applied sciences

MDPI

Article

Non-Destructive Early Detection of Drosophila Suzukii Infestation in Sweet Cherries (c.v. *Sweet Heart*) Based on Innovative Management of Spectrophotometric Multilinear Correlation Models

Giuseppe Altieri [1,*], Mahdi Rashvand Avaei [2], Attilio Matera [1], Francesco Genovese [1], Vincenzo Verrastro [3], Naouel Admane [3], Orkhan Mammadov [1], Sabina Laveglia [1] and Giovanni Carlo Di Renzo [1]

[1] DAFE, Department of Agricultural, Forestry, Food and Environmental Sciences, University of Basilicata, 85100 Potenza, Italy; attilio.matera@unibas.it (A.M.); francesco.genovese@unibas.it (F.G.); orkhan.mammadov@unibas.it (O.M.); sabina.laveglia@unibas.it (S.L.); giovanni.direnzo@unibas.it (G.C.D.R.)
[2] National Centre of Excellence for Food Engineering, Sheffield Hallam University, Howard Street, Sheffield S1 1WB, UK; m.rashvand@shu.ac.uk
[3] CIHEAM-IAMB—International Centre for Advanced Mediterranean Agronomic Studies, 70010 Bari, Italy; verrastro@iamb.it (V.V.); admane@iamb.it (N.A.)
* Correspondence: giuseppe.altieri@unibas.it

Abstract: *Drosophila suzukii* (Matsumura), also known as spotted wing drosophila (SWD), is invasive, with a preference for infesting commercially viable soft berries, particularly cherries. SWD infestations in sweet cherries are difficult to detect and remove in the field, packing houses, and processing lines, causing significant economic losses and reducing yields significantly, necessitating early detection of insect infestation in fruits during primary decaying stages. Few publications have addressed the use of non-destructive techniques for the detection of insect infestation in cherries. Based on the advantages and effectiveness of the spectrophotometric techniques, an attempt was made to use the spectrophotometry to rapidly detect postharvest SWD infestations of intact sweet cherry fruit, to employ it in sweet cherry fruit selection and grading processes. The main purpose of this study was to apply spectrophotometry as a rapid and non-destructive method in detecting and classifying healthy sweet cherry fruit versus that infested with SWD eggs. To model the data fit/prediction, principal components regression and partial least squares regression algorithms were considered. The external cross-validation set was initially set to 20% of the overall available samples and subsequently increased to 50% in the final selected optimal model. The identified procedure of management of regression algorithms allowed the selection of a very performant and robust model using the partial least squares regression algorithm: its false negative rate and false positive rate, after 500 Monte Carlo runs, were 0.004% +/− 0.003 and 0.02% +/− 0.01, respectively, and, in addition, the 50% of samples were used for the external cross-validation set.

Keywords: non-destructive methods; insect infestation; sweet cherry; spotted wing drosophila; spectrophotometry; correlation model; Monte Carlo method

check for updates

Academic Editor: Małgorzata Ziarno

Received: 24 November 2024
Revised: 17 December 2024
Accepted: 27 December 2024
Published: 29 December 2024

Citation: Altieri, G.; Avaei, M.R.; Matera, A.; Genovese, F.; Verrastro, V.; Admane, N.; Mammadov, O.; Laveglia, S.; Di Renzo, G.C. Non-Destructive Early Detection of Drosophila Suzukii Infestation in Sweet Cherries (c.v. *Sweet Heart*) Based on Innovative Management of Spectrophotometric Multilinear Correlation Models. *Appl. Sci.* **2025**, *15*, 197. https://doi.org/10.3390/app15010197

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The invasive polyphagous pest *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae), also known as spotted wing drosophila (SWD), spread from its native distribution in eastern and southeastern Asia [1] and emerged as a critical invasive insect pest of wild

and domesticated berries and stone fruits in the Americas and Europe in the late 2000s [2]. Human activities facilitated SWD colonization, particularly the movement and export of infested fruits, climatic conditions similar to the pest native range [3], and the low effectiveness of natural enemies [4].

Insect feeding damages occur without an apparent external symptom until fruits are nearly fully mature. In most cases, the serrated ovipositor of female SWD allows it to lay eggs in ripening, and ripe intact soft-skinned fruits, particularly strawberries, raspberries, blueberries, and sweet cherries; in this regard, SWD has a wide range of hosts [5,6]. Several authors have also indicated sweet cherry as a high-risk host for SWD because it bears fruit when few alternative host fruits are available [7,8].

Once the fruit is infested, the internal larval feeding causes direct damage, making the fruit unmarketable and unsuitable for consumption and processing industries. In addition, oviposition exposes fruit to secondary infection by providing access to pathogens, including fungi, yeasts, and bacteria [9,10]. Consequently, the infested fruits become soft and rot rapidly after harvest, resulting in reduced yields and significant economic losses [11–13].

The insect damages mainly occur inside the fruits without any visual symptoms during the larval stage. For this reason, it is difficult to detect and remove these fruits in packing houses and processing lines; the low-quality fruits enter the storage and cause damage to the surrounding healthy ones. Therefore, any method of early detection of the infestation is welcomed as it can improve productivity [14,15].

Countries allocate considerable resources to limit the spread and control of transboundary pests, also adapting plant health services and activities and cooperating regionally and globally for prevention, early warning, and control. As zero-tolerance regulations make the shipment unmarketable for just a few infested cherries [13], the early identification of the infestation before the shipment is essential. However, most of the current monitoring methods rely on visual inspection; these traditional sorting techniques are laborious, time-consuming, subjective, and inadequate for detecting and removing cherry fruits with hidden internal damages [16].

Consequently, non-destructive detection methods for insect infestation of agricultural commodities have emerged, including spectroscopic methods [17,18]. Commercial spectrometers usually have variable ranges: visible/shortwave near-infrared (Vis/SWNIR), near-infrared (NIR), or visible/near-infrared (Vis/NIR), using wavelengths between 350–1100 (Vis/SWNIR), 780–2500 nm (NIR), or 350–2500 nm (Vis/NIR), respectively.

Spectroscopy is widely accepted as a non-destructive technique for qualitative and quantitative analyses in the agriculture industry, especially for the assessment of the quality of fruits [19–25]. This technology is environmentally friendly, safe, and cost-effective. In addition, as spectroscopic techniques normally require no sample preparation, they can be used for on/in-line inspection [26,27] and have been investigated for non-destructive detection of internal insect infestation in fruits like sour cherries, wild blueberries, jujube, olive, and pomegranate [28]. Peshlov et al. tested three NIR instruments to classify fruit fly larvae damage in wild blueberries, achieving an infestation detection rate between 58% and 82% [29]. Xing et al. employed a genetic algorithm (GA) to select the most suitable spectral bands for detecting internal insect infestation in sour cherries [30]. Moscetti et al. addressed the problem of damage caused by insect feeding in chestnut trees, achieving very low classification error rates, with 16.81% false negatives, 0.00% false positives, and a total error of 8.41% [31]. Sudarjat et al. approached the problem of detecting insect infestation in mango fruits achieving a total explained variance of 99% using only two principal components [18]. Jamshidi et al. explored Vis/NIR spectroscopy to detect pomegranate fruits with internal infestation caused by carob moth (*Ectomyelois ceratoniae*) larvae, showing an overall classification rate of 90.6% [15]. Xing et al. used visible and

near-infrared spectroscopy (550 to 950 nm) to identify internal insect infestation in cherries, achieving an 85.0% classification accuracy [32]. Li et al. developed a Vis/NIR spectroscopy system to analyze the location of *Bactrocera dorsalis* infestation in citrus at different stages of infestation achieving 95.2%, 80.1%, and 100% accuracy for not infested, lightly infested, and severely infested citrus, respectively [33]. More recently [34], Zheng et al. developed a model to classify infestation in wild apple trees achieving a model specificity and accuracy of 96.32% and 95.94%, respectively.

