12 neurons and 350 epochs, and in field (canopy) trial 4×4 neurons and 250 epochs, while for SVIs in field (leaf) trial 4×4 neurons and 150 epochs, and in field (canopy) trial 6×6 neurons and 350 epochs.

2.4. Conclusions

Our study has sought to uncover the potential of hyperspectral reflectance data to discriminate between healthy and fire blight-infected apple and pear. The increasing importance of fire blight, due to its severe damage, lack of effective control, asymptomatic infections, and continuously fast dissemination, make it imperative for researchers to develop modern early detection methods on large-scale to help fighting preventively against the disease. However, the only study that addressed the use of Near Infrared spectral data for fire blight detection was failed to discriminate between healthy and fire blight-inoculated pear (cv. Abbé Fétél) under controlled conditions (Spinelli *et al.*, 2006). The current study was set out to explore (I) the hyperspectral difference between fire blight-infected (symptomatic and asymptomatic) and non-infected apple and pear in glasshouse and open field trials, and (II) to develop classification models capable of predicting fire blight infection.

Showing distinctive spectral signatures, hyperspectral analysis demonstrated high capacity of separating the spectral signatures of fire blight-infected and non-infected apple and pear not only under controlled conditions, but also in open fields. The most important advantage was the indisputable capability of hyperspectral analysis to recognize the asymptomatic fire blight-inoculated plants which made this analysis extremely useful and promising for early detection procedures.

Good predictability could be achieved by using hyperspectral data and non-parametric classification methods, such as Partial Least Square-Discriminant Analysis (PLS-DA) and Artificial Neural Networks (ANNs), to classify apple and pear plants infected and non-infected with *E. amylovora*. However, the two multivariate analysis-based methods, PLS-DA and ANNs, resulted in different classification performance in each glasshouse and field trial. PLS-DA classification models were more efficient and accurate than ANNs models in the hyperspectral classification of healthy and fire blight-infected apple and pear for several reasons. Firstly, the optimization process of the best network architecture in ANNs models is arbitrary and time consuming due to its algorithm inflexibility and extremely long procedure. Dissimilarly, the optimization procedure of PLS-DA classification models is fast, smooth, and more efficient. Secondly, although the developing of ANNs models was preceded by an optimization step, overfitting is still a serious issue when training these type of supervised classification models. Finally, elaborating a large number of multicollinear variables, such as the hyperspectral reflectance data in our study, leads to totally invalid ANNs models. Even the accidental noise in the collected spectral signature was fitted in the calculated ANNs classification model, and ended up with misleading classification models. Unlike ANNs, the overfitting issue was relatively less effective in PLS-DA classification models, and it was reduced noticeably by evaluation procedure.

The non-parametric classification models, PLS-DA and ANNs, calculated from Spectral Vegetation Indices (SVIs) for fire blight detection in apple and pear are fairly better than the models calculated from Reflectance Wavebands (RWs). The reasons behind that are: (I) the smaller number of needed hyperspectral data for the indices models calculation comparing to the huge number in the RWs ones, (II) the ratio of a linear combination of reflectance reduces the multicollinearity, and (III the higher classification accuracy and performance of the models calculated from the indices comparing to the RWs models. However, it must be noted that the vegetation indices need to be tested on various sensors and different varieties of apple and pear in order to be used perfectly in precision farming, also atmospheric constrains, sensor view, the sun angles, soil background, and canopy architecture must be accounted for if the vegetation indices are calculated from canopy hyperspectral data.

Based on the excellent classification results and the robustness shown in this study, the calculated models can be further developed to be integrated in many fire blight management applications and activities. Phytosanitary inspectors can use these developed classification models to monitor frequently fire blight faster and easier than the conventional laborious procedures. Moreover, the developed models can be improved to monitor the disease on a large scale through replacing the spectroradiometer by a special camera mounted on an Unmanned Aerial Vehicles (drones), which collect the necessary hyperspectral data from many trees at once. Furthermore, the performance of fire blight risk assessment models, such as *CougarBlight* and MARYBLYTE, can be improved by using the classification models developed in the current study.

Chapter 3

Molecular typing of some Mediterranean isolates of *Erwinia amylovora* the causal agent of fire blight

Abstract

Erwinia amylovora, the causal agent of fire blight, is the most devastating pathogen that destructs economically important rosaceous plants like apple and pear. Molecular epidemiology and evolutionary studies of *E. amylovora* are challenging due to pathogen low genetic diversity and the lack of simple and high resolution genotyping techniques. Several genotyping methods can distinguish between *E. amylovora* isolates from *Rubus* and *Spiraeoideae* host plants, though they fail to distinguish isolates within *Spiraeoideae* group. Recently, a Multiple-Locus VNTR Analysis (MLVA) based on six Variable Number of Tandem Repeats (VNTR) loci was successfully used to genotype different *E. amylovora* strains and identify previously unresolved populations.

In this study, MLVA analysis based on a developed set of nine VNTR loci, six of them were reported earlier and the other three were revealed in the current research, was used to evaluate the DNA of 46 *E. amylovora* strains collected from different host plants in 16 different countries mainly Mediterranean. The developed MLVA recognized 38 haplotypes clustered in seven Clonal Complexes, or populations. The study demonstrated also the relatively high genetic diversity among the Mediterranean strains particularly from the Balkan Peninsula and the Eastern Mediterranean countries. Furthermore, our results showed significant diversity among the studied Italian strains which confirmed the continuous multiple introductions of the disease. The developed MLVA can be used effectively as a fast, cheap, and simple tool to track *E. amylovora* infection sources, to have a detailed insight into the geographic diversity, and to understand the dynamic evolution of the pathogen.

3.1. Introduction

Fire blight, which is caused by the bacterium *E. amylovora*, is the most serious disease on apple and pear in Italy. More than half a million infected fruit trees were eradicated in a hopeless attempt to stop this capricious disease. Despite its broad host plant range within the plants of *Rosaceae* family, *E. amylovora* strains are mainly divided into *Spiraeoideae*- and *Rubus*-infecting groups according to their host specificity (Braun and Hildebrand, 2005). The first group, *Spiraeoideae*-infecting, infects member plants of the subfamily *Spiraeoideae* such as apple and pear species, hawthorn, and cotoneaster. The second group, which is referred to as *Rubus*-infecting strains, infects host plants in the genus *Rubus* such as raspberry and blackberry. *Rubus*-infecting strains are non-pathogenic to apple, but some variations are capable of infecting immature pear fruit, with some strains being weakly virulent and others unable to cause any symptoms (Powney *et al.*, 2011). The main phenotypic differences between the *Spiraeoideae*- and *Rubus*-infecting strains include variations in exopolysaccharide composition, carbon utilization, and profiles of secreted protein (Maes *et al.*, 2001; Braun and Hildebrand, 2005).

3.1.1. Homogeneity and molecular typing in *Erwinia amylovora*

Genetic diversity of *E. amylovora* strains is extremely low. Comparison of complete genome sequences of two *Spiraeoideae*-infecting strains of *E. amylovora*, from apple and pear, revealed 99.99% sequence identity (Sebaihia *et al.*, 2010; Smits *et al.*, 2010). Another study confirmed that the genomic identity of *E. amylovora* is ranging from ~ 99.5% within the *Rubus*-infecting strains to 99.98% within the *Spiraeoideae*-infecting strains (Mann *et al.*, 2013). This low genetic diversity makes the characterization of different *E. amylovora* isolates by the routine typing methods challenging. Furthermore, searching for new discriminatory markers, which can generally subdivide pathogens, has become indispensable for this pathogen because these markers are useful to: I) understand whether outbreaks have a common source, II) track particular strains for their epidemic and pandemic spread, and III) reconstruct strains evolutionary history (Achtman, 2008).

Several molecular genotyping methods, which have been used for the characterization of *E. amylovora*, had confirmed the genetic and the phenotypic homogeneity of its strains (Donat *et al.*, 2007; Smits *et al.*, 2010). These genotyping techniques include ribotyping, rep-PCR tools: BOX, ERIC, and REP (McManus and Jones, 1995), fAFLP and AFLP (Rico *et al.*, 2004; Yaich *et al.*, 2011), RAPD (Momol *et al.*, 1997), Pulsed-Field Gel Electrophoresis (PFGE) after *Xba*I restriction (Jock *et al.*, 2002), and CRISPR analysis (Rezzonico *et al.*, 2011). Although a few of these typing methods can discriminate between *Spiraeoideae*- and *Rubus*-infecting isolates of *E. amylovora*, they usually fail to distinguish isolates within the *Spiraeoideae* host plants (Kim and Geider, 1999; Jock *et al.*, 2002; Gehring and Geider, 2012).

3.1.2. Multiple-Locus Variable Number of Tandem Repeats Analysis (MLVA)

Recently, a new molecular assay for *E. amylovora* typing, based on the VNTR, was reported to have the greatest discriminatory power up to now. MLVA is a real revolution in the genotyping techniques. Used to study population structure, geographical distribution, and outbreaks of several human and plant pathogens, MLVA is widely adopted by microbiologist as a simple and capable tool for evolutionary studies.

VNTR loci were firstly identified in humans for parental identification, forensics, and DNA fingerprinting. These polymorphic markers have been identified also in bacteria and have been efficiently used as a molecular typing method especially in genetically monomorphic species (van Belkum, 1999). Depending on the size of their repeating sequence, VNTRs are classified in two main categories: microsatellites and minisatellites. Microsatellites, which are also known as Simple Sequence Repeats (SSRs), are short DNA motifs (1–6 nucleotides) that are tandemly repeated in variable frequency. Minisatellites, on the other hand, are repeat units in the range of 6–100 bp (Gur-Arie *et al.*, 2000; Le Fleche *et al.*, 2001). In spite of the fact that the border between microsatellites and minisatellites is undefined, repeat units of about 6–9 nucleotides are often used to differentiate between them. Slipped-Strand Mispairing (SSM) is the principal mechanisms of VNTR variability (Torres-Cruz and van der Woude, 2003). Tandem Repeat (TR) loci are unstable and vary among strains with respect to the number of their repeat units and their primary structure. Moreover, they play a major role in generating and maintaining bacterial diversity. The hyper mutable TR loci help in bacterial adaptation, though they maintain a stable genome with a low overall genomic mutation rate (Legendre *et al.*, 2007). TR loci can therefore be used to study strains relatedness and serve as targets for epidemiological typing where the variable number of repeat units at each locus is characterized by using locus-specific primers to PCR-amplify the DNA region.

Based on six repeats only, MLVA allowed the distinguish of 227 haplotypes among 833 *E. amylovora* strains of different origin and collected from different host plants (Bühlmann *et al.*, 2014). However, these VNTR markers failed to identify isolates by distance in Europe and also failed to detect significant diversity of haplotypes within outbreaks between years.

Objective

The disastrous outbreaks of fire blight, in the Mediterranean region in general and particularly in Italy, made the study of the genetic diversity indispensable in order to identify the inoculum source of the causal agent and subsequently to prevent new introductions in the future. Furthermore, genotyping of *E. amylovora* is definitely important to assign the studied strains to predefined populations which help the epidemiological studies and source tracking. The objective of the present study was therefore to use the MLVA assay to analyze the DNA of a number of *E. amylovora* strains isolated from different host plants in Italy, some Mediterranean countries, and other regions.

3.2. Materials and methods

3.2.1. Bacterial strains and DNA quantification

Crude DNA from 46 *E. amylovora* strains were kindly provided by Zurich University for Applied Science (ZHAW), University of Belgrade, Emilia-Romagna Regional Phytosanitary Services, and by other laboratories (Table 10). All the molecular analyses in this study were carried out in the molecular laboratory of the Department of Agriculture, Forestry, Nature, and Energy Science and Technology (DAFNE), Tuscia University, Viterbo, Italy. Firstly, the DNA of the studied strains were genotypically confirmed as *E. amylovora* by nested PCR that target the ubiquitous plasmid pEA29 (Llop *et al.*, 2000). The collected DNA samples were subsequently quantified by a Qubit® Fluorometer (Invitrogen, Monza, Italy), adjusted to 40 ng/µL by SDW, and stored at -20°C.

Table 10 A list of the DNA from *Erwinia amylovora* strains which are collected used in this study.

No.	Strain's ID	Geographic origin	Host plant	Year of isolation
1	**DZC4A**	Algeria	Pear	2010
2	**E1E**			
3	**C1A**			
4	**Ea668/05**	Austria	Apple	2005
5	**Ea4/82**	Egypt	Pear	1982
6	**CFBP2301**	Haute-Savoie, France	Firethorn	1981
7	**CFBP1430**	Lille, France	Hawthorn	1972
8	**EA1**	Greece	Pear	2014
9	**Ea33**	Iran	Quince	2009
10	**Ea263**	Israel	Quince	1997
11	**E204**	Italy	Pear	2002
12	**EAI208**			
13	**OMPBO1077,7/94**	Bologna, Italy	Pear	1994
14	**OMPER25,1**	Ferrara, Italy	Pear	2000
15	**OMPBO8630,1/10**			2010
16	**OMPBO347,1/06**	Parma, Italy	Hawthorn	2006
17	**OMPBO329,2/02**	Ravenna, Italy	Apple	2002
18	**OMPER22481,1A**		Pear	2013
19	**KOS36**	Kosovo	Apple	2012
20	**KOS45**			
21	**KOS43**			
22	**KOS67**			
23	**KOS61**			
24	**CFBP4781**	Lebanon	Apple	1998
25	**LebA3**			1998
26	**LEB1**			2013
27	**LEB2**			
28	**LEB3**		Pear	
29	**LEB4**		Apple	
30	**LEB6**		Pear	
31	**LEB7**		Apple	
32	**LEB8**			
33	**KFB146**	Serbia	Apple	1998
34	**KFB150**			2005
35	**KFB187**			2003
36	**KFB153**		Pear	1998
37	**KFB161**			2005
38	**KFB162**			2005
39	**UPN527**	Spain	Pear	1997
40	**ACW56400**	Fribourg, Switzerland	Pear	2007
41	**ACW63579**	St. Gallen, Switzerland	Apple	2008
42	**TK119**	Turkey	Pear	2011
43	**CFBP1232**	UK	Pear	1959
44	**UTFer2**	Utah, USA	Apple	N.A*
45	**EA153**	Oregon, USA	Apple	1989
46	**JL1185**	Washington, USA	Pear	1988

*N.A not available

3.2.2. VNTR loci mining and primer designing

The Whole-Genome Sequence (WGS) of CFBP1430, the European reference strain of *E. amylovora* (Smits *et al.*, 2010), was used in the Tandem Repeats Finder (TRF) program version 4.08 (Benson, 1999) for *In-silico* VNTR analysis. The alignment parameters were set as -2, 7, 7-, the minimum alignment score 30, and the maximum period size 20 bp. Then, the TRs of > 91% match percentage were selected to continue the analysis. The selected loci were confirmed by comparing their sequences to the genome sequence of the American reference strain ATCC49946 (Sebaihia *et al.*, 2010). Only six of the selected TRs were reported to be polymorphic VNTRs in a previous study (Bühlmann *et al.*, 2014). Up and downstream flanking regions, of 500 bp each, were used to design the primer pairs for PCR amplification through the software Primer3plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi).

3.2.3. PCR amplification and capillary electrophoresis

The selected VNTR loci were amplified by using a C1000™ Thermal Cycler (Bio-Rad, USA) in a final volume of 25 µL. Each PCR reaction mix contained 12.5 µL 2X GoTaq® master mix (Promega, Madison, USA), 9.5 µL nuclease-free water (Promega, Madison, USA), 2 µL (10 pmol mL^{-1}) corresponding primers synthesized by a commercial company (Macrogen Europe, Amsterdam, the Netherland), and 40 ng of template DNA. Based on the size of the expected amplicon, the following parameters were used for amplification: (1) amplicons below 300 bp 95°C for 10 min, 39 cycles of 95°C for 30 s, 56°C for 30 s, and 72°C for 30 s, and 72°C for 5 min, (2) amplicons between 300 bp and 600 bp 95°C for 10 min, 39 cycles of 95°C for 30 s, 56.5°C for 30 s, and 72°C for 36 s, and 72°C for 5 min, and (3) amplicons larger than 600 bp 95°C for 10 min, 39 cycles of 95°C for 30 s, 58°C for 30 s, and 72°C for 1 min, and 72°C for 5 min. The PCR products were then analyzed by using a QIAxcel multicapillary electrophoresis system (QIAGEN, Irvine, USA). A DNA High-Resolution gel cartridge was used for estimating amplicons sizes below 600 bp and a DNA screening cartridge for amplicons larger than 600 bp. Finally, the sizes of the amplicons were determined by QIAxcel ScreenGel software.

3.2.4. MLVA analysis

The numbers of the repeat units were confirmed through sequencing. The amplicons of 11 suspected polymorphic VNTR loci, which were produced different amplicons in the model strain CFBP1430 and in four other strains (KFB150, KFB153, DZC4A, and LEB8) representative to the collected list, were sequenced by a commercial company (Macrogen, Amsterdam, the Netherlands). Thereafter, MEGA6 software (Tamura *et al.*, 2013) was used align and analyze the obtained sequences, and the number of the repeat units were calculated precisely based on the sequence from each primer. Next, the numbers of repeat units were compared with the results from the capillary electrophoresis in each locus and the sequence differences in each repeat unit were carefully checked. Then, the number of all alleles and the frequency distribution of the repeats were determined for all studied strains in each locus by subtracting the right and the left flanking regions from the amplicons, which resulted from the PCR using the corresponding primer pairs, and dividing them by the size of the corresponding repeat unit in that locus.