However, with regard to the use of spectroscopy for the detection of internal insect infestation in fruits, the results depend on the type and stage of fruit ripening, type and age of the insect, and level of fruit damage or insect infestation [28], as well as on the spectrum acquisition conditions, such as the orientation of the light source [33].

In general, the spectrophotometric models are obtained by constructing a statistical chemometric model able to extract information from the sample's spectrum [23]. However, as spectrophotometric data are influenced by many factors producing different effects, it is difficult to assess the correlation between the spectra and the parameter of interest; therefore, the correlation needs to be quantified, from the statistical point of view, assessing the model prediction error on unknown samples [35] (i.e., samples that have not been used in the building of the statistical model).

Some algorithms widely used for this aim are: principal components regression (PCR), relying on the singular values decomposition factorization along the maximum variance axes, proposed in 1901 by Karl Pearson [36], and partial least squares regression (PLS), developed in 1975 by Herman Wold and Svante Wold [37,38]. PLS builds the model using as new variables the decomposition in spectral scores and loading matrices.

To estimate the prediction error, the cross-validation (CV) technique is used [39–41]. In general, the so-called k-fold CV allows the estimate of the prediction performance—assessed by the coefficient of determination (R2P) on unknown samples—through the measure of the coefficient of determination of CV (R2CV), being widely accepted as a statistical estimate of R2P.

However, the ability of the model to predict unknown samples should be assessed through the external cross-validation set (EXTCV), directly measuring the R2P; but, the external validation set is rarely used: it consists of taking apart a subset of the samples (expressed as a percentage of the overall samples, generally 20–30%) and using it to verify and measure the true prediction ability of the model on truly unknown samples.

Furthermore, some wavelengths negatively affect the model performance [42], impeding the achievement of a suitable predictive error. To address this problem, the method of the "coefficient of variation algorithm" (CVA) was developed. CVA allows for the identification of the interfering wavelengths bringing to spectrophotometric models of great predictive performance [43–45]. The use of k-fold CV involves k estimates of the model coefficients; CVA is related to the observation that the final model arises from the mean of these k estimates of the coefficients. Therefore, because the standard deviation of the mean for each coefficient measures its uncertainty, the coefficient with the greatest uncertainty is likely related to a wavelength that is not as useful for the prediction model but instead impairs the model's prediction.

To further evaluate and quantify the predictive statistical performance of the model, an additional step was considered in this work. The models' robustness is further tested using a Monte Carlo approach, randomizing the samples with which the algorithms are built [46,47]. The Monte Carlo method involves the running of several simulations to obtain a probabilistic estimate of an unknown parameter; modern computers have made this method very popular as, nowadays, this type of calculation is feasible in an acceptable time (from a few days to a few weeks, depending on the complexity of the problem).

2. Problem Statement

Few publications have addressed the use of non-destructive techniques for the detection of insect infestation in cherries. Hansen et al. used X-ray CT imaging and film X-ray to detect western cherry fruit fly larvae in cherries [48]. Moreover, Xing et al. identified the internal insect infestation in tart cherry using visible and NIR spectroscopy [32]. However, both used non-destructive techniques detecting only seriously infested fruits, with no difference between healthy and slightly infested fruits.

Considering the advantages and effectiveness of the spectrophotometric techniques, this work explored the development of a non-destructive spectrophotometric method for the early identification of SWD eggs in intact sweet cherries for use in selection and grading processes in packing houses.

3. Materials and Methods

The experiment was carried out using organic sweet cherries (c.v. *Sweet Heart*), harvested at the end of June from a commercial organic farm, "Tenute D'Onghia" located in Taranto province, Southern Italy.

The harvested cherries were transported directly to the CIHEAM-Bari insectarium (Valenzano, Puglia, Italy), where a simulation of natural infestation was performed. Fruits were carefully checked under a stereoscope (Nikon SMZ 745T) to exclude damaged and unhealthy cherries. The selected healthy cherries were divided into three identical batches. The first batch was added to the rearing cage to allow *D. suzukii* adults to lay eggs on them. After two days, corresponding to the ovipositional period determined by the life cycle estimation [16], infested fruits were moved from the cage and checked for the number of eggs per cherry under the stereoscope.

The infested cherry fruits batch, composed of 100 infested fruits, was divided into four groups, composed of 25 fruits, according to the number of eggs by fruit:

- 1st group: from 1 to 5 eggs per cherry fruit;
- 2nd group: from 6 to 10 eggs per cherry fruit. Each subgroup with the same number of eggs was composed of five fruits;
- 3rd group: from 11 to 30 eggs per cherry fruit. Each subgroup with the same number of eggs was composed of five fruits;
- 4th group: more than 30 eggs per cherry fruit.

The second batch was composed of 100 healthy and non-infested cherries.

The Vis/SWNIR measurement on the fruits of the two batches (infested and healthy fruits) was done on the same day, 48 h from the harvest, and all the fruits were maintained in the same conditions at room temperature and RH.

Subsequently, all the fruits were stored in a cold room for 24 h (0 °C, 90–95% RH) and their Vis/SWNIR spectrum was further collected on the next day. This permitted the spectral variations due to temperature changes to be taken into account [49].

A VIS-NIR spectrophotometer (model AvaSpec-2048-UA, nominal 200–1100 nm, 2048 pixels InGaAs CCD detector, Avantes, The Netherlands) was used to collect the spectra from the samples using an integrating sphere (IS) (model AvaSphere-80-REFL, diameter 80 mm, nominal reflection range 250–2500 nm, sample port 15 mm, Avantes, The Netherlands), to provide diffuse reflectance measurements, equipped with a 100 W tungsten-halogen source light (model ASBN-100W-L, SP Spectral Products, Putnam, CT, USA). On each fruit, the diffuse reflectance spectrum was measured, placing the fruit below the sampling port of the IS; four spectra were collected, turning each fruit around its symmetry axis by 90 degrees and subsequently averaging the four acquired spectra. For the spectra collection, the dedicated software Avasoft Basic ver. 7.3 from Avantes was

used. The data were analyzed using in-house scripts programmed using MATLAB software (MATLAB R2024b, MathWorks, Natick, MA, USA).

To model the data fit/prediction, PCR and PLS algorithms were selected, being both available from the "Statistics and Machine Learning Toolbox" of MATLAB software. Furthermore, to speed up calculations, the "Parallel Computing Toolbox" of MATLAB was used and the calculations were carried out on a Dell Precision T3610 computer (Intel Xeon, 12 cores, 3.7 GHz).

The chemometric parameter Y to be predicted was categorical, HEALTHY versus INFECTED. To these two categories, the Y values of −1 and +1 were arbitrarily assigned for healthy and infected fruit, respectively.

As the fitting model of the Y prediction is a continuous value, a threshold value of 0 has been considered to revert back to the two categories (HEALTHY versus INFECTED). When the predicted Y value is greater than or equal to the threshold then an INFECTED fruit is predicted; on the contrary, when the predicted Y value is less than the threshold, then the fruit is predicted to be a HEALTHY one.

The EXTCV was set to 20% (the choice of samples was randomized) of the overall available samples. The methods were validated by a 2-fold CV.

The statistical performance of the chemometric models was evaluated by the squared coefficient of correlation (R2CR), and the squared coefficient of correlation of CV (R2CV), using 50 Monte Carlo runs [23,45].

Additional parameters of prediction performance were taken into account: the 'limit of confident detection' (LODP@95%) (i.e., 1.96 times the root mean square error of prediction), the 'range of confident prediction error percent' (RANGEPERR%@95%) and the widespread ratio of the standard deviation of calibration data to standard error of prediction data (RPD), introduced by Williams and Norris [50], as extensively described in the literature and previous works of the authors [22,43–45,51].

3.1. First Step

The identification of features (i.e., wavelengths) statistically explaining the sample property has been successfully addressed by considering the regression coefficients coming out from the model cross-validation to which the CVA method was applied. Once a problematic wavelength is identified, this is eliminated from the calculation loop, and another run is performed (each new run uses a new randomized set of calibration samples). This loop continues until no more wavelengths are identified as problematic. After this phase, a study was performed on the recorded RPD values, rewinding back to the RPD maximum value and to the corresponding selected wavelengths subset. Finally, 50 Monte Carlo runs were performed using the subset of selected wavelengths; in this manner, the RPD was statistically assessed at its maximum measured value (both its mean and standard deviation of the mean).

After the selection of the optimal wavelengths using the CVA method, the EXTCV was used to verify the model performance of prediction. To assess this, the usual parameters were used: the false positive rate (FPR) and the false negative rate (FNR).

FPR is the ratio between wrongly predicted samples as inoculated and overall healthy samples in EXTCV.

FNR is the ratio between wrongly predicted samples as healthy and overall inoculated samples in EXTCV.

While the FPR represents the fruits that could be discarded as inoculated despite being healthy, the FNR is a very critical parameter representing a risk; indeed, FNR represents the fruits that could arrive to consumers being, unfortunately, inoculated.

The two algorithms were coded as M = PCR or PLS.

In addition, to improve the prediction capability of the models, some pre-treatments of the spectra were considered:

(a) raw spectrum or standardized spectrum (i.e., by centering around the mean value and dividing by the standard deviation), coded as OPTN = 0 or OPTN = 6, respectively;
(b) application of a 0th, 1st, or 2nd-degree derivative (coded as D = 0 (none), 1 (first derivative), or 2 (second derivative)) (i.e., using the Savitzky–Golay algorithm [52] with a width of 15 points and a polynomial order of two).

The identification and removal of outliers were not considered.

The final result was coded as M-D-OPTN, i.e., model name, derivative degree, and specific pre-treatment considered.

This represented the first step (see inner Monte Carlo loop in Figure 1).

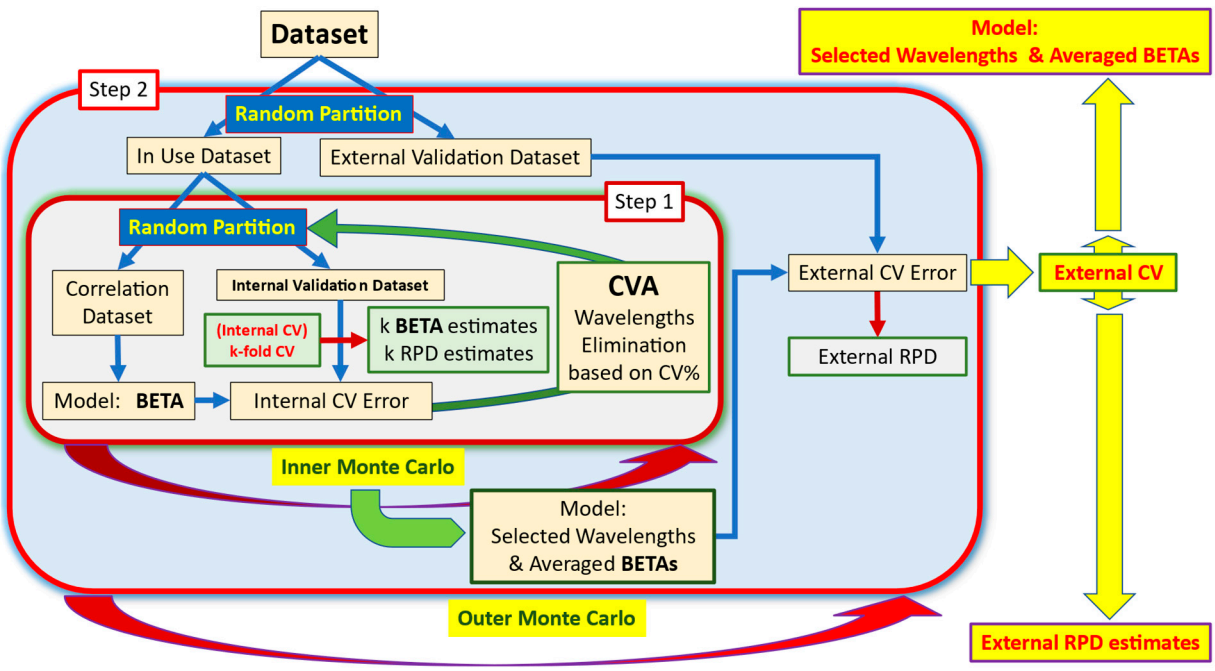

Figure 1. This figure depicts general Step1 and Step 2 layout, as previously described, of programmed script of MATLAB evaluating models.

3.2. Second Step

The second step consisted of repeating 10 times the first step (setting EXTCV to 20%), performing 10 Monte Carlo estimations of the selection algorithms on randomized samples (see outer Monte Carlo loop in Figure 1).

Therefore, for each M-D-OPTN model, the FPR and FNR parameters were assessed from the statistical point of view (i.e., the mean and standard deviation of the mean are evaluated). From these results, the optimal M-D-OPTN model was selected.

3.3. Third Step

The third step allowed evaluation of the robustness of the selected optimal model and selection of the optimal subset of wavelengths that best predicted the required property. Indeed, each of the 10 replicates, as performed in the second step, selected a different subset of wavelengths.

3.4. The Final Model

The final model was built considering the most frequently selected wavelengths. To this aim, a threshold number (S) was introduced; therefore, to each wavelength, the S value

was assigned, measuring the number of replicates of the second step (from 1 to 10) in which the specific wavelength was selected.

For example, $S \geq 5$ implies the identification of all the wavelengths that the algorithm selected in 5 to 10 of the considered 10 replicates; $S \geq 10$ implies the identification of the wavelengths that have been used in all 10 replicates.

The measure of the robustness of the selected optimal model was carried out by varying S (from 1 to 10, selecting a decreasing number of wavelengths) and changing EXTCV from 20% to 70% with increments of 5%.

The values of FNR and FPR were represented against EXTCV. An exponential fit ($y = a * \exp(b * x)$) was considered on these data and a threshold of 0.01% and 0.1% was used to express a SCORE (with regard to FNR and FPR respectively) for each subset of wavelengths (arising from the variation of S). A higher SCORE measured a better statistical performance.