The discriminatory power of the nine VNTR loci in the collected DNA samples was evaluated by Simpson's index of diversity (D; Simpson, 1949) which was calculated by the online tool available on the Comparing Partitions website (htttp://Darwin.phyloviz.net/Comparingpartions/index.php?link=Home). Finally, the scores were imported into the BioNumerics Version 7.5 (Applied Maths, Belgium) for the Minimum Spanning Tree (MST) analysis.

3.3. Results and discussions

3.3.1. VNTR loci selection and MLVA optimization

Seven hundred ninety nine candidate VNTR loci (790 on the chromosome and nine on the almost ubiquitous plasmid pEA29) were obtained by analyzing the sequence of the European reference strain (CFBP1430) through TRF program. The alignment analysis of the candidate loci against the available genome sequence of the American reference strain (ATCC49946) showed that less than 1% of these loci varied in size between the two strains. However, a larger number of the candidate loci were included in this study because only two *E. amylovora* genome sequences were accessible and used in the analysis, and the variation in repeat numbers may appear in different loci if other genome sequences were available and used in this study. A total of 56 candidate VNTR loci (54 on the chromosome and two on the plasmid pEA29), which had a match percentage > 91% each, were therefore selected as shown in Table 11 and their designed primer pairs are shown in Table 12. The candidate VNTR loci and their designed primer pairs were named with the prefix Ea. Six VNTR loci from the 56 candidate ones were the same as the recently reported loci in Bühlmann *et al*. (2014) paper, four on the chromosome (Ea27, Ea30, Ea34, and Ea39) and two others on the ubiquitous plasmid pEA29 (Ea55 and Ea56).

Table 11 The selected VNTR loci in the *Erwinia amylovora* CFBP1430 genome sequence by the Tandem Repeats Finder program.

Candidate VNTR	Indices in CFBP1430	Period size	No. of copies	Percent of matches	Alignment score	%A	%C	%G	%T
Ea01	135815–135886	36	2.0	100	144	25	33	30	11
Ea02	217608–218136	157	3.4	98	1015	26	26	25	21
Ea03	483539–483831	125	2.3	97	568	13	30	24	31
Ea04	1246592–1247285	157	4.4	94	1221	23	29	32	13
Ea05	1246592–1247285	313	2.2	97	1307	23	29	32	13
Ea06	1386439–1387256	117	7.0	98	1559	23	35	26	14
Ea07	1404394–1404761	108	3.4	95	664	12	39	36	11
Ea08	1762956–1763624	223	3.0	99	1320	20	26	26	27
Ea09	1481192–1481310	45	2.6	95	220	31	25	31	11
Ea10	1906833–1907077	51	4.8	97	454	41	26	23	8
Ea11	1404394–1404753	54	6.7	91	585	13	39	36	11
Ea12	1844182–1844269	39	2.3	98	169	5	39	27	27
Ea13	2165930–2166403	184	2.6	98	921	22	28	27	21
Ea14	2028610–2029545	168	5.6	95	1637	22	33	24	19
Ea15	2544581–2545230	157	4.1	98	1246	23	29	33	13
Ea16	2706983–2707298	156	2.0	100	632	26	25	25	22
Ea17	2916188–2916338	61	2.5	100	302	20	32	25	21
Ea18	3596300–3596536	125	1.9	98	456	22	23	37	16

Candidate VNTR	Indices in CFBP1430	Period size	No. of copies	Percent of matches	Alignment score	%A	%C	%G	%T
Ea19	3197128–3198127	235	4.3	94	1788	23	23	27	25
Ea20	3197128–3198127	353	2.8	95	1813	23	23	27	25
Ea21	535681–535730	25	2.0	100	100	24	24	28	24
Ea22	908185–908216	16	2.0	100	64	37	31	12	18
Ea23	1536275–1536318	18	2.4	100	88	9	27	29	34
Ea24	882531–882570	21	1.9	94	71	20	17	40	22
Ea25	771309–771342	18	1.9	94	59	11	23	29	35
Ea26	1535260–1535490	18	12.8	91	336	13	41	29	15
Ea27*	2668417–2668535	18	6.6	92	121	17	20	36	25
Ea28	3156249–3156296	23	2.1	92	78	14	35	29	20
Ea29	3747554–3747590	18	2	94	NA	NA	NA	NA	NA
Ea30*	457256–457310	6	9.2	100	110	18	0	16	65
Ea31	624545–624570	12	2.2	100	52	7	15	53	23
Ea32	628542–628571	15	2.0	100	60	20	13	53	13
Ea33	1112313–1112338	9	2.9	100	52	11	23	34	30
Ea34*	1254089–1254135	6	7.8	100	94	31	17	34	17
Ea35	2178342–2178370	15	1.9	100	58	6	31	41	20
Ea36	2233004–2233029	8	3.3	100	52	11	26	38	23
Ea37	2430651–2430676	12	2.2	100	52	34	42	23	0
Ea38	2944967–2945001	6	5.8	100	70	34	0	17	48
Ea39*	3517824–3517893	9	7.8	100	140	21	11	34	32
Ea40	3567697–3567726	11	2.7	100	60	50	10	26	13
Ea41	2033430–2033459	9	3.3	95	51	30	36	13	20
Ea42	420109–420126	6	3.0	100	36	0	33	50	16
Ea43	1013861–1013880	6	3.3	100	40	20	45	35	0
Ea44	1736270–1736284	5	3.0	100	30	0	40	20	40
Ea45	1876910–1876925	5	3.2	100	32	0	43	37	18
Ea46	2051788–2051806	6	3.2	100	38	0	31	36	31
Ea47	2326928–2326945	6	3.0	100	36	16	50	33	0
Ea48	2787772–2787791	6	3.3	100	40	0	35	50	15
Ea49	2860959–2860975	5	3.4	100	34	0	41	41	17
Ea50	3134342–3134356	5	3.0	100	30	0	60	40	0
Ea51	3169307–3169323	5	3.4	100	34	17	41	41	0
Ea52	3575763–3575777	4	3.8	100	30	26	0	0	73
Ea53	3613913–3613933	6	3.5	100	42	38	0	47	14
Ea54	3772868–3772885	6	3.0	100	36	16	50	16	16
Ea55* [Ŧ]	26041–26084	8	5.5	100	88	50	11	11	27
Ea56* [Ŧ]	14703–14728	8	3.3	94	43	11	46	11	30

*Ea reported previously as VNTRs; [Ŧ]Ea located on the plasmid pEA29 and the others are located on the chromosome.

Table 12 List of designed primer pairs for the selected VNTR loci which were revealed from the European reference strain CFBP1430 by the TRF program.

Candidate VNTRs	Primers		Candidate VNTRs	Primer	
Ea01	F	TGTGGATCGCACATTTTCAT	Ea15	F	AATCAATTGGTCGCAGGTTC
	R	ATTGACGGTAGCGTTCAGGT		R	AAAGAAGTGATAAGAAAGTGAAAGGTG
Ea02	F	CTTGGTAGCGCAACTGGTTT	Ea16	F	TCTGCACCGCTCAAAATTC
	R	GTTTCTGGCTTGGCCTGTTA		R	TGGCGTTATCAACAGAACAAA
Ea03	F	GAATGCGTCCTGCCATAAAT	Ea17	F	AACCGGCTATCCGAAATACC
	R	TAGTTTTCAGGCCGTGCTTC		R	ATAAACCGCAAGCGATCAAC
Ea04	F	CCACCATATCGGGTCGTTAG	Ea18	F	ACGGTAATATGCCGCAAAAG
	R	GCATCTCTGGTTAGCGGAAG		R	GTTATGGGGTGAGGGAACC
Ea05	F	CCACCATATCGGGTCGTTAG	Ea19	F	GGTTTGCACTGCTGGTAGC
	R	TACGCCGTTAGCTCAGCTTT		R	TGACTGTCCACGACAGAACC
Ea06	F	ATGGCTATGAAGTGCCCAAC	Ea20	F	TAATCCCGAGTGGTTTGCAC
	R	TGACGCGTCTGGTAACACAT		R	TGACTGTCCACGACAGAACC
Ea07	F	TCCTGTCGAACCCCCTGTTTA	Ea21	F	AACGGTCAGCATACGGTGTT
	R	CTGATCGGCTTCAGGTGAGT		R	ATCGATAACGGTGGGTTCAA
Ea08	F	AGGTCCAGTCGATGCTCTGT	Ea22	F	CACGGTAGTCACCAACCAGA
	R	CGCTATATAACGCCCCCAATG		R	GCTGCATCAGGTTCCTCTTC
Ea09	F	GAGAAAGCCATTGCCGATAA	Ea23	F	TGCGACGATTGTTATTACGG
	R	CTCTTTCCAGTGCGGTTTTC		R	CCGGCAGTTGAAGAGAAAAG
Ea10	F	AAGCATGTTAAGAAACATAAGAAAGC	Ea24	F	GCTTTGCAGGTAGCCAATTC
	R	TACGCTTTAGTGGCGGTTTT		R	TAGCCTGTGAAGGCAAAGGT
Ea11	F	TCCTGTCGAACCCCCTGTTTA	Ea25	F	ATACATCCGCTGACCATCCT
	R	CTGATCGGCTTCAGGTGAGT		R	GCGCAAAGTCGATCATCATA
Ea12	F	TGGGGGTTTATCAGTCTGGA	Ea26	F	TCTACAGCTTCGGCCTGTTC
	R	CGATATCTTGCGCGTCTTG		R	CAGCGTGACGGAAAGCAT
Ea13	F	GGCAACGTATCGAGGGTTC	Ea27	F	GTAGGCAGCTCACCAGCAG
	R	GGGATCACTCGCTGATGACT		R	TGCGGTAGAGAGCAGCAATA
Ea14	F	AATCTGCCCGTCATCGACTTC	Ea28	F	GTAGGGGTGCTGCGTAAATCG
	R	GAATTCGCGCCAATAAAGAA		R	CGGAAAGGGCAATTAACAGA

Table 12, supplement list of designed primer pairs for the selected VNTR loci which were revealed from the European reference strain CFBP1430 by the TRF program.

Candidate VNTRs	Primers		Candidate VNTRs	Primers	
Ea29	F	AGCCATGATCACTCCCGTAG	Ea43	F	GCGGGTTTTATGGCTGACTA
	R	CTGACGACAACTCACGGCTA		R	CGTAAACCGCAAGAAGAAGC
Ea30	F	CTTTATCATGGGTGTCAAGCA	Ea44	F	GGTTATTTTCCCACGTTCCA
	R	GCCACTTTGCATTGAAAGAA		R	CGACATATCCACCACAATGC
Ea31	F	CTTACTCACCGGCATGATGA	Ea45	F	CACCAGTTTCGAGCCATTTT
	R	CAGCGAACCGCCTAACATA		R	AATATCCTGGCGCTTAATGCT
Ea32	F	GGAAATCAAGCGGTGAGGTA	Ea46	F	ATGGCAGGAGTCTTGTTGCT
	R	CAGTGCCTCTTCCGGTACTT		R	CCTCCCGCCATTTGAATAC
Ea33	F	GAGTATCAACCCGCAGAGGA	Ea47	F	GCTGGTCATTGGTGCCTAAC
	R	TTATCGTTGAAGCGGACATTT		R	GGGGTGAAAAACATCATTGC
Ea34	F	CCGATATAAAAACCCTCAGGAA	Ea48	F	GCGTATGTATGGTGGCACTG
	R	CACATCACCGATGTCGATTT		R	GGAAGCAGGAGACCAGCTC
Ea35	F	GTGTCGGTACAGGGATTTGC	Ea49	F	CGCTATCAACAGCAGCAGAA
	R	GCACGGCACCATTTTTATTC		R	GTGGTTAACCGGTCGTATCG
Ea36	F	CGGGCTGATTGTTTTCTGTC	Ea50	F	GACGTAAACGGCCAAACTGT
	R	TGGCGTGATCTGAAGCTTTT		R	GCCGCTGTTAATCCCTGATA
Ea37	F	GTGGGAGGTATCCAGAATCG	Ea51	F	CGGGTAATGCTGCCTGAG
	R	CTACTCGCTGGTGATGATGC		R	GTTTAGCCCGCTTGTGGAT
Ea38	F	GCTGATGAACAATGCCGATA	Ea52	F	CATCAGCAAACCGATCAACA
	R	GAGGTGAGGTTAAGGGCTGA		R	CTCCAGGAGCGCTCTCAAT
Ea39	F	GGTGAATAGGAGGCGCAATA	Ea53	F	AAGCTATGCGGCAATGAAAG
	R	TGGATCAATCAGCTCCGTTT		R	GAAACGCGTCCAATCACTTC
Ea40	F	TGTCACAGGGTAAGCGTCAG	Ea54	F	GCAGGCTATGGTTGGTCAGT
	R	GGTGGATTAACCCCATTCCT		R	TGCAGGAGAAGCTGGTGAC
Ea41	F	CGGTCACGTCCTTCATTTTC	Ea55[Ŧ]	F	GCCAGCGGATATCCCTAAA
	R	CCTGTGCTAGTTTCTGCCTGA		R	AAGCCGGATGTCTTTCCTTT
Ea42	F	CGGTCAAAAACAACTGCAAA	Ea56[Ŧ]	F	GGAGACCCGTGATACAGGAA
	R	ATGGCCATAATGACCGTAGC		R	GCAGACCACCGGGTAAAATA

[Ŧ]Ea located on the plasmid pEA29.

Thereafter, a QIAxcel multicapillary electrophoresis system was used to assess the 56 candidate VNTRs in the DNA of the 46 *E. amylovora* strains. Figure 33 illustrates an example of the amplicons estimation process using the primer pairs for Ea41 locus in the DNA of two *E. amylovora* strains LEB8 and OMPER25.1. A total of nine VNTR loci, seven of them are located on the chromosome (Ea01, Ea09, Ea27, Ea30, Ea34, Ea39, and Ea41) and the other two are on the almost ubiquitous plasmid pEA29 (Ea55 and Ea56), varied clearly in size between the tested isolates, with one to nine alleles, whereas the rest 46 loci showed no variation at all and subsequently they were excluded from the study. As a result, a set of nine VNTR loci were found to be polymorphic and promising for the genotyping of *E. amylovora* strains of different origins, year of isolation, and host plants. Six of these polymorphic VNTRs were reported previously (Bühlmann *et al.*, 2014). Table 13 shows the estimated sizes of the amplicons produced by the selected VNTRs in the DNA of the studied strains.

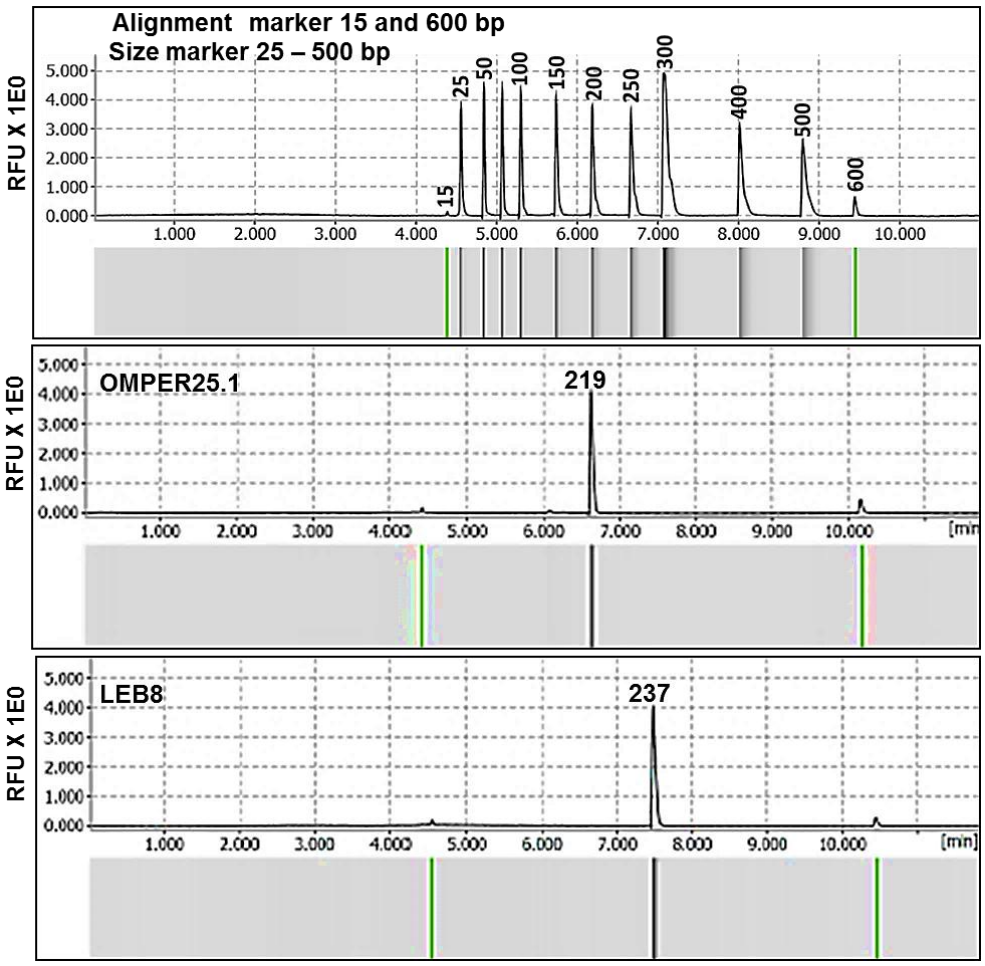


Figure 33 Analysis the DNA of two *Erwinia amylovora* strains with the VNTR locus Ea41 by QIAxcel multicapillary electrophoresis system (QIAGEN, USA); the green bands in the gel images (the gray areas) represent the two bands (15 and 600 bp) of the alignment marker.