For example, for each fixed S value, a subset of wavelengths was selected, so the SCORE for FNR represents the EXTCV value at which the FNR reaches 0.01% (using the exponential fit of the data), while the SCORE for FPR represents the EXTCV value at which the FPR reaches 0.1% (using the exponential fit of the data).

In conclusion, while the second step identified the optimal model to use, the third step identified the optimal subset of wavelengths to consider.

This allowed the final model coefficients and their statistical reliability to be determined, expressed through their percent coefficient of variation (CV%).

4. Results and Discussion

4.1. First and Second Step Results and Discussion

Figure 2 shows the results of the second step, consisting of repeating the first step 10 times on randomized samples and setting EXTCV to 20%, performing a Monte Carlo estimation of the selection algorithms on randomized samples. The results are expressed as FNR and FPR in percent.

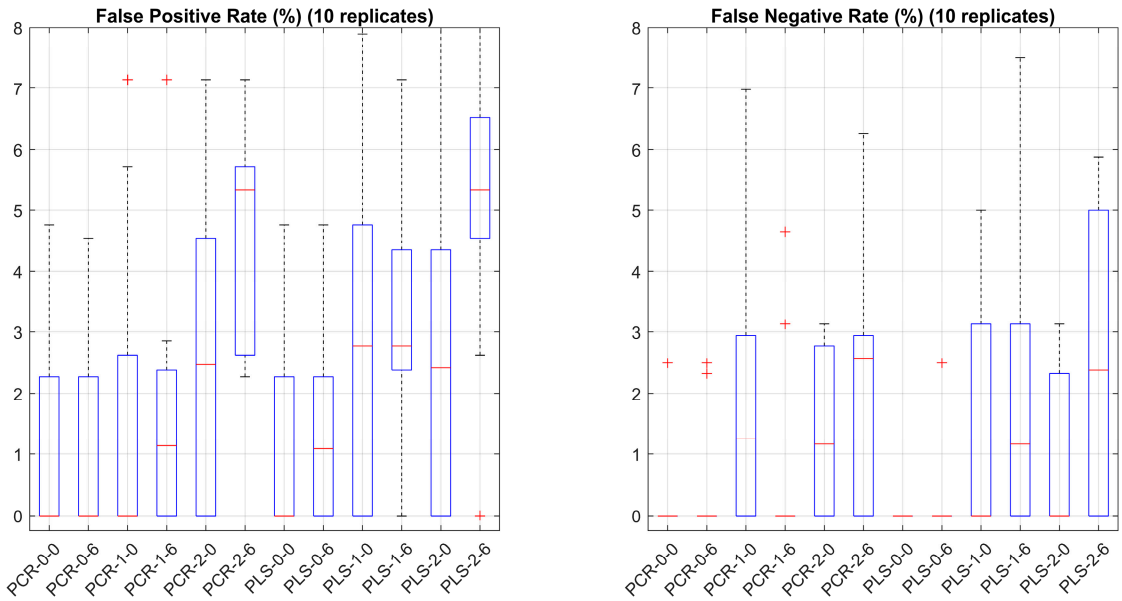

Figure 2. Boxplot of results of the second step, consisting of repeating the first step 10 times on randomized samples and setting EXTCV to 20%. Results are expressed as FNR and FPR in percent. The red "+" represent the data outliers.

Figure 3 shows the bar plot of the mean values of the FNR and FPR of Figure 2 and their standard deviation of the mean (bar). Furthermore, the PLS-0-0 model (i.e., raw spectra

without any derivative) is statistically estimated to have an FNR of 0% in all 10 randomized replicates. The red arrow in Figure 3 indicates the optimal model with regard to FNR (i.e., the consumer risk).

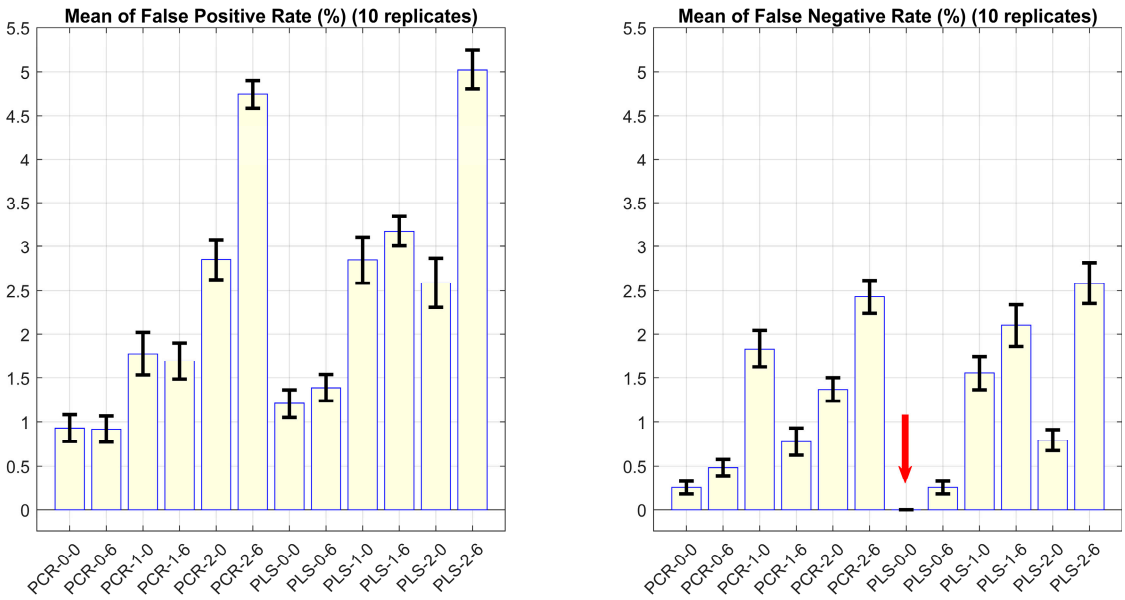

Figure 3. Bar plot, as mean and standard deviation of mean, of results of the second step, consisting of repeating the first step 10 times on randomized samples and setting EXTCV to 20%. Results are expressed as FNR and FPR in percent. Red arrow indicates optimal model with regard to FNR (i.e., consumer risk).

Figure 4 shows a replicate of the PLS-0-0 model, belonging to the first step. It is to be highlighted that the algorithm does not know the EXTCV; all the plots regarding the EXTCV are only for informative purposes. The categorical nature of the model is shown by the FPR and FNR values appearing in one of the plots in Figure 4.