Table 13 The amplicon sizes of the selected set of the VNTR loci in the studied list of *Erwinia amylovora* strains.

Strain's ID	PCR product size (bp)								
	Ea01	**Ea09**	**Ea27**	**Ea30**	**Ea34**	**Ea39**	**Ea41**	**Ea55**	**Ea56**
DZC4A	166	241	235	141	204	236	221	167	231
E1E	176	258	237	146	205	239	231	167	240
C1A	175	256	238	146	205	238	231	166	237
Ea668/05	178	256	240	159	245	232	232	169	232
Ea4/82	178	259	240	154	226	231	234	177	240
CFBP2301	131	240	236	155	219	227	220	182	232
CFBP1430	177	258	238	159	225	230	233	176	239
EA1	167	240	237	135	199	236	222	199	231
Ea33	177	258	240	153	226	231	234	160	240
Ea263	177	212	239	159	225	230	234	176	239
E204	174	257	237	141	223	231	231	182	240
EAI204	175	260	237	144	223	234	231	182	243
OMPBO1077,7/94	166	239	235	144	202	227	219	183	231
OMPER25.1	166	239	236	143	202	227	219	182	232
OMPBO8630,1/10	177	254	240	148	206	230	231	175	236
OMPBO347,1/06	179	252	240	121	205	229	231	174	236
OMPBO329,2/02	177	252	240	146	204	229	230	166	236
OMPER 22481,1A	178	253	240	147	206	229	231	167	237
KOS36	168	244	239	141	224	229	224	215	237
KOS45	168	241	239	141	223	230	221	215	237
KOS43	178	254	241	141	223	230	231	175	237
KOS67	180	254	241	141	224	230	231	201	239
KOS61	178	254	240	141	223	230	231	215	238
CFBP4781	176	257	240	158	231	229	233	175	239
LebA3	178	256	241	160	226	232	232	187	241
LEB1	167	239	235	160	190	227	220	191	231
LEB2	177	254	242	166	195	231	231	193	240
LEB3	180	257	244	168	197	233	235	195	241
LEB4	174	208	237	157	229	228	232	181	241
LEB6	173	255	237	162	193	228	233	182	241
LEB7	177	254	238	163	193	228	235	183	241
LEB8	175	208	228	151	229	228	237	175	241
KFB146	178	254	240	148	248	230	231	160	238
KFB150	184	244	238	154	219	227	224	184	231
KFB187	167	196	240	134	248	230	220	208	237
KFB153	167	194	236	153	219	230	220	191	231
KFB161	177	208	240	135	241	239	230	231	237
KFB162	177	253	239	157	222	229	231	190	237
UPN527	176	257	241	158	229	230	233	160	243
ACW56400	176	258	205	145	206	229	232	168	236
ACW63579	176	258	241	158	224	229	233	183	238
TK119	176	257	241	159	224	230	232	160	240
CFBP1232	176	258	239	158	206	229	233	167	239
UTFer2	176	259	242	171	218	230	233	209	233
EA153	177	259	239	159	231	230	234	175	227
JL1185	177	256	222	141	195	205	223	169	231

3.3.2. VNTRs characteristics

The allelic variability at the nine VNTR loci among the 46 isolates is summarized in Table 14. The highest number of different alleles was revealed at locus Ea55 (9), followed by the loci Ea30 (8), Ea34 (8), Ea01 (3), Ea09 (3), Ea27 (3), Ea39 (3), Ea41 (3), and Ea56 (3) (Table 15). Simpson's D index that measures the discriminatory power of the studied discriminators, which are the VNTR loci in our case, defines the ability of these loci to distinguish between unrelated strains. Simpson's D value ranges between 0 and 1, and the higher the value of the index in a given locus, the greater the discriminatory power of that locus. The highest diversity was observed in VNTR loci Ea55, Ea34, and Ea30 with diversity indices of 0.86, 0.84, and 0.82 respectively, whereas the lowest discriminatory power was in Ea27 (0.14).

Table 14 The list of the studied *Erwinia amylovora* strains and their MLVA profile.

Strain's ID	MLVA allelic profile								
	Ea01	**Ea09**	**Ea27**	**Ea30**	**Ea34**	**Ea39**	**Ea41**	**Ea55**	**Ea56**
DZC4A	2	2	7	7	8	9	3	6	4
E1E	3	3	7	8	8	9	4	6	5
C1A	3	3	7	8	8	9	4	6	4
Ea668/05	3	3	7	10	15	8	4	6	4
Ea4/82	3	3	7	9	12	8	5	7	5
CFBP2301	3	3	7	10	11	8	4	7	5
CFBP1430	3	3	7	10	8	8	4	6	5
EA1	2	2	7	6	7	9	3	10	4
Ea33	3	3	7	9	12	8	5	5	5
Ea263	3	2	7	10	11	8	5	7	5
E204	3	3	7	7	11	8	4	8	5
EAI204	3	3	7	7	11	9	4	8	5
OMPBO1077,7/94	2	2	7	7	8	8	3	8	4
OMPER25.1	2	2	7	9	8	8	3	8	4
OMPBO8630,1/10	3	3	7	8	8	8	4	7	4
OMPBO347,1/06	3	3	7	4	8	8	4	7	4
OMPBO329,2/02	3	3	7	8	8	8	4	6	4
OMPER 22481,1A	3	3	7	8	8	8	4	6	4
KOS36	2	2	7	7	11	8	3	12	4
KOS45	2	2	7	7	11	8	3	12	4
KOS43	3	3	7	7	11	8	4	7	4
KOS67	3	3	7	7	11	8	4	10	5
KOS61	3	3	7	7	11	8	4	12	5
CFBP4781	1	2	7	9	10	8	3	8	4
LebA3	3	3	7	10	12	8	4	8	5
LEB1	3	3	7	11	6	8	4	9	5
LEB2	3	3	7	11	7	9	5	9	5
LEB3	3	2	7	10	12	8	4	7	5
LEB4	3	3	7	10	6	8	4	8	5
LEB6	3	3	7	11	6	8	5	8	5
LEB7	3	2	6	9	12	8	5	7	5
LEB8	2	2	7	10	6	8	3	9	4
KFB146	3	3	7	8	15	8	4	5	5
KFB150	3	2	7	9	10	8	3	8	4
KFB187	2	1	7	6	15	8	3	11	4
KFB153	2	1	7	9	10	8	3	9	4
KFB161	3	2	7	6	14	9	4	14	4
KFB162	3	3	7	10	11	8	4	9	4
UPN527	3	3	7	10	12	8	4	5	5
ACW56400	3	3	5	8	8	8	4	6	4
ACW63579	3	3	7	10	11	8	4	8	5
TK119	3	3	7	10	11	8	4	5	5
CFBP1232	3	3	7	10	12	8	4	7	5
UTFer2	3	3	7	12	10	8	4	11	4
EA153	3	3	7	10	12	8	5	7	3
JL1185	3	3	6	7	6	5	3	6	4

Table 15 Characteristics of the selected set of VNTR loci in which D: Simpson's diversity index. VNTRs Ea55 and Ea56 are on the plasmid pEA29 and the other VNTRs are on the chromosome.

VNTR locus	Target region	No. of alleles		D
Ea01	Hypothetical protein	3	All	0.36
		2	Italy	0.43
Ea09	*agp*	3	All	0.47
		2	Italy	0.43
Ea27	*yfgA*	3	All	0.13
		1	Italy	0
Ea30	Hypothetical protein	8	All	0.82
		4	Italy	0.79
Ea34	Hypothetical protein	8	All	0.84
		2	Italy	0.43
Ea39	Hypothetical protein	3	All	0.30
		2	Italy	0.25
Ea41	Hypothetical protein	3	All	0.58
		2	Italy	0.43
Ea55	Repeat region on pEA29	9	All	0.86
		3	Italy	0.71
Ea56	Repeat region on pEA29	3	All	0.53
		2	Italy	0.43

3.3.3. MLVA genotyping

The DNA from all the studied *E. amylovora* strains was genotyped reproducibly at the nine VNTR loci since the duplicate strains, within the same run or between runs, produced fully consistent results. Minimum Spanning Trees (MSTs) have proven to be a powerful tool to understand and decode the MLVA profiles. Each node in the MST represents one strain or a group of strains that have identical MLVA haplotype, and the size of the node reflects the number of the strains in the group. A total of seven Clonal Complexes (CCs) and 13 singletons were distinguished among the 38 haplotypes (Figure 35). The biggest complex CC-I comprised 12 different strains isolated mainly from apple and pear: eight of them Mediterranean (Egypt, France, Israel, Kosovo, Lebanon, Serbia, Spain, and Turkey), one from Switzerland, one from Iran, one from the UK, and one North American. However, no Italian strains were found in this common CC. The second complex CC-II was formed by seven strains: four of them were isolated from Kosovo, two from Italy, and one from Lebanon. The third one CC-III contained five strains isolated from Italy (four strains) and France (one strain). The fourth complex CC-IV contained three Algerian strains isolated from pear plants only. The fifth one CC-V contained two Italian strains isolated from pear plants and had an identical MLVA profile. Finally, the sixth and the seventh complexes, CC-VI and CC-VII respectively, were formed by two Lebanese strains each. Although the strains of the sixth and the seventh CCs had an identical MLVA profile, they were isolated form apple solely in one complex (CC-VI) and from apple and pear in the other (CC-VII).

On the Mediterranean basin countries scale, the highest diversity was noticed among the strains from the Balkan Peninsula region (Southeast Europe) with 11 haplotypes, whereas less diversity was noticed among the Southern European (eight haplotypes) and the Eastern Mediterranean strains (seven haplotypes) as shown in Figure 34.

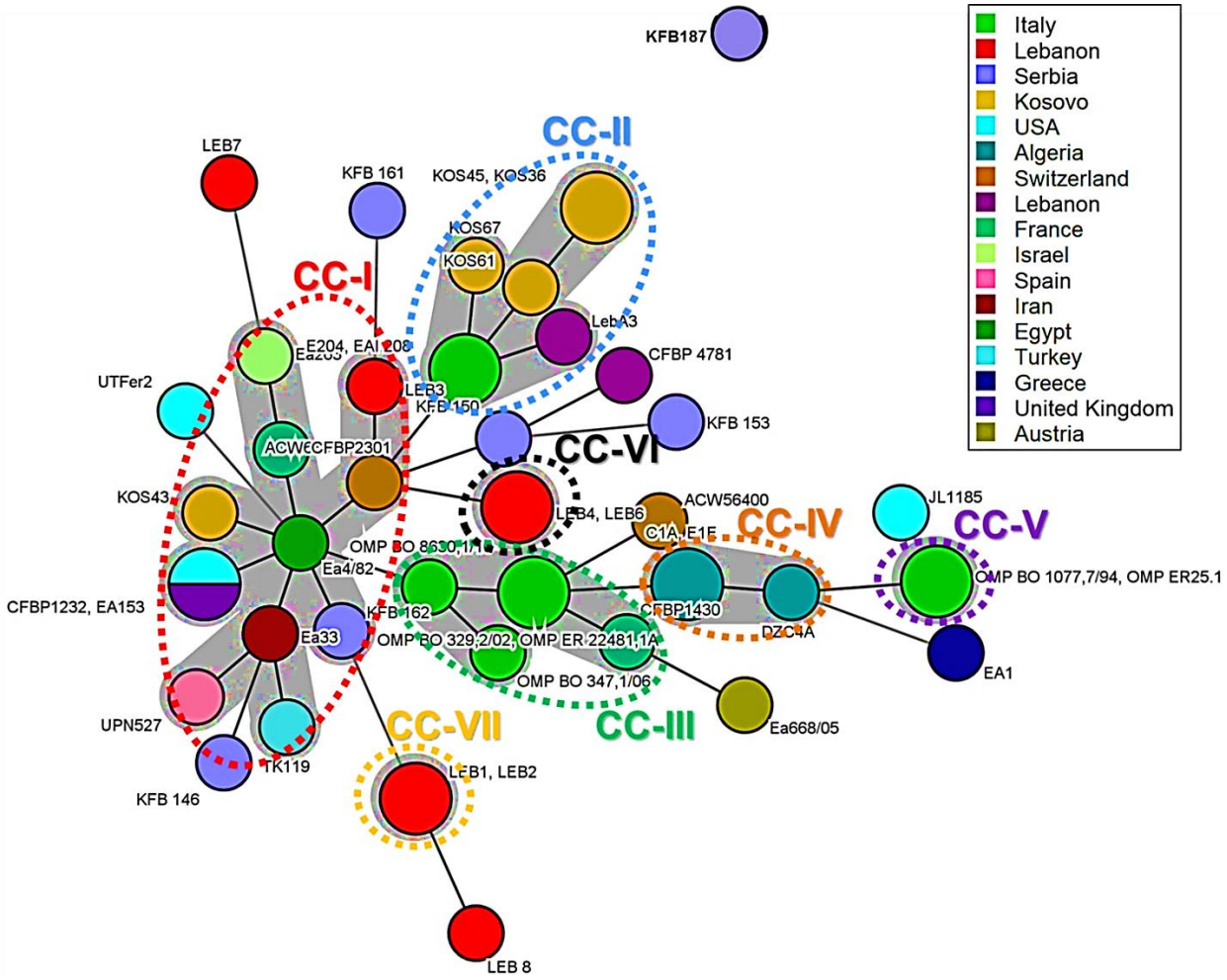


Figure 34 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains and organized by their origin. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci.

On the national scale, the DNA of only eight *E. amylovora* strains collected from different Italian territories were accessible and used in this study. Three Clonal Complexes and five haplotypes were identified from these studied strains (Figure 34). Lower diversity was observed compared with the global collection of strains in this study (Table 15). The relatively higher diversity within the Italian strains reflects the multiple introduction events of the pathogen in a relatively short period, and it may also indicates the globally high commercial trafic of the fire blight-infected plant materials with no effective control on the introduction.

The MST organized by the year of isolation showed that the oldest strains in our *E. amylovora* list (EA153, CFBP1232, Ea4/82, and CFBP2301) were clustered in the same Clonal Complex, CC-I, and in a temporally logical way supporting the reports about the history of the disease dissemination from the USA, through the UK, to EU countries and Egypt (Figure 35). The Italian strain OMPBO1077.7/94, which was isolated during the first outbreaks in the early 1990s, appeared to be introduced from a different *E. amylovora* population than CC-I. This OMPBO1077.7/94 strain seemed more closely related to CC-IV, which contained only Algerian strains, as demonstrated in Figure 35 suggesting that the disease was introduced into Italy in the same period as into Algeria, with no official reports from Algeria, or the disease was imported from North Africa to Italy as it was reported from Egypt 30 years earler. However, these results need to be further investigated and confirmed by other genotyping methods like the Next-Generation Sequencing (NGS) or the Single Nucleotide Polymorphism (SNP) analysis. Furthermore, no apparent pattern of year to year difference was found on the national scale (Figure 35) suggesting that the same strains reinfect the orchards and cause the disease every year.

Regarding to the MST organized by the host plants. Although the isolated strains from hawthorn and quince are separately clustered in the same Clonal Complex, Figure 35 shows that the different host species, mainly apple and pear in our study, did not drive any divergence of *E. amylovora* populations in which no obvious trends or patterns were revealed. This result is congruent with Bühlmann *et al*. (2014) finding that the bacterium is well adapted to the studied hosts.

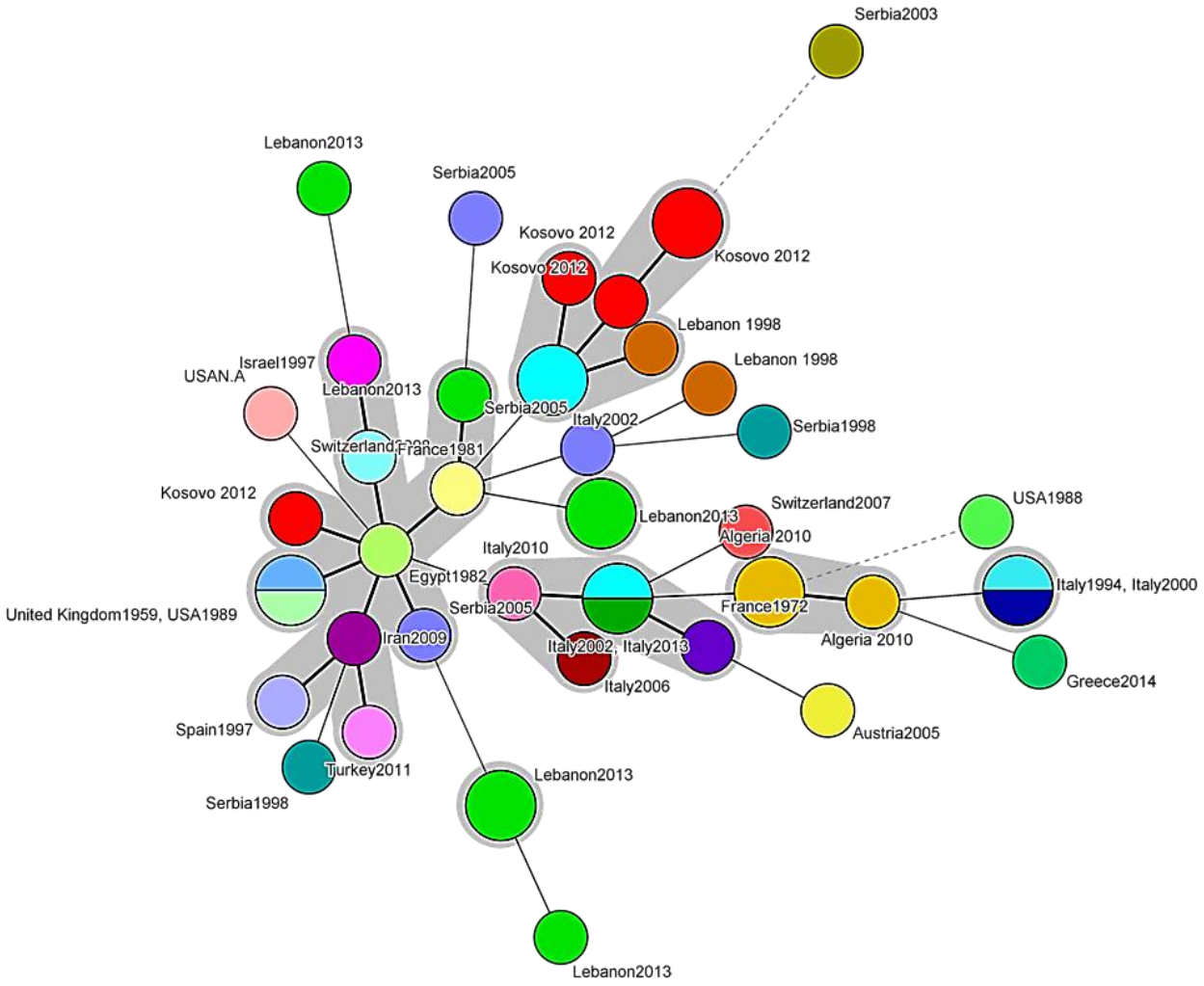


Figure 35 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains and organized by their year of isolation. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci.

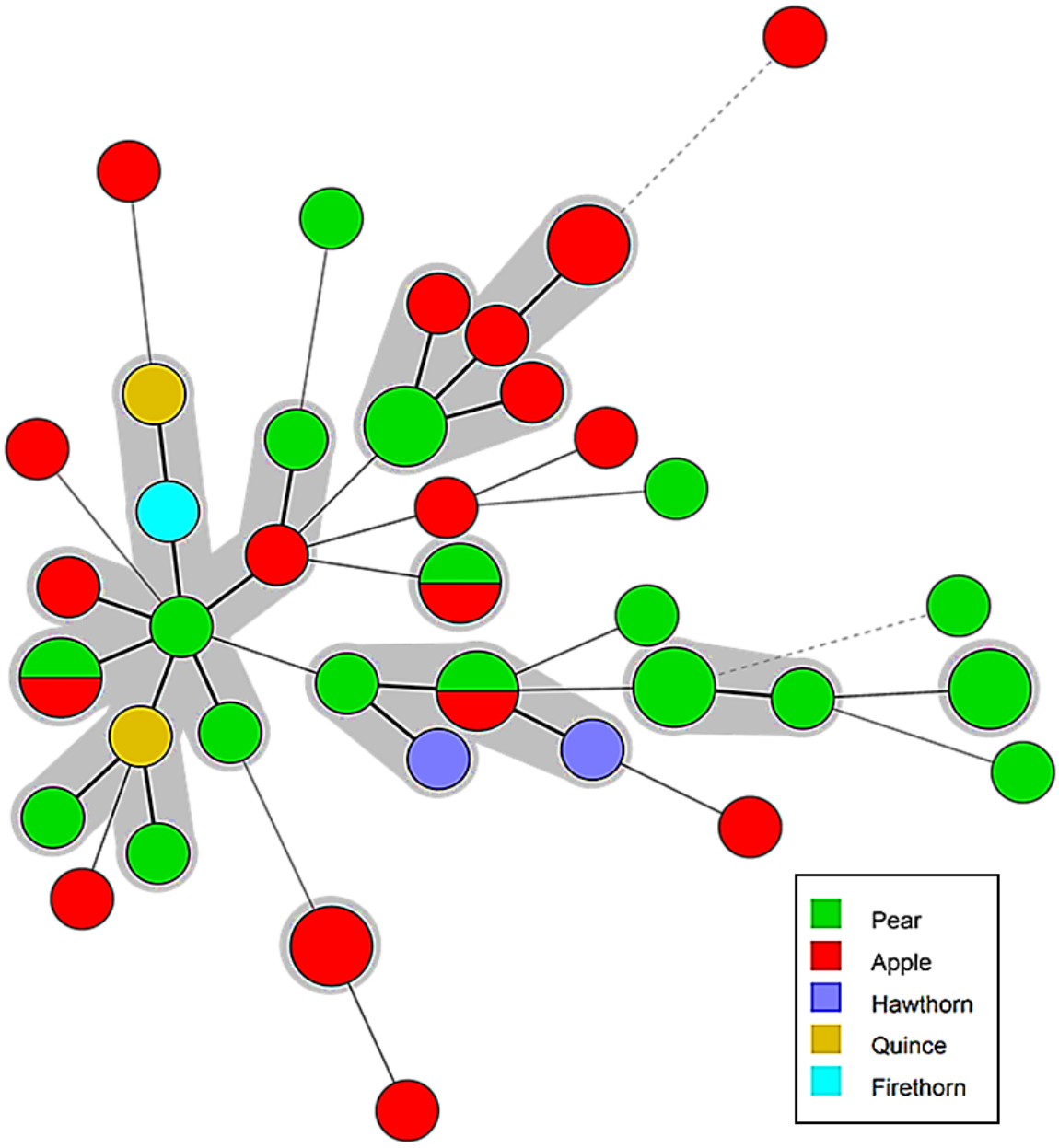


Figure 36 Minimum Spanning tree of MLVA profiles from the DNA of the 46 *Erwinia amylovora* strains organized by their host plants. Branch thickness indicates how many loci are different in the MLVA profiles of the connected nodes. Bold solid lines connect nodes that differ by one MLVA locus, thin solid lines connect nodes that differ by two MLVA loci and dashed lines connect nodes that differ by three MLVA loci.

3.4. Conclusions

In the current study, an improved MLVA assay was developed and successfully used for genotyping different *Erwinia amylovora* strains mainly collected from Mediterranean basin countries. The Whole-Genome Sequence (WGS) data of only two reference strains were used to develop a panel of nine VNTR markers capable of discriminating between *E. amylovora* strains collected from different hosts in different geographic origins with a good resolution comparing with other genotyping techniques. Three VNTR loci of the newly developed panel were revealed by this study and the other six were reported in a previous study in 2014.

The results from *In-silico* and MLVA analyses showed that all the studied VNTR loci are stable enough for any future epidemiological, evolutional, or biological studies. Additionally, the varied number of alleles in each VNTR loci, which was ranged from three to nine, can theoretically discriminate a huge number of haplotypes.

The relatively high diversity among the studied Italian strains confirmed also the multiple introduction events of the pathogen from different countries and call for increasing attention to the plant materials traffic from and to these regions. At least three *E. amylovora* populations are currently present and active in Italy; two of these populations are closely related to strains from North Africa (Algeria and Egypt) and Central Europe (Switzerland).

Unless WGS data becomes cost effective, MLVA analysis will not fade away in the near future. Since adding new polymorphic VNTR loci is one of the biggest advantages for the MLVA technique, the current panel can be further improved based on the availability of new WGS data from other strains isolated from different host plants in different countries around the globe. These data will increase the resolution of the MLVA analysis of the genetically monomorphic pathogens. Moreover, MLVA assay can be facilitated by including all the VNTRs primer pairs in one multiplex PCR to save the time.

References

(2005). NPPO of Italy. 02.

(2009). NPPO of Italy. 07.

Achtman M. (2008). Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. *Annual Review of Micorobiolgy*, 62: 53-70.

Alnaasan Y. (2012). *Evaluation of sampling methods for assessing Citrus tristeza virus incidence in Apulian citrus groves*. CIHEAM-MAIB, Bari, Italy. Master of Science. Integrated Pest Management

Asner G. P. (1998). Biophysical and Biochemical Sources of Variability in Canopy Reflectance. *Remote Sensing of Environment*, 64(3): 234-253.

Ballabio D. and Consonni V. (2013). Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods*, 5: 3790-3798.

Ballabio D. and Todeschini R. (2008). Multivariate Classification for Qualitative Analysis. In: (ed). *Infrared Spectroscopy for Food Quality Analysis and Control*. Elsevier, Da-Wen Sun (ED), pp. 83-104.

Ballabio D. and Vasighi M. (2012). A MATLAB Toolbox for Self Organizing Maps and supervised neural network learning strategies. *Chemometrics and Intelligent Laboratory Systems*, 118: 24-32.

Ballabio D., Vasighi M., Consonni V. and Kompany-Zareh M. (2011). Genetic Algorithms for architecture optimisation of Counter-Propagation Artificial Neural Networks. *Chemometrics and Intelligent Laboratory Systems*, 105(1): 56-64.

Barker M. and Rayens W. (2003). Partial least squares for discrimination. *Journal of chemometrics*, 17(3): 166-173.

Barnes J. D., Balaguer L., Manrique E., Elvira S. and Davison A. W. (1992). A Reappraisal of the Use of DMSO for the Extraction and Determination of Chlorophylls-a and Chlorophylls-B in Lichens and Higher-Plants. *Environmental and Experimental Botany*, 32(2): 85-100.

Benson G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Research*, 27(2): 573-580.

Billing E. and Baker L. A. E. (1963). Characteristics of Erwinia-like Organisms found in Plant Material. *Journal of Applied Microbiology*, 26(1): 58-65.

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.

Blackmore B. S. (1994). Precision Farming: an introduction. *Outlook on agriculture*, 23(4): 275-280.

Bonn W. G. and van der Zwet T. (2000). Distribution and econonmic importance of fire blight. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 37-53.

Braun P. G. and Hildebrand P. D. (2005). Infection, carbohydrate utilization, and protein profiles of apple, pear, and raspberry isolates of *Erwinia amylovora*. *Canadian Journal of Plant Pathology*, 27(3): 338-346.

Bühlmann A., Dreo T., Rezzonico F., Pothier J. F., Smits T. H. M., Ravnikar M., Frey J. E. and Duffy B. (2014). Phylogeography and population structure of the biologically invasive phytopathogen *Erwinia amylovora* inferred using minisatellites. *Environmental Microbiology*, 16(7): 2112-2125.

Burdon J. J. (1987). *Diseases and Plant Population Biology*. Cambridge University Press.

Calzolari A. (1984). Survey for the presence of fire blight in some Italian fruit growing areas. *Acta Horticulturae*, 151: 329-334.

Calzolari A., Finelli F. and Mazzoli G. L. (1999). A severe unforeseen outbreak of fire blight in the Emilia-Romagna region. *Acta Horticulturae*, 489: 171-176.

Calzolari A., Peddes P., Mazzucchi U., Mori P. and Garzena C. (1982). Occurrence of *Erwinia amylovora* in buds of asymptomatic apple plants in commerce. *Phytopathologische Zeitschrift*, 103: 156-162.

Cariddi C. (1990). Colpo di fuoco sul pero. *Terra e vita*, 34: 67-69.

Cariddi C. (1993). Further observations on fire blight in Italy. (eds). *6th International Workshop on Fireblight, Athens, Greece*. Acta Holticulturae, Athens, Greece. 338, 107-110.

Cariddi C. and Piglionica V. (1990). First record of fire blight in Italy. (eds). *Proceeding of the 8th congress of the Mediterranean Phytopathological Union, Agadir, Morocco,* Agadir, Morocco.

Chen J. M. (1996). Evaluation of vegetation indices and a modified simple ratio for boreal applications. *Canadian Journal of Remote Sensing*, 22(3): 229-242.

Cierniewski J. and Verbrugghe M. (1997). Inferring soil surface roughness from soil bidirectional reflectance data. *International Agrophysics*, 11(147-157).

Clabassi I., Petris G. and Belloni S. (2002). Il colpo di fuoco batterico su pero: prima segnalazione nelle regione Friuli Venezia Giulia. Notiziario ERSA, Agenzia regionale per lo sviluppo rurale. 3**:** 2.

Crosse J. E., Bennett M. and Garrett C. M. E. (1958). Fire-blight of pear in England. *Nature*, 182: 1530.

Daughtry C. S. T., Walthall C. L., Kim M. S., de Colstoun E. B. and McMurtrey Iii J. E. (2000). Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment*, 74(2): 229-239.

Deckers T. and Daemen E. (1993). Influence of chemical growth regulation on the host susceptibility of pear trees for fire blight. *Acta Horticulturae*, 338: 205-215.

Deckers T., Porreye W. and Maertens P. (1990). Three years of experience in chemical control of fire blight in pear orchards in Belgium. *Acta Horticulturae*, 273: 367-376.

Denning W. (1794). On the decay of apple trees. Transactions of the Society for the Promotion of Agriculture, Arts and Manufacturers. 1**:** 185-187.

Dewdney M. M., Biggs A. R. and Turechek W. W. (2007). A Statistical Comparison of the Blossom Blight Forecasts of MARYBLYT and *Cougarblight* with Receiver Operating Characteristic Curve Analysis. *Phytopathology*, 97(9): 1164-1176.

Donat V., Biosca E. G., Peñalver J. and López M. M. (2007). Exploring diversity among Spanish strains of *Erwinia amylovora* and possible infection sources. *Journal of Applied Microbiology*, 103(5): 1639-1649.

EPPO (2014). Plant Quarantine data Retrieval system PQR.

FAO (2015). https:// www.faostat.fao.org.

Fourty T., Baret F., Jacquemoud S., Schmuck G. and Verdebout J. (1996). Leaf optical properties with explicit description of its biochemical composition: Direct and inverse problems. *Remote Sensing of Environment*, 56(2): 104-117.

Gamon J. A., Peñuelas J. and Field C. B. (1992). A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency. *Remote Sensing of Environment*, 41(1): 35-44.

Gausman H. W. (1977). Reflectance of leaf components. *Remote Sensing of Environment*, 6(1): 1-9.

Gehring I. and Geider K. (2012). Differentiation of *Erwinia amylovora* and *Erwinia pyrifoliae* strains with single nucleotide polymorphisms and by synthesis of dihydrophenylalanine. *Current Microbiology*, 65: 73-84.

Geider K., Auling G., Du Z., Jakovljevic V., Jock S. and Volksch B. (2006). *Erwinia tasmaniensis* sp. nov., a non-phytopathogenic bacterium from apple and pear trees. *International Journal of Systematic and Evolutionary Microbiology*, 56: 2937-2943.

Geladi P. and Kowalski B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185(0): 1-17.

Gianetti G., Garofalo M. C., Morone C. and Scortichini M. (2004). Primo rinvenimento di *Erwinia amylovora* in Piemonte. *L'Informatore agrario*, 5: 62-63.

Gitelson A. and Merzlyak M. N. (1994). Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves. *Journal of Photochemistry and Photobiology B: Biology*, 22(3): 247-252.

Gitelson A. A., Merzlyak M. N., Zur Y., Stark R. and Gritz U. (2001). Non-destructive and remote sensing techniques for estimation of vegetation status. In: (eds). *Third European Conference on Precision Agriculture, Montpellier, France*, Montpellier, France. 1.

Griffo R., Tartaglia A., Capriolo F. and Adamo F. (1998). Il colpo di fuoco batterico in Campania. *Informatore Agrario*, 54(29): 73-75.

Gur-Arie R., Cohen C. J., Eitan Y., Shelef L., Hallerman E. M. and Kashi Y. (2000). Simple sequence repeats in *Escherichia coli*: abundance, distribution, composition, and polymorphism. *Genome Research,* 10(1): 62-71.

Guyot G., Baret F. and Major D. J. (1988). High spectral resolution: Determination of specral shifts between the red and the near infrared. (eds). *International Archives of Photogrammetry and Remote Sensing, XVIth ISPRS Congress, Technical Commission VII: Interpretation of Photographic and Remote Sensing Data, July 1-10, 1988, Kyoto, Japan.*, XVIth ISPRS Congress, Technical Commission VII: Interpretation of Photographic and Remote Sensing Data, July 1-10, 1988, Kyoto, Japan. 11, 750-760.

Hand D. J. (1997). *Construction and Assessment of Classification Rules*. Wiley, Chichester, UK.

Hauben L., Moore E. R. B., Vauterin L., Steenackers M., Mergaert J., Verdonck L. and Swings J. (1998). Phylogenetic Position of Phytopathogens within the Enterobacteriaceae. *Systematic and Applied Microbiology*, 21(3): 384-397.

Heald C. M., Thames W. H. and Wiegand C. L. (1972). Detection of *Rotylenchulus renifromis* infestations by aerial infrared photography. *Journal of Nematology*, 4: 298-300.

Henze N. and Zirkler B. (1990). A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*, 19(10): 3595-3617.

Huete A. R. and Liu H. Q. (1994). An error and sensitivity analysis of the atmospheric- and soil-correcting variants of the NDVI for the MODIS-EOS. *Geoscience and Remote Sensing, IEEE Transactions on*, 32(4): 897-905.

Jock S., Donat V., López M. M., Bazzi C. and Geider K. (2002). Following spread of fire blight in Western, Central and Southern Europe by molecular differentiation of *Erwinia amylovora* strains with PFGE analysis. *Environmental Microbiology*, 4(2): 106-114.

Johnson K. B. (2000). *Fire blight of apple and pear*. American Phytopathological Society net.

Jolliffe I. T. (2002). *Principal Component Analysis*. Springer-Verlag, New York.

Jordan C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, 50(4): 663-666.

Kim W. and Geider K. (1999). Analysis of variable short-sequence DNA repeats on the 29 kb plasmid of *E. amylovora* strains. *European Journal of Plant Pathology*, 105: 703-713.

Kim W. S., Gardan L., Rhim S. L. and Geider K. (1999). *Erwinia pyrifoliae* sp. nov., a novel pathogen that affects Asian pear trees (*Pyrus pyrifolia* Nakai). *International Journal of Systematic Bacteriology*(49): 899-906.