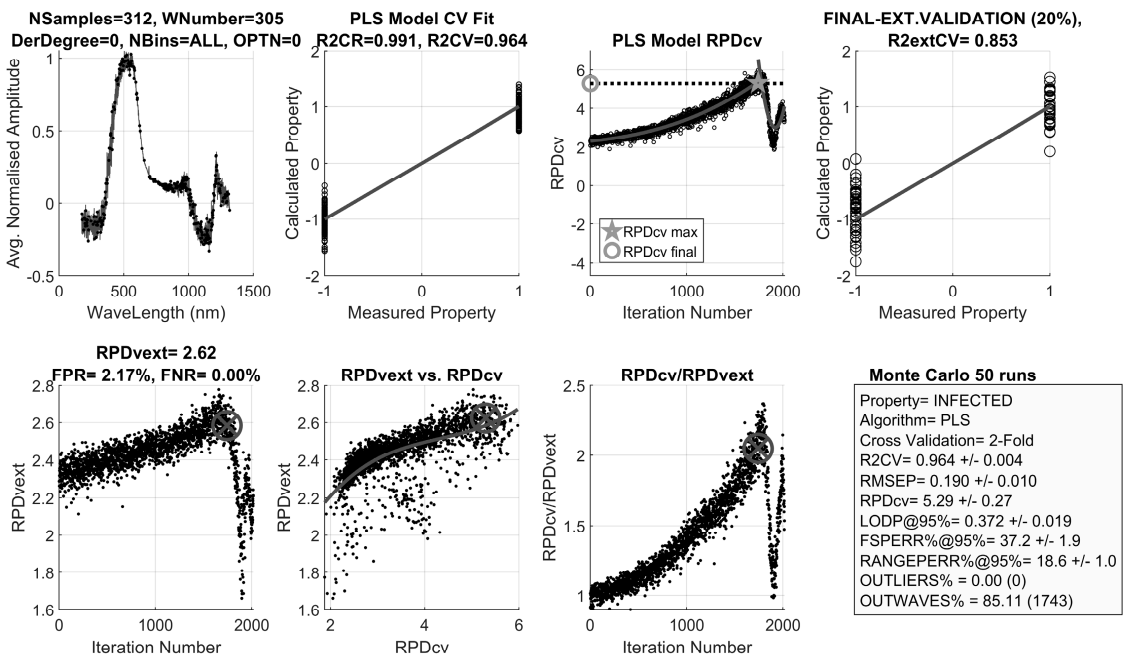

Figure 4. Replicate of PLS-0-0 model. It is to be highlighted that the algorithm does not know the EXTCV; all plots regarding the EXTCV are only for informative purposes. The categorical nature of the model is shown by the FPR and FNR values appearing in one of the plots.

The first step allowed comparison of the performance of the different M-D-OPTN models under test. To this aim, Figure 5 shows, from the same replicate previously considered, the boxplot of performances expressed as RANGEPER%@95%.

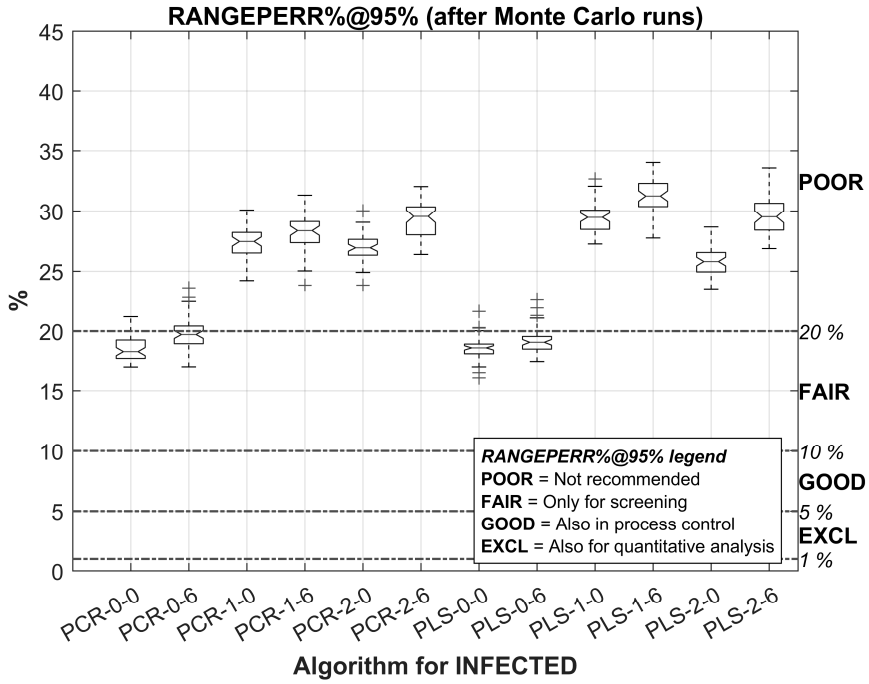

Figure 5. Boxplot graphs of a replicate of the first step, allowing the comparison of the performance of the M-D-OPTN models under evaluation. Model performance is expressed as RANGEPER%@95%. Each replicate was performed using 50 Monte Carlo runs in its final evaluation, once maximum RPD and the subset of the optimal wavelengths were identified.

Finally, the goodness of the overall underlying statistics must be assessed. Figure 6a allows to evaluate this aspect. Indeed, Figure 6a shows, for the considered replicate, the relationship between the RPD and R2CV for all the runs performed, all the considered models (PCR, PLS), and all the grouped Monte Carlo evaluations (where the standard deviation of the mean of each group is shown).

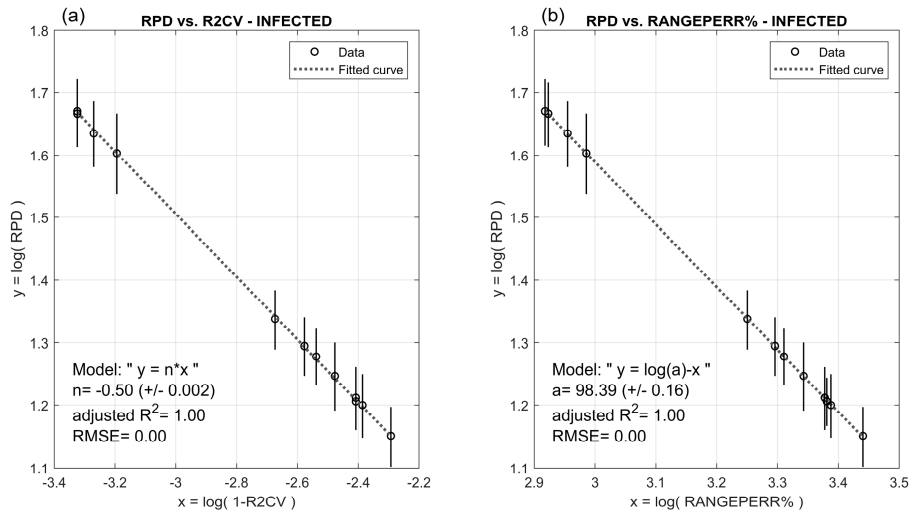

Figure 6. (**a**) Relationship between RPD and R2CV, for considered replicates, for all runs performed, for all considered models (PCR, PLS) and all grouped Monte Carlo runs. (**b**) Relationship between RANGEPERR%@95% and RPD. Correlation is hyperbolic; RPD is inversely correlated with RANGEPERR%@95%.

According to [53,54], the relation is 1/sqrt (1-R2CV); therefore, RPD and R2CV are two transposable parameters, but RPD is more discriminant than R2CV [53]. To this aim, it is normally accepted that a prediction model can be used in quality control when RPD $\geq$ 5.0 and to quantitatively measure the property of interest when RPD $\geq$ 8.1 [50,55]. However, there is no definite accord among various authors as the RPD parameter does not allow the model's statistical reliability to be evaluated in terms of the statistically guaranteed maximum measurement error on unknown samples, this last being the value needed by the final user of the model to quantify its measuring reliability.