Kim W. S., Jock S., Paulin J. P., Rhim L. S. and Geider K. (2001). Molecular detection and differentiation of *Erwinia pyrifoliae* and host range analysis of the Asian pear pathogen. *Plant disease*, 85(11): 1183-1186.

Klecka W. R. (1980). *Discriminant analysis*. Sage Publications, Beverly Hills and London. Quantitative Applications in the Social Sciences.

Kunz S., Schmitt A. and Haug P. (2011). Development of strategies for fire blight control in organic fruit growing. *Acta Horticulturae*, 896: 431-436.

Laala S., Manceau C., Valentini F., Kerkoud M. and Kheddam M. (2012). Fireblight survey and first characterization of *Erwinia amylovora* isolates from Algeria. *Journal of plant patholohy*, 94(3): 693-696.

Laudien R., Bareth G. and Doluschitz R. (2003). Analysis of hyperspectral field data for detection of sugar beet diseases. In: (eds). *Proceedings of the EFITA conference, Debrecen, Hungary*, Debrecen, Hungary.

Le Fleche P., Hauck Y., Onteniente L., Prieur A., Denoeud F., Ramisse V., Sylvestre P., Benson G., Ramisse F. and Vergnaud G. (2001). A tandem repeats database for bacterial genomes: application to the genotyping of *Yersinia pestis* and *Bacillus anthracis*. *BMC Microbiology*, 1(2): 30.

le Maire G., Francois C. and Dufrene E. (2004). Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements. *Remote Sensing of Environment*, 89(1): 1-28.

Leardi R. and Nørgaard L. (2004). Sequential application of backward interval PLS and genetic algorithms for the selection of relevant spectral regions. *Journal of chemometrics*, 18(11): 486-497.

Legendre M., Pochet N., Pak T. and Verstrepen K. J. (2007). Sequence-based estimation of minisatellite and microsatellite repeat variability. *Genome Research*, 17(12): 1787-1796.

Lenney M. P., Woodcock C. E., Collins J. B. and Hamdi H. (1996). The status of agricultural lands in Egypt: The use of multitemporal NDVI features derived from landsat TM. *Remote Sensing of Environment*, 56(1): 8-20.

Lichtenthaler H. K. (1996). Vegetation Stress: an Introduction to the Stress Concept in Plants. *Journal of Plant Physiology*, 148(1–2): 4-14.

Lichtenthaler H. K. and Rinderle U. (1988). The Role of Chlorophyll Fluorescence in The Detection of Stress Conditions in Plants. *C R C Critical Reviews in Analytical Chemistry*, 19(sup1): S29-S85.

Lillesand T. M., Kiefer R. W. and Chipman J. W. (2008). *Remote sensing and image interpretation*. John Wiley & Sons.

Llop P., Bonaterra A., Peñalver J. and López M. M. (2000). Development of a Highly Sensitive Nested-PCR Procedure Using a Single Closed Tube for Detection of *Erwinia amylovora* in Asymptomatic Plant Material. *Applied and Environmental Microbiology*, 66(5): 2071-2078.

López M. M., Roselló M., Llop P., Ferrer S., Christen R. and Gardan L. (2011). *Erwinia piriflorinigrans* sp. nov., a novel pathogen that causes necrosis of pear blossoms. *International Journal of Systematic and Evolutionary Microbiology*, 61(3): 561-567.

Lorenzen B. and Jensen A. (1989). Changes in leaf spectral properties induced in barley by cereal powdery mildew. *Remote Sensing of Environment*, 27: 201-209.

M.D. (1999). Regolamento recante misure per la lotta obbligatoria contro il colpo di fuoco batterico (*Erwinia amylovora*) D.M. 356/99. Ministero delle politiche agricole e forestali Gazzetta ufficiale della repubblica Italiana: GU Serie Generale n.243 del 15-10-1999.

Maes M., Orye K., Bobev S., Devreese B., Van Beeumen J., De Bruyn A., Busson R., Herdewijn P., Morreel K. and Messens E. (2001). Influence of amylovoran production on virulence of *Erwinia amylovora* and a different amylovoran structure in *E. amylovora* isolates from Rubus. *European Journal of Plant Pathology*, 107(8): 839-844.

Mann R. A., Smits T. H. M., Bühlmann A., Blom J., Goesmann A., Frey J. E., Plummer K. M., Beer S. V., Luck J., Duffy B. and Rodoni B. (2013). Comparative Genomics of 12 Strains of *Erwinia amylovora* Identifies a Pan-Genome with a Large Conserved Core. *PLoS ONE*, 8(2): e55644.

Mansfield J., Genin S., Magori S., Citovsky V., Sriariyanum M., Ronald P., Dow M. A. X., Verdier V., Beer S. V., Machado M. A., Toth I. A. N., Salmond G. and Foster G. D. (2012). Top 10 plant pathogenic bacteria in molecular plant pathology. *Molecular Plant Pathology*, 13(6): 614-629.

Matsuura T., Mizuno A., Tsukamoto T., Shimizu Y., Saito N., Sato S., Kikuchi S., Uzuki T., Azegami K. and Sawada H. (2012). *Erwinia uzenensis* sp. nov., a novel pathogen that affects European pear trees (*Pyrus communis* L.). *International Journal of Systematic and Evolutionary Microbiology*, 62(Pt 8): 1799-1803.

Mazzucchi U. (1994). The fireblight monitoring network in northern Italy: origin and development. *EPPO Bulletin*, 24(4): 783-791.

Mazzucchi U., Stefani E. and Bazzi C. (1993). Setting up of a fire blight monitoring network in Northern Italy. *Acta Horticulturae*, 338: 177-182.

McManus P. S. and Jones A. L. (1995). Genetic fingerprinting of *Erwinia amylovora* strains isolated from tree-fruit crops and *Rubus* spp. *Phytopathology*, 85: 1547-1553.

McMichael J. K. (1890). Pear Blight. Annual reports of the fruit growers' association and entomological society of Ontario, Toronto, Canada. 22**:** 75-78.

Mergaert J., Hauben L., Cnockaert M. C. and Swings J. (1999). Reclassification of non-pigmented *Erwinia herbicola* strains from trees as *Erwinia billingiae* sp. nov. *International Journal of Systematic Bacteriology*, 49(2): 377-383.

Merton R. N. (1993). Monitoring community hysteresis using spectralshiftanalysis and the red-edgevegetationstressindex. In: (eds). *Seventh Annual JPL Airborne Earth Science Workshop, Pasadena, California*, Pasadena, California.

Merzlyak M. N., Gitelson A. A., Chivkunova O. B. and Rakitin V. Y. U. (1999). Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening. *Physiologia Plantarum*, 106(1): 135-141.

Mohan S. K. and Thomson S. V. (1996). An outbreak of fire blight in plums. *Acta Horticulturae*, 411: 73-76.

Momol M. T., Momol E. A., Lamboy W. F., Norelli J. L., Beer S. V. and Aldwinckle H. S. (1997). Characterization of *Erwinia amylovora* strains using random amplified polymorphic DNA fragments (RAPDs). *Journal of Applied Microbiology*, 82(3): 389-398.

Momol M. T., Norelli J. L., Piccioni D. E., Momol E. A., Gustafson H. L., Cummins J. N. and Aldwinckle H. S. (1998). Internal movement of *Erwinia amylovora* through symptomless apple scion tissues into the rootstock. *Plant disease*, 82(6): 646-650.

Norelli J. L., Aldwinckle H. S., Beer S. V. and Lamb R. C. (1987). The effects of virulence of *Erwinia amylovora* on the evaluation of fire blight resistance in Malus. *Phytopathology*, 77(11): 1551-1555.

Nørgaard L., Saudland A., Wagner J., Nielsen J. P., Munck L. and Engelsen S. B. (2000). Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from Near-Infrared spectroscopy. *Applied Spectroscopy*, 54: 413-419.

Norman G. G. and Fritz N. L. (1965). Infrared photography as an indicator of disease and decline in citrus trees. (eds). *Proceedings of the Florida State Horticulture Society*. 78, 59-63.

Notomi T., Okayama H., Masubuchi H., Yonekawa T., Watanabe K., Amino N. and Hase T. (2000). Loop-mediated isothermal amplification of DNA. *Nucleic Acids Research*, 28.

Nuclo R., Johnson K. B., Sugar D. and Stockwell V. O. (1996). Importance of secondary spread of bacterial antagonists in the biological control of *Erwinia amylovora*. *Acta Horticulturae*, 411: 297.

Oerke E. C. (2006). Crop losses to pests. *The Journal of Agricultural Science*, 144(1): 31-43.

Paulin J. P. (2000). *Erwinia amylovora*: General Characteristics, Biochemistry and Serology. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 87-115.

Penuelas J., Baret F. and Filella I. (1995). Semi-empirical indices to assess carotenoids/chlorophyll a ratio from leaf spectral reflectance. *Photosynthetica*, 31: 221-230.

Penuelas J., Filella I., Biel C., Serrano L. and SavÉ R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. *International Journal of Remote Sensing*, 14(10): 1887-1905.

Powney R., Beer S. V., Plummer K. M., Luck J. and Rodoni B. (2011). he specificity of PCR-based protocols for detection of *Erwinia amylovora*. *Australas Plant Pathology*, 40: 87-97.

Pucci N., L'Aurora A. and Loreti S. (2013). Fire blight: first report in Latium, Italy. *journal of plant Pathology*, 95(3): 659-668.

Rama Rao N., Garg P. K., Ghosh S. K. and Dadhwal V. K. (2008). Estimation of leaf total chlorophyll and nitrogen concentrations using hyperspectral satellite imagery. *The Journal of Agricultural Science*, 146(1): 65-75.

Ramsey R. D., Falconer A. and Jensen J. R. (1995). The relationship between NOAA-AVHRR NDVI and ecoregions in Utah. *Remote Sensing of Environment*, 53(3): 188-198.

Reby D., Lek S., Dimopoulos I., Joachim J., Lauga J. and Aulagnier S. (1997). Artificial neural networks as a classification method in the behavioural sciences. *Behavioural Processes*, 40(1): 35-43.

Reyniers M., Walvoort D. J. J. and 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(19): 4159-4179.

Rezzonico F., Smits T. H. M. and Duffy B. (2011). Diversity, Evolution, and Functionality of Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Regions in the Fire Blight Pathogen *Erwinia amylovora*. *Applied and Environmental Microbiology*, 77(11): 3819-3829.

Rhouma A., Helali F., Chettaoui M., Hajjouji M. and Hajlaoui P. (2014). First report of fireblight caused by *Erwinia amylovora* on pear in Tunisia. *Plant disease*, 98(1): 158.

Richardson A. D., Duigan S. P. and Berlyn G. P. (2002). An evaluation of noninvasive methods to estimate foliar chlorophyll content. *New Phytologist*, 153(1): 185-194.

Rico A., Ortiz-Barredo A., Ritter E. and Murillo J. (2004). Genetic characterization of *Erwinia amylovora* strains by amplified fragment length polymorphism. *Journal of Applied Microbiology*, 96(2): 302-310.

Roselló M., García Vidal S., Tarín A., Llop P., Gorris M. T., Donat V., Chartier R., Paulin J. P., Gardan L., Peñalver J. and López M. M. (2002). Characterization of an *Erwinia* sp. isolated from necrotic pear blossoms in Valencia, Spain. *Acta Horticulturae*, 590: 139-142.

Rosen H. R. and Groves A. B. (1928). Studies of fire blight: Host range. *Journal of Agricultural Research*, 37: 493-505.

Roujean J.-L. and Breon F.-M. (1995). Estimating PAR absorbed by vegetation from bidirectional reflectance measurements. *Remote Sensing of Environment*, 51(3): 375-384.

Rouse J. W., Jr., Haas R. H., Schell J. A. and Deering D. W. (1973). Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation. (eds). *Remote Sensing Center, Texas A&M Univ., College Station*. 93.

Rumelhart D. and McClelland J. (1986). *Parallel Distributed Processing.* MIT press, Cambridge, MA.

Santoro F. (pers comm.). *Post Processing Spectroradiometer*.

Santoro F., Gualano S., Djelouah K., Guario A. and D'Onghia A. M. (2009). Remote sensing to support the monitoring of *Citrus tristeza virus* (CTV) infected areas. *Options méditerranéennes, Série B: Studies and Research,*(35): 165-171.

Santoro F., Tarantino E., Figorito B., Gualano S. and D'Onghia A. M. (2011). A tree counting algorithm for precision agriculture tasks. *International Journal of Digital Earth*: 1-9.

Sebaihia M., Bocsanczy A. M., Biehl B. S., Quail M. A., Perna N. T., Glasner J. D., DeClerck G. A., Cartinhour S., Schneider D. J., Bentley S. D., Parkhill J. and Beer S. V. (2010). Complete Genome Sequence of the Plant Pathogen *Erwinia amylovora* Strain ATCC 49946 *Journal of Bacteriology*, 192(7): 2020-2021.

Sherman V. T. (2000). Epidemiology of fire blight. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 9-36.

Simpson E. H. (1949). Measurement of diversity *Nature*, 163(4148): 688.

Smith R. C. G., Adams J., Stephens D. J. and Hick P. T. (1995). Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite. Australian Journal of Agricultural Research. 46**:** 113-125.

Smith T. J. (1999). Report on the development and use of *CougarBlight* 98C- a situation-specific fire blight risk assessment model for apple and pear. *Acta Horticulturae*, 489: 429-436.

Smits T. H. M., Rezzonico F., Kamber T., Blom J., Goesmann A., Frey J. E. and Duffy B. (2010). Complete Genome Sequence of the Fire Blight Pathogen *Erwinia amylovora* CFBP 1430 and Comparison to Other Erwinia spp. *Molecular Plant-Microbe Interactions*, 23(4): 384-393.

Sobiczewski P., Mikiciński A. and Stępowska A. (2013). Necrotrophic survival of *Erwinia amylovora* in apple leaf tissue. (eds). *13th ISHS International Fire Blight Workshop, Zurich, Switzerland*, Zurich, Switzerland. 26.

Spinelli F., Noferini M. and Costa G. (2006). Near Infrared Spectroscopy (NIRs): Perspective of Fire Blight Detection in Asymptomatic Plant Material. *Acta Horticulturae*, 704: 87-90.

Steiner P. W. (1990a). Predicting apple blossom infections by *Erwinia amylovora* using the MARYBLYT model. *Acta Horticulturae*, 273: 139-148.

Steiner P. W. (1990b). Predicting canker, shoot and trauma blight phases of apple fire blight epidemics using the MARYBLYT model. *Acta Horticulturae*, 273: 149-158.

Steiner P. W. (2000). Integrated Orchard and Nursery Management for the Control of Fire Blight. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI Publishing, New York, pp. 339-358.

Suzuki K., Takamatsu K., Okuno T., Ohuchi A. and Kakazu Y. (2002). A Planning System for Precision Farming Based on an Autonomous Tractor. In: Kozan E. and Ohuchi A. (ed). *Operations Research/Management Science at Work*. Springer US, pp. 363-373. International Series in Operations Research & Management Science.

Tamura K., Stecher G., Peterson D., Filipski A. and Kumar S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30 2725-2729.

Thenkabail P. S., Smith R. B. and de Pauw E. (2000). Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sensing of Environment*, 71: 158-182.

Thomson S. V. (2000). Epidemiology of Fire Blight. In: Vanneste J. L. (ed). *Fire Blight: The Diseae and its Causative Agent, Erwinia amylovora*. Cabi Publishing, Wallingford, UK, pp. 9-36.

Timur Momol M. and Aldwinckle S. H. (2000). Genetic Diversity and Host Range of *Erwinia amylovora*. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 55-72.

Torres-Cruz J. and van der Woude M. W. (2003). Slipped-Strand Mispairing Can Function as a Phase Variation Mechanism in Escherichia coli. *Journal of Bacteriology*, 185(23): 6990-6994.

van Belkum A. (1999). The role of short sequence repeats in epidemiologic typing. *Current Opinion in Microbiology*, 2(3): 306-311.

Van der Zwet T. and Bell R. L. (1995). Response of Central European pyrus germplasm to natural fire blight infection and artificial inoculation. *HortScience*, 30(6): 1287-1291.

Van der Zwet T. and Keil H. L. (1979). Fire Blight – A Bacterial Disease of Rosaceous Plants. *Agriculture Handbook 510, US Department of Agriculture, Washington, DC*: 200.

Vanneste J. L. (2000). What is Fire Blight? Who is *Erwinia amylovora*? How to Control It? In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 1-6.

Vanneste J. L., Cornish D. A., Spinelli F. and Yu J. (2004). Colonisation of apple and pear leaves by different strains of biological control agents of fire blight. *New Zeland Plant Protection*, 57: 49-53.

Vanneste J. L. and Eden-Green S. (2000). Migration of *Erwinia amylovora* in host plant tissues. In: Vanneste J. L. (ed). *Fire blight: The disease and its causative agent, Erwinia amylovora*. CABI *Publishing*, New York, pp. 73-83.

Vanneste J. L., Lex S., Vermeulen M. and Berger F. (2002). Ioslation of *Erwinia amylovora* from blighted plum (Prunus domestica) and potato toses (Rosa rugosa). *Acta Horticulturae*, 590: 89-94.

Végh A., Némethy Z., Hajagos L. and Palkovics L. (2012). First Report of *Erwinia amylovora* Causing Fire Blight on Plum (*Prunus domestica*) in Hungary. *Plant disease*, 96(5): 759.

Vogelmann J. E., Rock B. N. and Moss D. M. (1993). Red edge spectral measurements from sugar maple leaves. *International Journal of Remote Sensing*, 14(8): 1563-1575.

Waterston J. M. (1938). Report of the plant pathologist. Report of the Department of Agricalture. Bermuda**:** 22-34.

Winslow C. E. A., Broadhurst J., Buchanan R. E., Krumwiede J. C., Rogers L. A. and Smith G. H. (1920). The families and genera of the bacteria. Final report of the Committee of the Society of American Bacteriologists on characterization and classification of bacterial types. *Journal of Bacteriology*(5): 191-229.

Yaich M., Fatmi M., Bougsiba M., Valentini F., Scuderi G., D'Onghia A. M. and Cirvilleri G. (2011). Fire blight (*Erwinia amylovora* [Burrill] Winslow) in Morocco: importance, geographical distribution and characterization. 50(2).

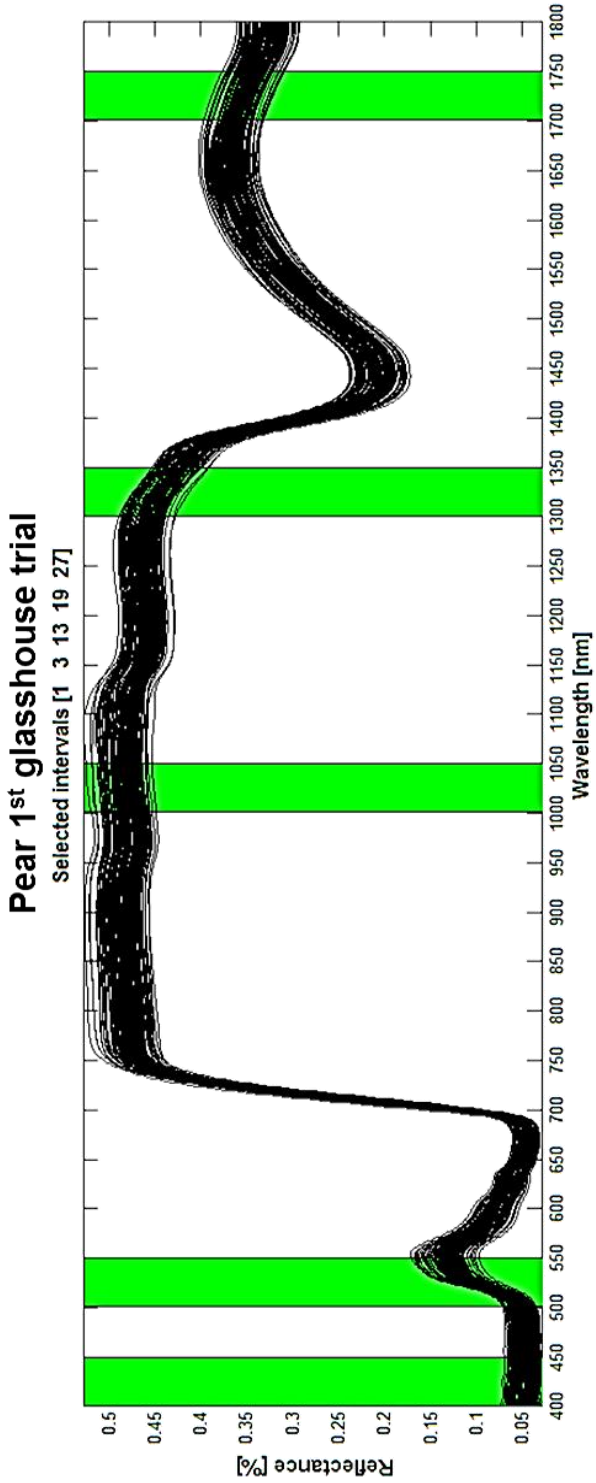
Zarco-Tejada P. J., Berjon A., Lopez-Lozano R., Miller R. J., Martin P., Cachorro V., Gonzalez M. R. and de Frutos A. (2005). Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sensing of Environment*, 99: 271-287.

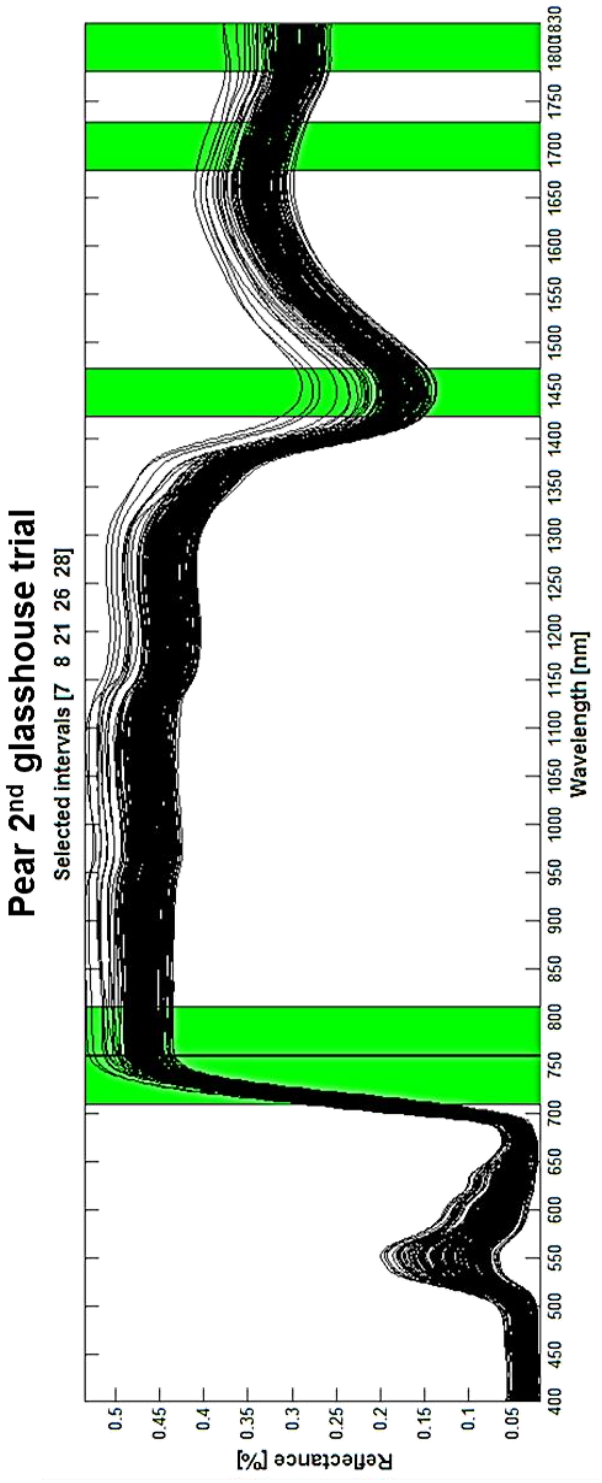
Zarco-Tejada P. J., Pushnik J. C., Dobrowski S. and Ustin S. L. (2003). Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects. *Remote Sensing of Environment*, 84(2): 283-294.

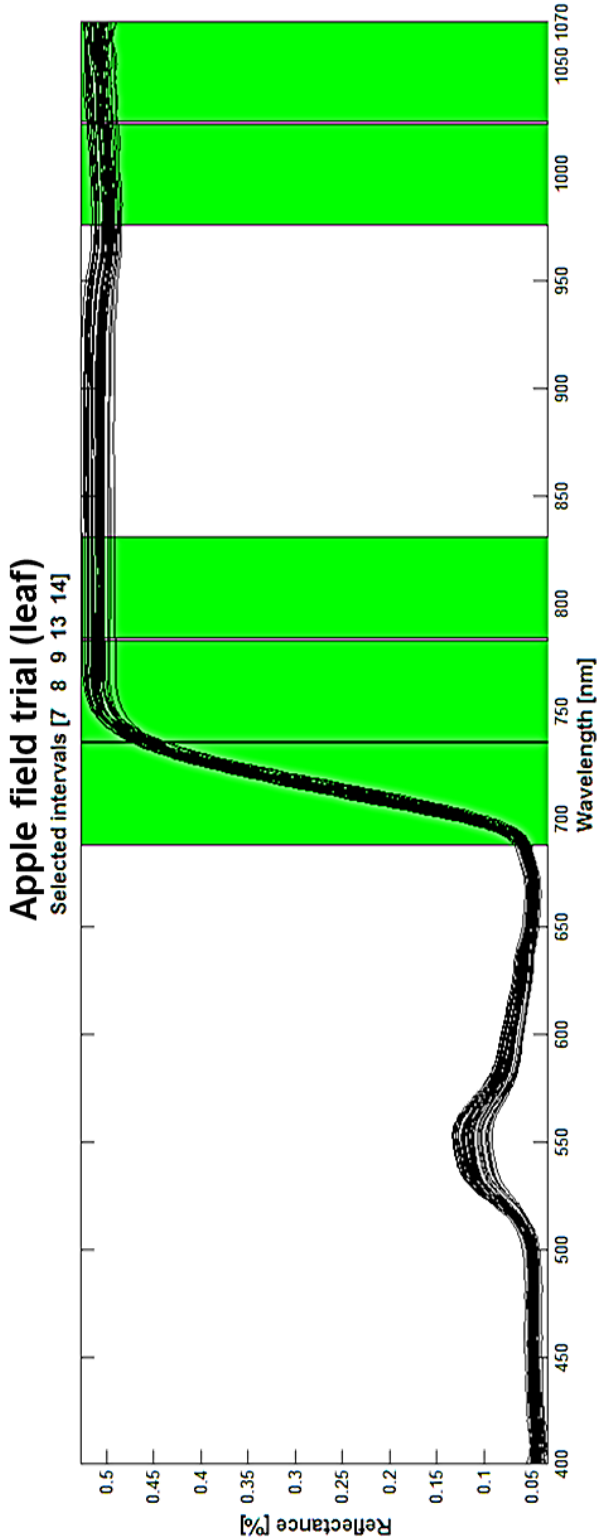
Zweig M. H. and Campbell G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4): 561-577.