Indeed, because any statistical model is used to forecast the property of interest on unknown samples, a prediction model must provide its error rate with a guaranteed significance level. The statistical parameter used to assess the model error rate on unknown samples is the RANGEPERR%@95%, allowing a clear interpretation of its meaning, being related to LODP%@95% and to the statistical concept of confidence interval.

Figure 6b shows the relationship existing between RANGEPERR%@95% and RPD. The correlation is hyperbolic; RPD is inversely correlated with RANGEPERR%@95%, i.e., their product is equal to a constant, and its value is 98.39 (+/−0.16).

The constant depends on the product under analysis. For example, from previous works, for kaki fruit, the mean value of the constant correlating RPD and RANGEPERR%@95% was equal to 36.39 (±3.70) [43], while for olive pomace the mean value of the constant was 45.60 (±1.78) [44]. With regard to donkey milk [45], the mean value of the constant was 45.95 (±0.52) for protein content, 70.35 (±1.32) for lactose content, and 73.84 (±0.82) for dry-matter content.

4.2. Third Step Results and Discussion

As the third step permitted the evaluation of the robustness of the selected optimal model and the identification of the optimal subset of the best wavelengths predicting the required property, the results are related to the measure of the reliability of the selected optimal model varying S (from 1 to 10, so selecting a decreasing number of wavelengths) and changing EXTCV from 20% to 70% with increments of 5%.

Figure 7 shows the FNR and FPR SCORE for the PLS-0-0 model. The optimal value for S is 5, where the higher SCORE is reached for both FNR and FPR. The S value selects the optimal wavelengths to be considered for the model, the subset being composed of 261 wavelengths.

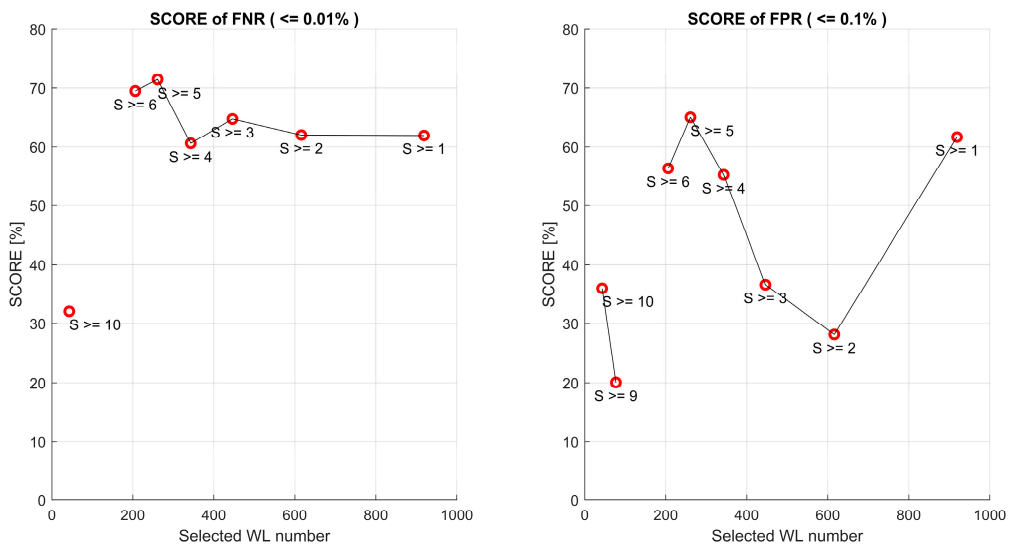

Figure 7. FNR and FPR SCORE for PLS-0-0 model. S value measures number of replicates of the second step (from 1 to 10) in which the specific wavelength was selected.

Figures 8 and 9 show the values of FNR and FPR, respectively, versus EXTCV for the optimal S $\geq$ 5. An exponential fit (y = a $*$ exp(b $*$ x)) has been considered on these data and the threshold of 0.01% and 0.1% have been used to express a SCORE (with regard to FNR and FPR respectively) for each subset of wavelengths. A higher SCORE measures a better statistical performance. The SCORE for FNR represents the EXTCV value at which the FNR reaches 0.01% (using the exponential fit of the data), while the SCORE for FPR represents the EXTCV value at which the FPR reaches 0.1%.

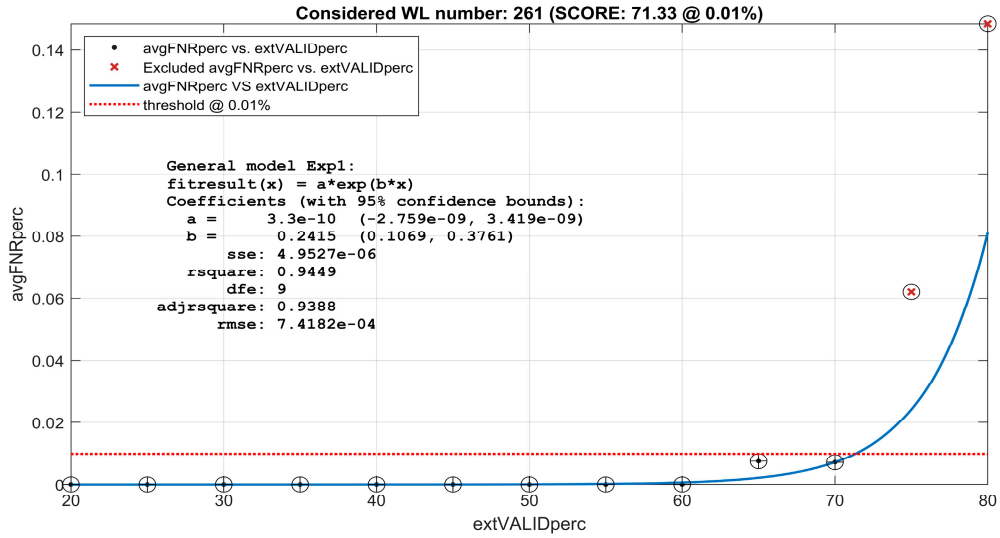

Figure 8. FNR values versus EXTCV for the optimal S $\geq$ 5. An exponential fit (y = a $*$ exp(b $*$ x)) has been considered on these data and a threshold of 0.01% has been used to express a SCORE. SCORE for FNR represents EXTCV value at which FNR reaches 0.01%.

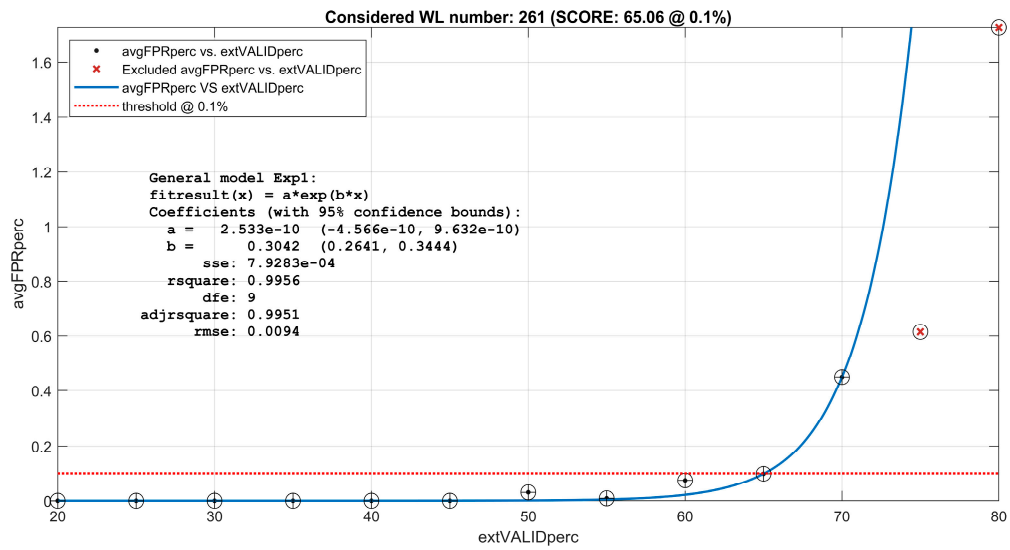

Figure 9. FPR values versus EXTCV for optimal S $\geq$ 5. An exponential fit (y = a $*$ exp(b $*$ x)) has been considered on these data and a threshold of 0.01% has been used to express a SCORE. SCORE for FPR represents EXTCV value at which FPR reaches 0.1%.