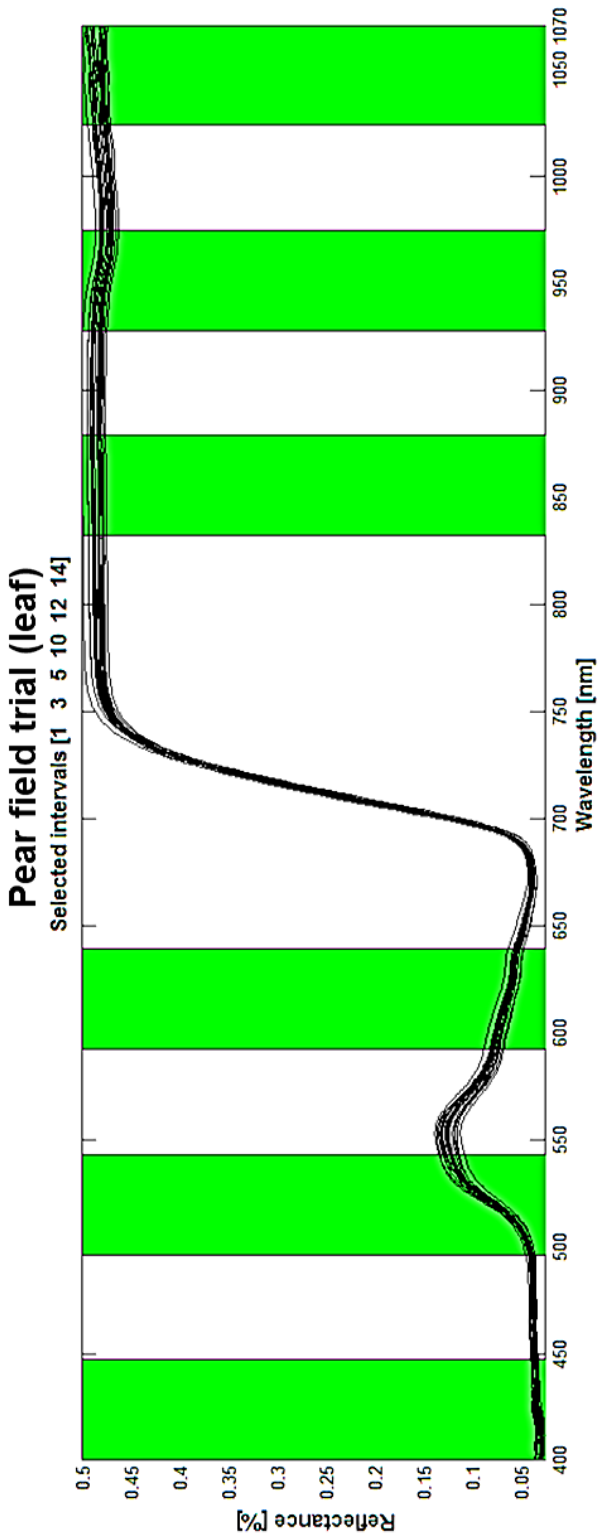
Annexes

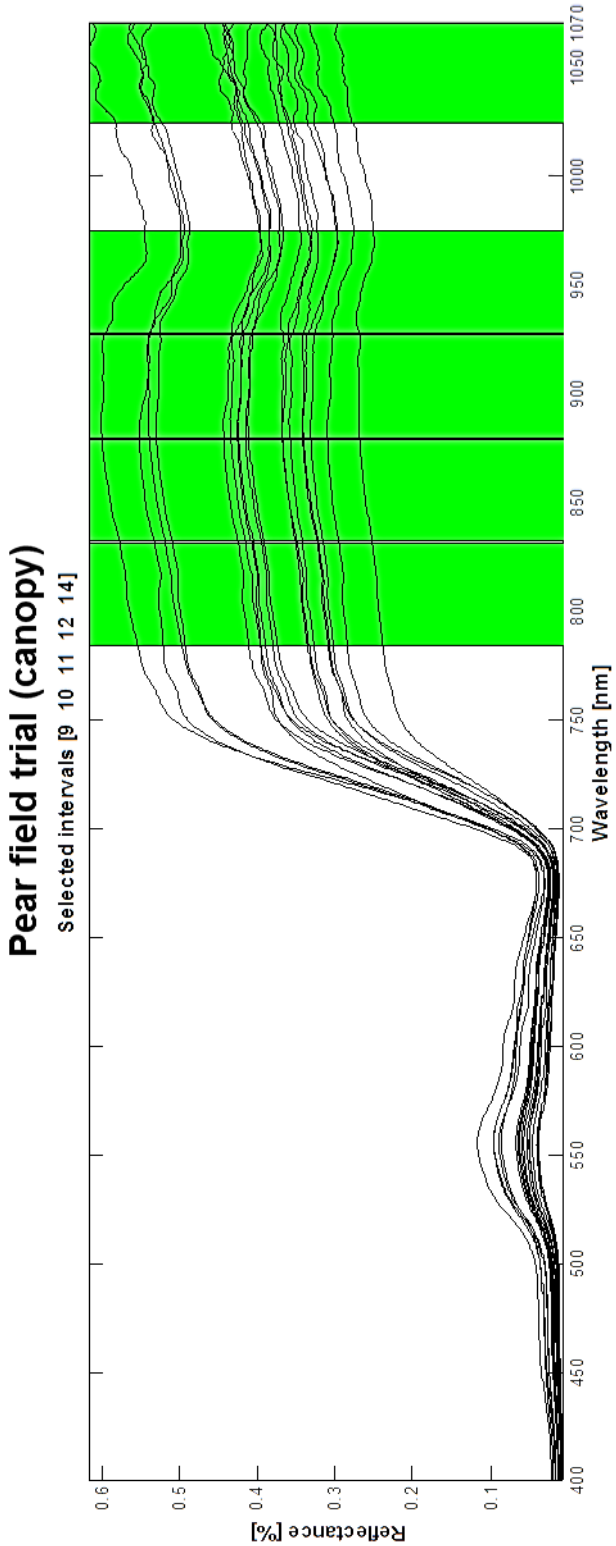
Annex I



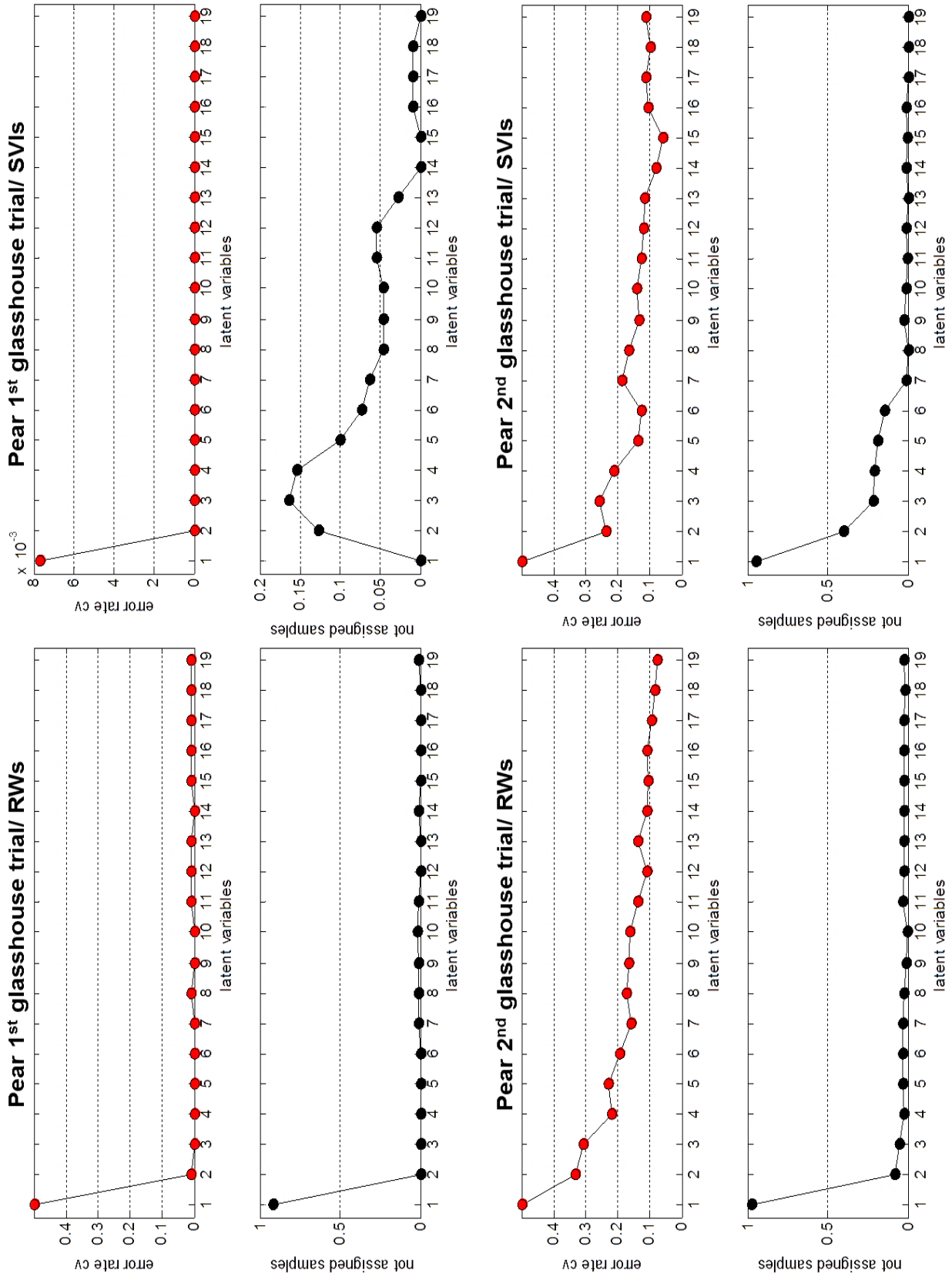


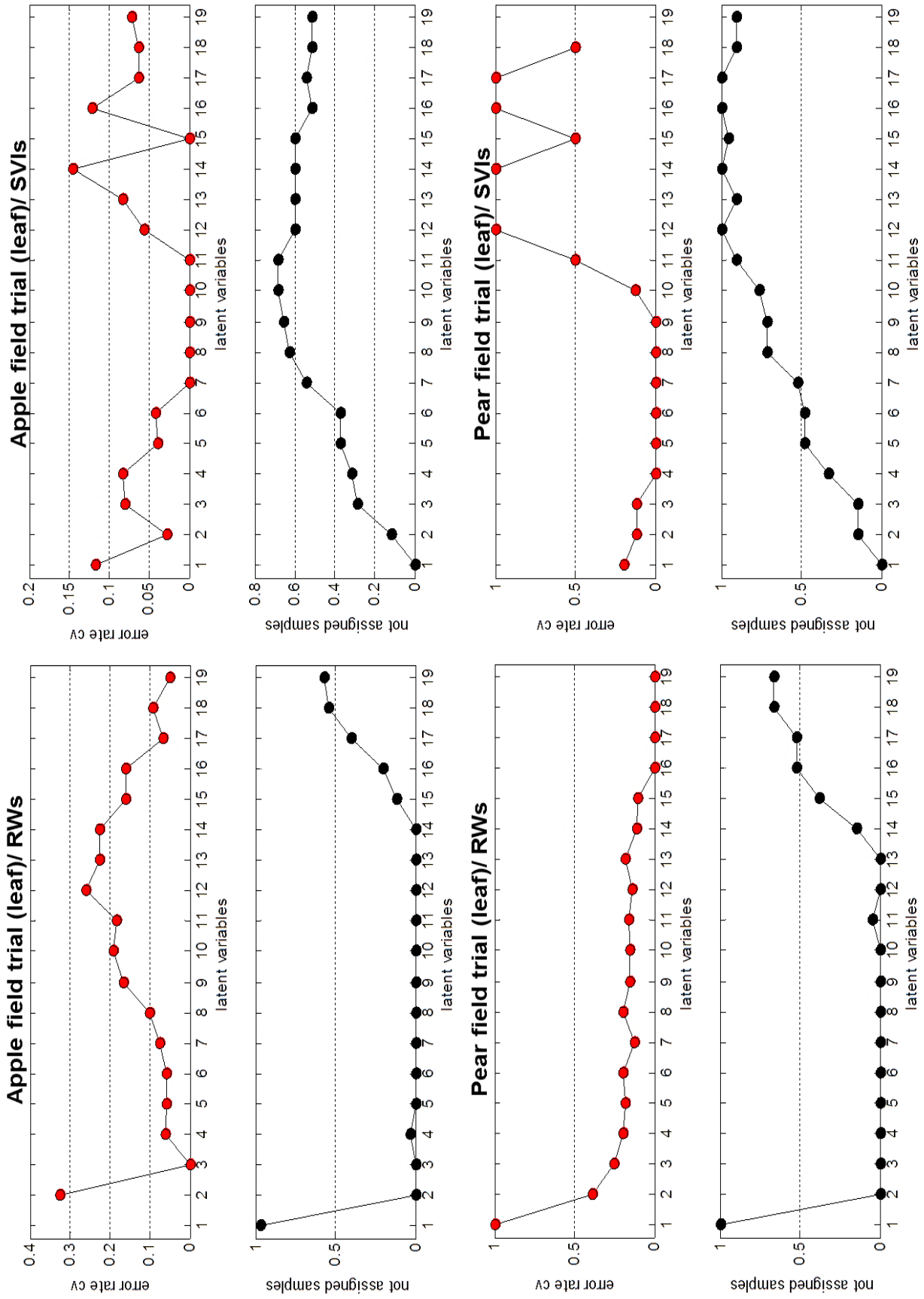

Apple field trial (leaf)
Selected intervals [7 8 9 13 14]
Reflectance [%]
Wavelength [nm]
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050
1070

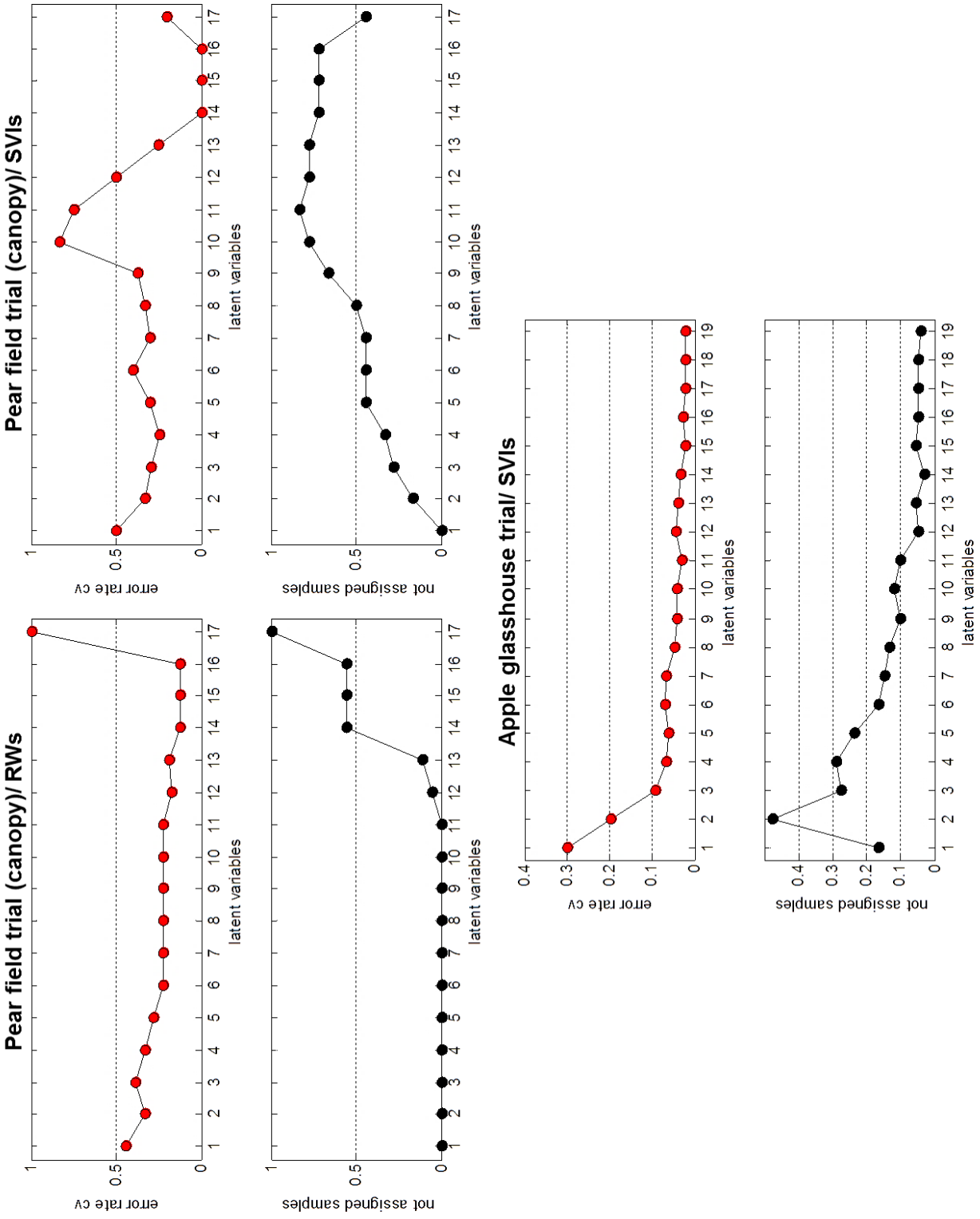

Pear field trial (leaf)
Selected intervals [1 3 5 10 12 14]
Reflectance [%]
Wavelength [nm]
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050
1070


Pear field trial (canopy)
Selected intervals [9 10 11 12 14]
Reflectance [%]
0.6
0.5
0.4
0.3
0.2
0.1
400
450
500
550
600
650
700
750
800
850
900
950
1000
1050 1070
Wavelength [nm]

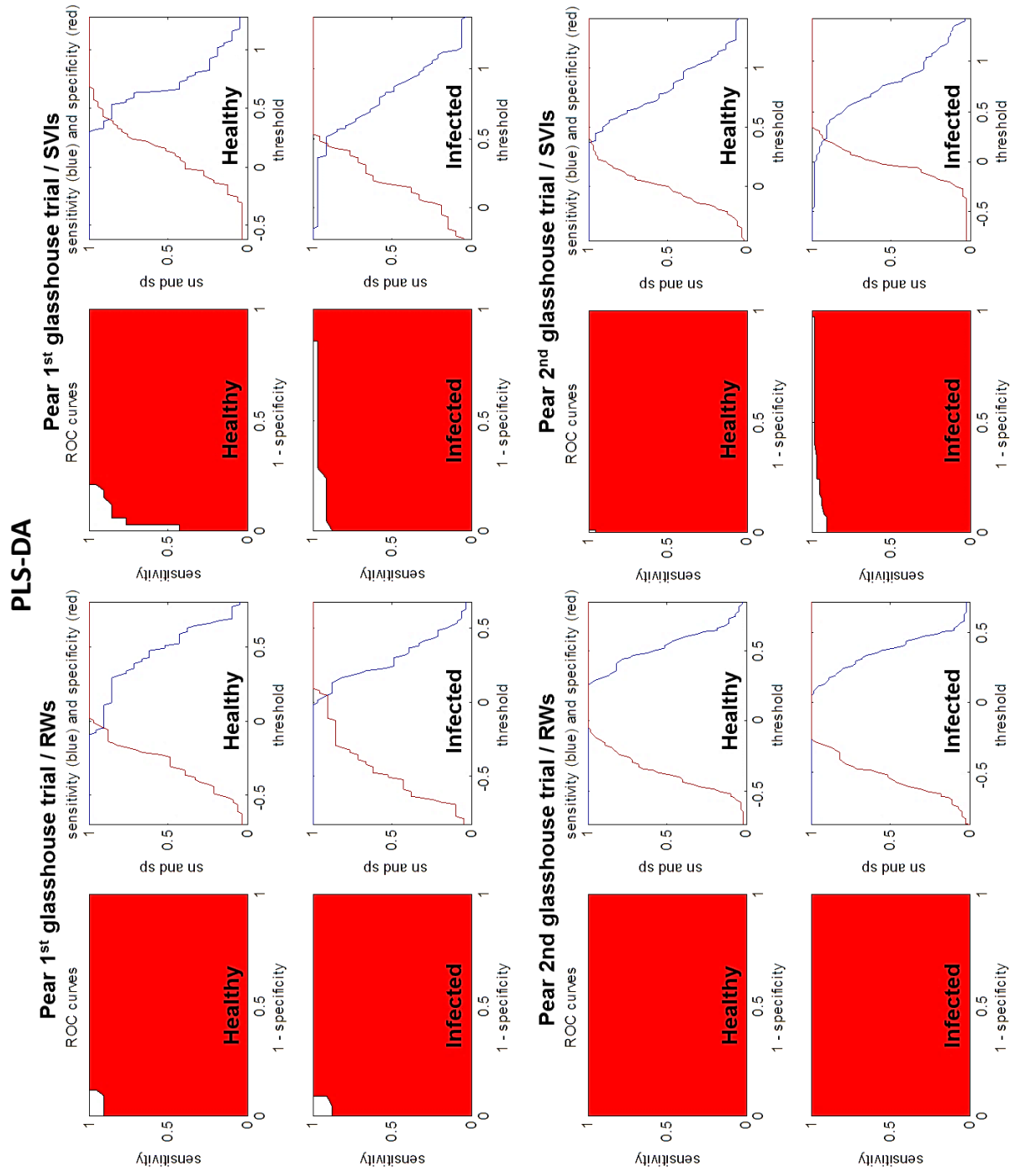
Annex II

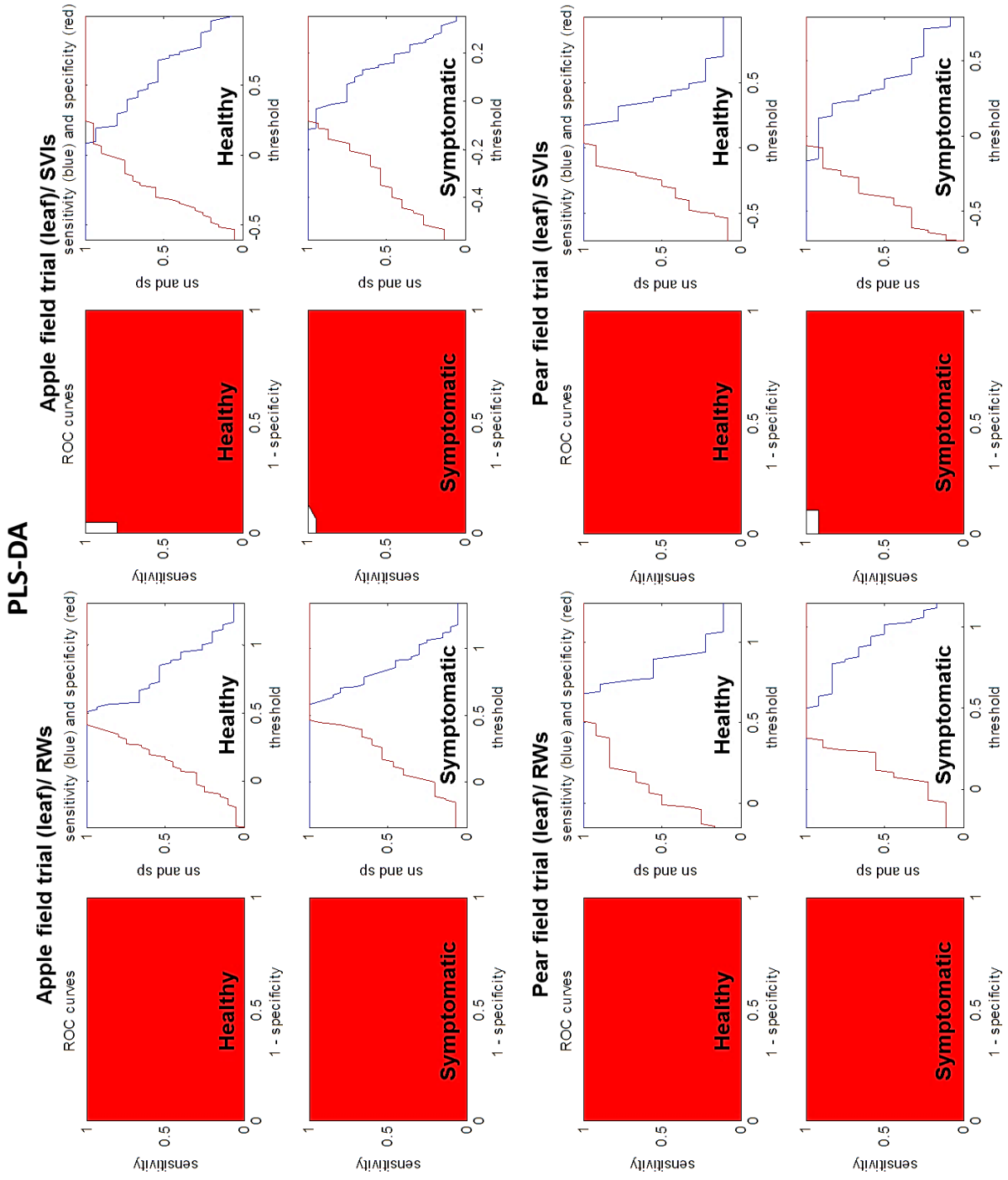

Pear 1st glasshouse trial/ SVIs
Pear 2nd glasshouse trial/ SVIs
Pear 1st glasshouse trial/ RWs
Pear 2nd glasshouse trial/ RWs
error rate cv
not assigned samples
latent variables
x 10-3