4.3. The Final Model Results and Discussion

Finally, while the second step allowed the identification of the optimal model to be used, the third step allowed the identification of the optimal subset of wavelengths to be considered.

The model coefficients (BETA or the wavelength multipliers) arise from the third step; in our case, S $\geq$ 5 is considered. The model calculation has calculated 11 estimates (the 11 values assigned to EXTCV from 20% to 70%) of each BETA coefficient. Therefore, the

model BETA coefficients are the mean of these values; in addition, the CV% is calculated to show the uncertainty of these BETA values.

Figure 10 shows the BETA coefficients value (262) and their uncertainty expressed as CV%. The average value of CV% is acceptable. However, two wavelengths have a CV% value higher than the others; this can be a symptom that the corresponding wavelength should be neglected because the value of the coefficient is near zero.

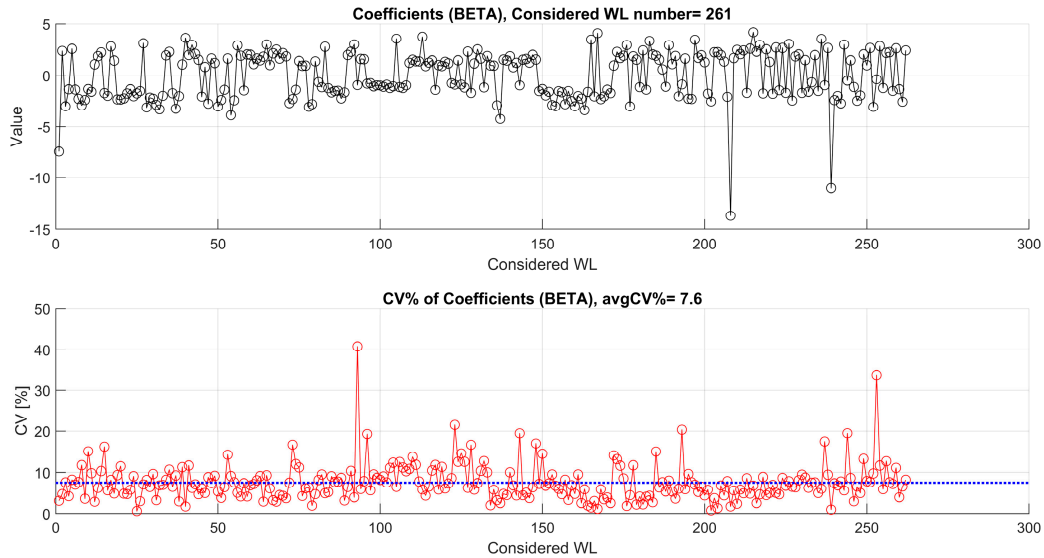

Figure 10. BETA coefficients of model arising from the third step and their uncertainty (in red). The dotted blue line represents the mean of CV%.

Because of this, the wavelengths with a CV% value greater than 15% have been further removed and the subsequent wavelengths subset was consequently obtained by purging the problematic wavelengths. Thus, the subset was reduced to 247 wavelengths.

Figure 11 shows the BETA coefficients value (247) and their uncertainty expressed as CV% for the final model. The data in Figure 11 have been obtained after 500 Monte Carlo runs assessing the FNR (0.004% +/− 0.003) and FPR (0.02% +/− 0.01) using an EXTCV of 50% of the overall randomized samples. The CV% of all the coefficients is less than 10%, showing a low uncertainty of the BETA coefficients over the overall of the Monte Carlo runs (the mean CV% is 3.1%).

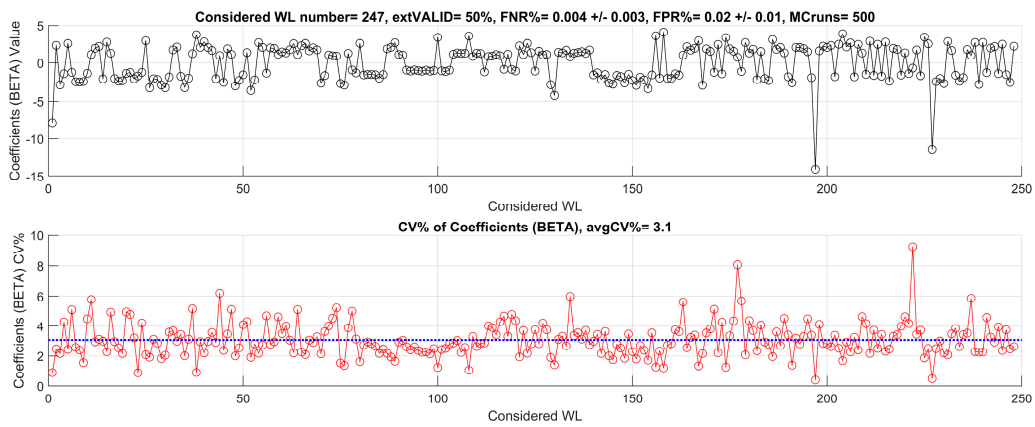

Figure 11. BETA coefficients of final model and their uncertainty; FNR and FPR, after 500 Monte Carlo runs, are 0.004% +/− 0.003 and 0.02% +/− 0.01 respectively. Considered EXTCV is set to 50% of randomized overall samples. The dotted blue line represents the mean of CV%.

Furthermore, in Figure 11, it can be visually observed that two wavelengths (1132 and 1223 nm) have a greater negative BETA coefficient. This could be related to their capacity to identify the HEALTHY cherries (HEALTHY cherries are identified by a negative Y prediction value); however, this could not be statistically demonstrated because the model considering only these two wavelengths and the constant term did not have acceptable FNR and FPR (49.8% and 44.8% respectively).

Finally, Figure 12 shows where the selected wavelengths are, as well as showing the average normalized spectrum of all HEALTHY and INFECTED cherries.