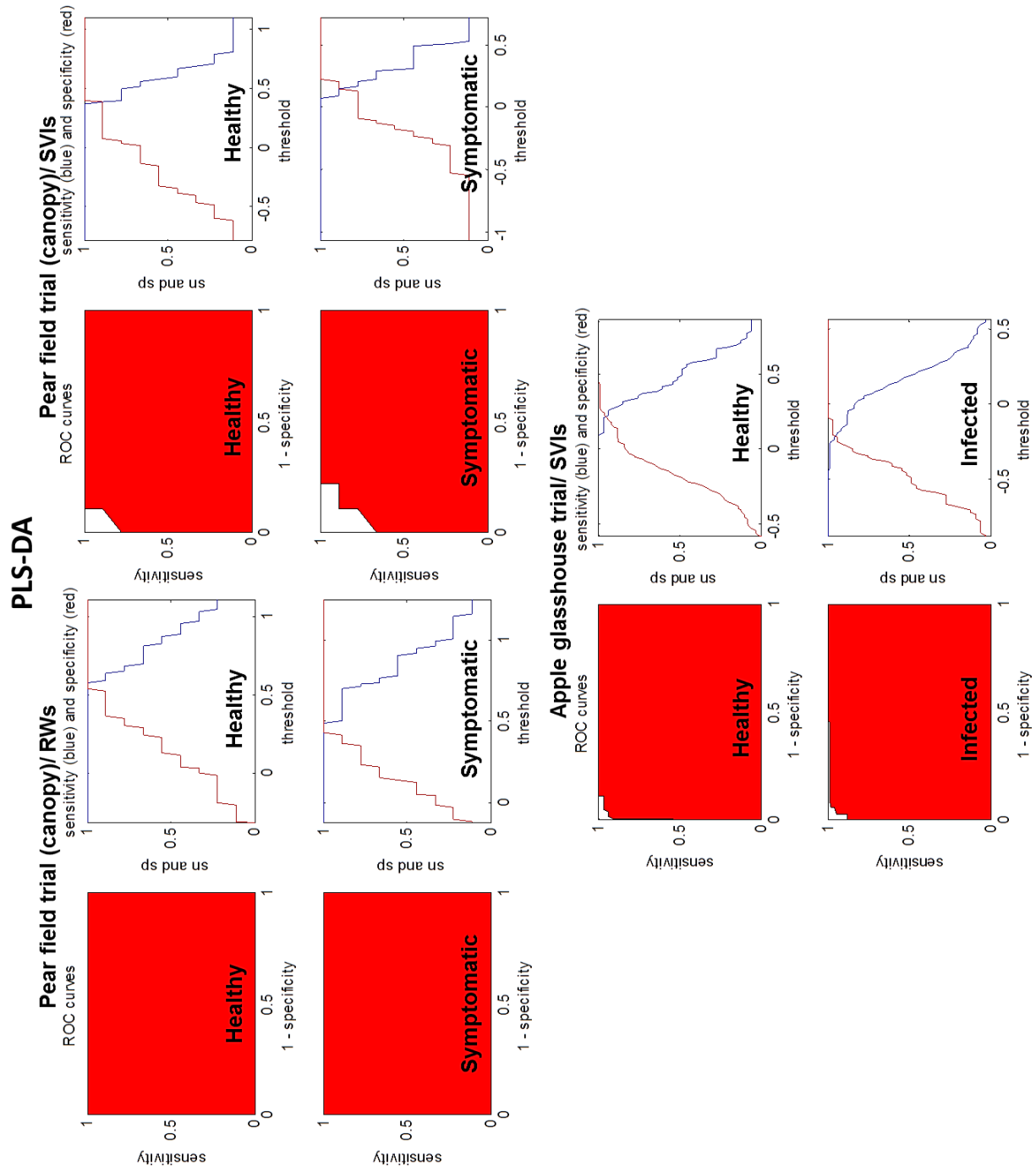

Apple field trial (leaf)/ SVIs
Pear field trial (leaf)/ SVIs
Apple field trial (leaf)/ RWs
Pear field trial (leaf)/ RWs
error rate cv
not assigned samples
latent variables


Pear field trial (canopy)/ SVIs
error rate cv
not assigned samples
latent variables
Pear field trial (canopy)/ RWs
error rate cv
not assigned samples
latent variables
Apple glasshouse trial/ SVIs
error rate cv
not assigned samples
latent variables

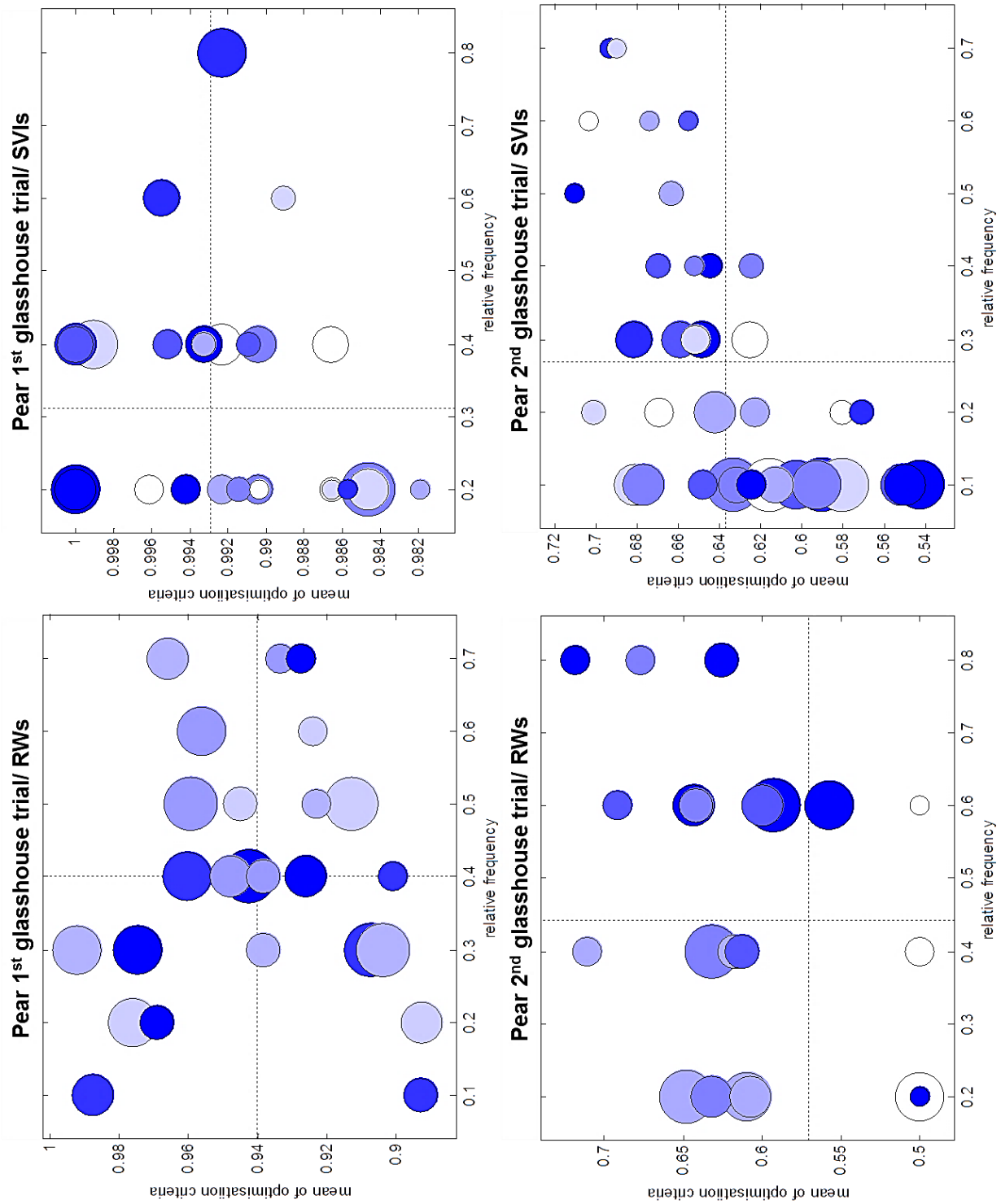
Annex III

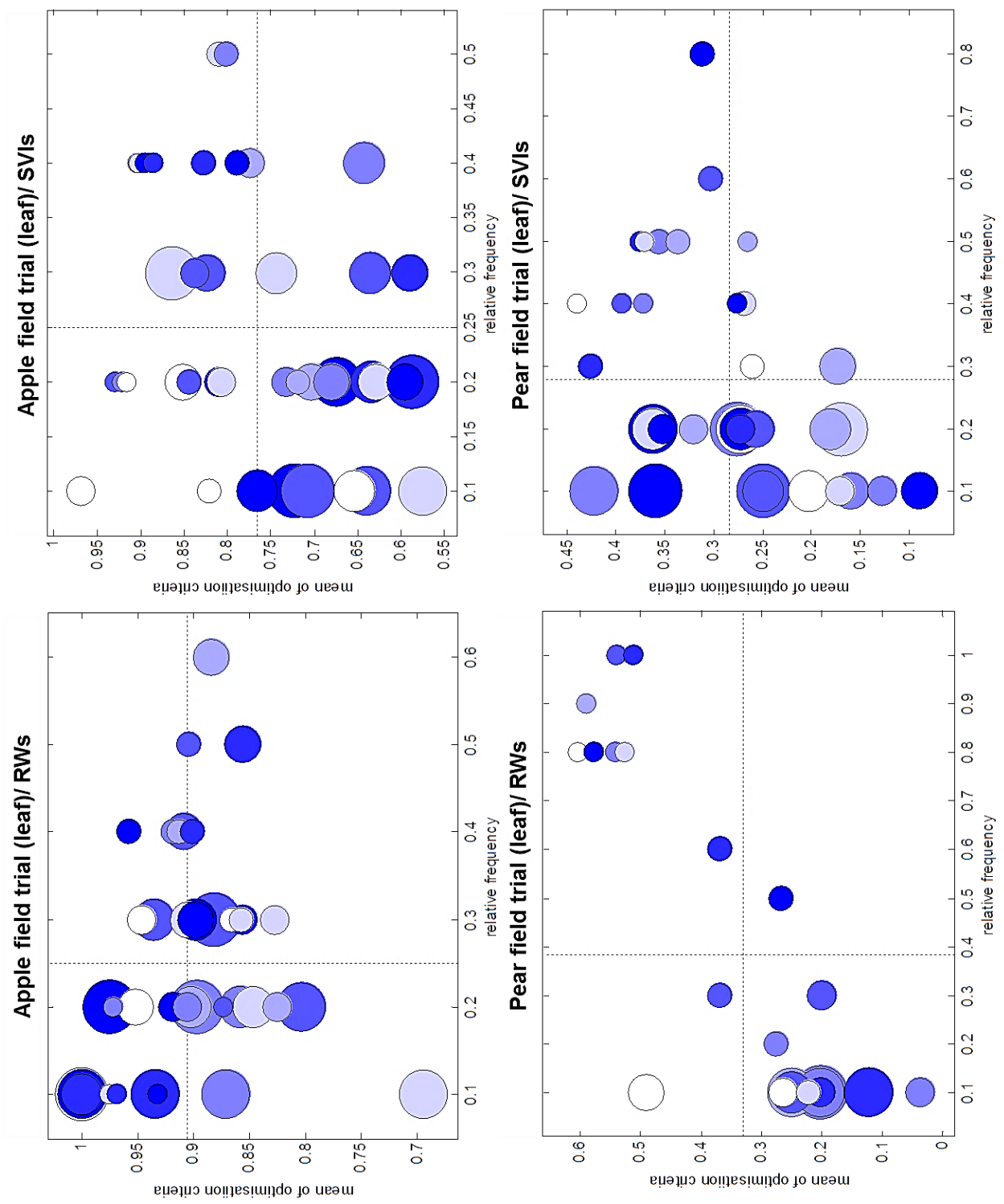

PLS-DA
Pear 1st glasshouse trial / RWs
Pear 1st glasshouse trial / SVIs
Pear 2nd glasshouse trial / RWs
Pear 2nd glasshouse trial / SVIs
ROC curves
sensitivity (blue) and specificity (red)
Healthy
Infected
sensitivity
1 - specificity
threshold
se and sp

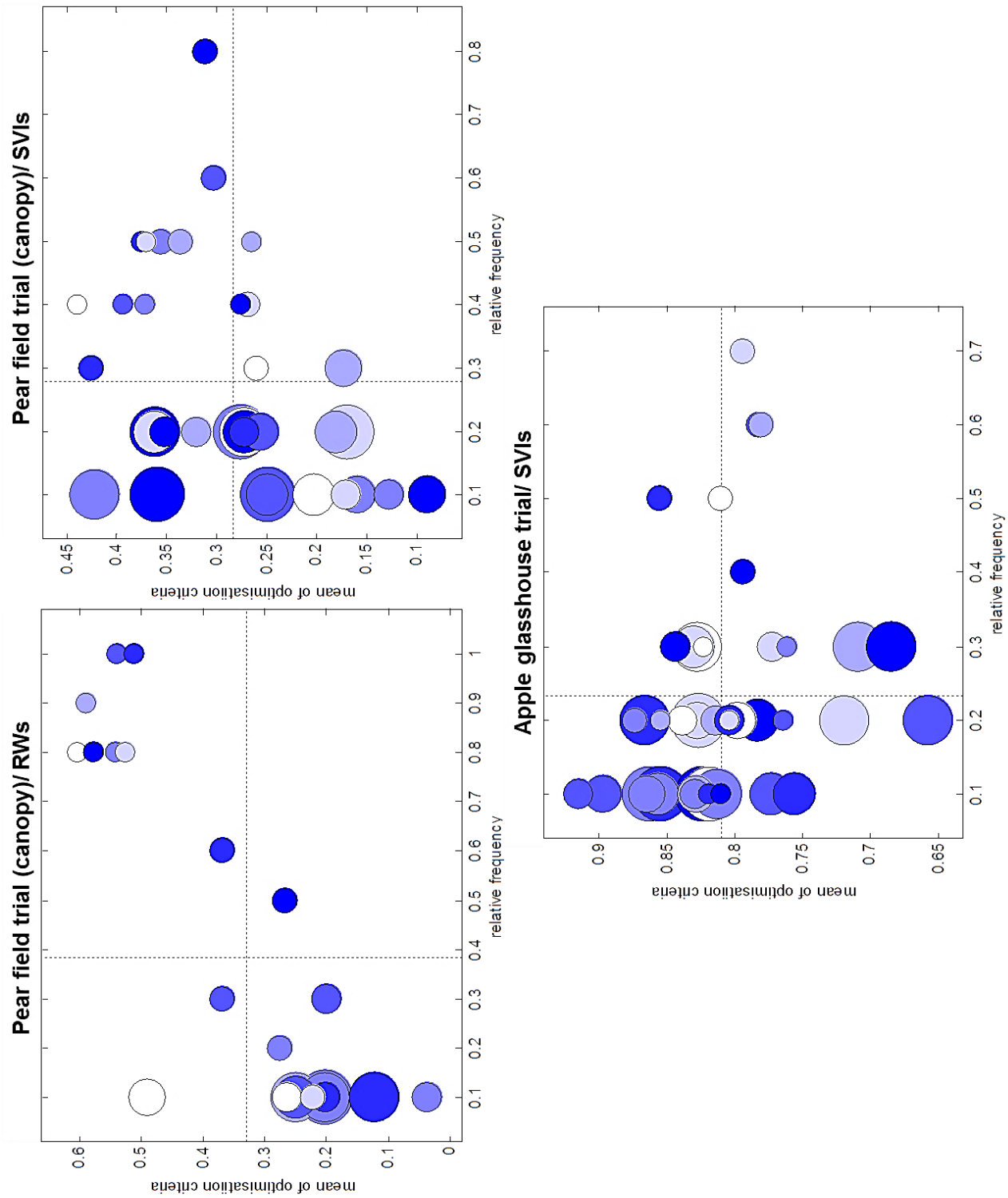

PLS-DA
Apple field trial (leaf)/ RWs
Apple field trial (leaf)/ SVIs
Pear field trial (leaf)/ RWs
Pear field trial (leaf)/ SVIs
ROC curves
sensitivity (blue) and specificity (red)
Healthy
Symptomatic
sensitivity
1 - specificity
threshold
sn and sp


PLS-DA
Pear field trial (canopy)/ RWs
ROC curves
sensitivity (blue) and specificity (red)
Pear field trial (canopy)/ SVIs
Apple glasshouse trial/ SVIs
Healthy
Symptomatic
Infected
sensitivity
1 - specificity
threshold

Annex IV


Pear 1st glasshouse trial/ SVIs
Pear 2nd glasshouse trial/ SVIs
Pear 1st glasshouse trial/ RWs
Pear 2nd glasshouse trial/ RWs
mean of optimisation criteria
relative frequency


Apple field trial (leaf)/ SVIs
Pear field trial (leaf)/ SVIs
Apple field trial (leaf)/ RWs
Pear field trial (leaf)/ RWs
mean of optimisation criteria
relative frequency


Pear field trial (canopy)/ SVIs
mean of optimisation criteria
relative frequency
Pear field trial (canopy)/ RWs
mean of optimisation criteria
relative frequency
Apple glasshouse trial/ SVIs
mean of optimisation criteria
relative frequency