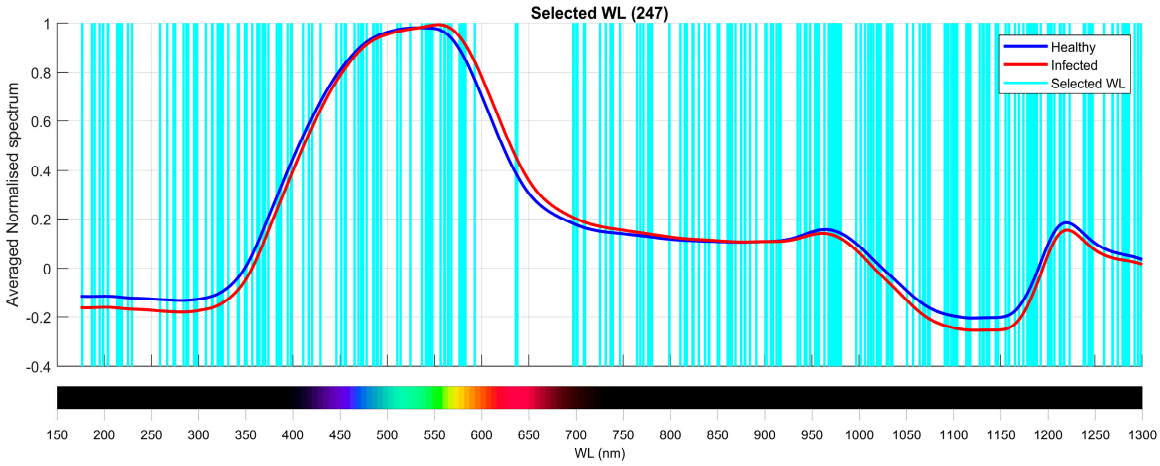

Figure 12. Final model selected wavelengths (247) and average normalized spectrum of all HEALTHY and INFECTED cherries.

There is an extended gap of not selected wavelengths between 600 and 700 nm. This corresponds to the orange-to-dark-red color range. Therefore, the orange-to-dark-red color range is uninformative with regard to the presence or absence of SWD eggs. Moreover, Figure 12 shows that the information concerning the SWD eggs identification is related to the 170–600 nm range (UV-orange) and 700–1300 nm range (shortwave near-infrared).

5. Conclusions

Due to the advantages and effectiveness of the spectrophotometric techniques, the use of the Vis/SWNIR spectral range was evaluated by developing a spectrophotometric method to rapidly detect postharvest SWD infestations of intact sweet cherry.

To model the data fit/prediction, PCR and PLS algorithms were selected.

The EXTCV was initially set to 20% of the overall available samples, then increased to 50% in the final selected optimal model.

The building of the final optimal model consisted of a three-step procedure. Each step was statistically evaluated using the Monte Carlo technique. This technique was extensively used in this work to statistically assess all the parameters and coefficients.

The identification of wavelengths best explaining the property of the samples was successfully addressed by considering the regression coefficients coming out from the model cross-validation, then applying the CVA method.

In conclusion, the following points can be highlighted.

(1) The CVA method demonstrated its usefulness in selecting the best-explaining wavelengths.
(2) The selection of the optimal M-D-OPTN model has been performed on randomized repetitions using the Monte Carlo technique.

(3) The procedure in three steps, adopted to manage the regression algorithms, allowed the selection of a very reliable prediction model; its FNR and FPR, after 500 Monte Carlo runs, were 0.004% +/− 0.003 and 0.02% +/− 0.01, respectively.
(4) A total of 50% of the samples have been used as EXTCV, and therefore the prediction model is highly performant; indeed, with regard to the FNR, the model wrongly identified only 4 cherries out of 100,000 as not infected (being however infected).
(5) The final model is robust from the point of view of the temperature variations of the samples; indeed, the model was built taking into account the spectra variations due to temperature changes with 50% of the samples at different temperatures.

The practical application of this study is related to the early on-line detection of postharvest SWD infestations of intact sweet cherry fruit, to be employed on processing lines in packing houses. To this end, because the use of the integrating sphere is difficult to apply on the processing lines, the same management procedure of the spectrophotometric models will be evaluated using a less expensive hyperspectral camera with a reduced discrete set of wavelengths.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This study was carried out within the Agritech National Research Center and received funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR)—MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4—D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions; neither the European Union nor the European Commission can be considered responsible for them.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available upon request to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Walsh, D.B.; Bolda, M.P.; Goodhue, R.E.; Dreves, A.J.; Lee, J.; Bruck, D.J.; Walton, V.M.; O'Neal, S.D.; Zalom, F.G. *Drosophila suzukii* (Diptera: Drosophilidae): Invasive Pest of Ripening Soft Fruit Expanding Its Geographic Range and Damage Potential. *J. Integr. Pest. Manag.* **2011**, *2*, G1–G7. [CrossRef]
2. Gutierrez, A.P.; Ponti, L.; Dalton, D.T. Analysis of the Invasiveness of Spotted Wing Drosophila (*Drosophila suzukii*) in North America, Europe, and the Mediterranean Basin. *Biol. Invasions* **2016**, *18*, 3647–3663. [CrossRef]
3. Wiman, N.G.; Walton, V.M.; Dalton, D.T.; Anfora, G.; Burrack, H.J.; Chiu, J.C.; Daane, K.M.; Grassi, A.; Miller, B.; Tochen, S.; et al. Integrating Temperature-Dependent Life Table Data into a Matrix Projection Model for *Drosophila suzukii* Population Estimation. *PLoS ONE* **2014**, *9*, e106909. [CrossRef] [PubMed]
4. Lee, J.C.; Wang, X.; Daane, K.M.; Hoelmer, K.A.; Isaacs, R.; Sial, A.A.; Walton, V.M. Biological Control of Spotted-Wing Drosophila (Diptera: Drosophilidae)—Current and Pending Tactics. *J. Integr. Pest Manag.* **2019**, *10*, 13. [CrossRef]
5. Lee, J.C.; Bruck, D.J.; Dreves, A.J.; Ioriatti, C.; Vogt, H.; Baufeld, P. In Focus: Spotted Wing Drosophila, *Drosophila suzukii*, across Perspectives. *Pest Manag. Sci.* **2011**, *67*, 1349–1351. [CrossRef] [PubMed]
6. Ioriatti, C.; Walton, V.; Dalton, D.; Anfora, G.; Grassi, A.; Maistri, S.; Mazzoni, V. *Drosophila suzukii* (Diptera: Drosophilidae) and Its Potential Impact to Wine Grapes During Harvest in Two Cool Climate Wine Grape Production Regions. *J. Econ. Entomol.* **2015**, *108*, 1148–1155. [CrossRef] [PubMed]
7. Beers, E.H.; Van Steenwyk, R.A.; Shearer, P.W.; Coates, W.W.; Grant, J.A. Developing *Drosophila suzukii* Management Programs for Sweet Cherry in the Western United States. *Pest Manag. Sci.* **2011**, *67*, 1386–1395. [CrossRef]
8. Kenis, M.; Tonina, L.; Eschen, R.; van der Sluis, B.; Sancassani, M.; Mori, N.; Haye, T.; Helsen, H. Non-Crop Plants Used as Hosts by *Drosophila suzukii* in Europe. *J. Pest Sci.* **2016**, *89*, 735–748. [CrossRef]
9. Cini, A.; Ioriatti, C.; Anfora, G. A Review of the Invasion of *Drosophila suzukii* in Europe and a Draft Research Agenda for Integrated Pest Management. *Bull. Insectology* **2012**, *65*, 149–160.

10. Hamby, K.A.; Hernández, A.; Boundy-Mills, K.; Zalom, F.G. Associations of Yeasts with Spotted-Wing Drosophila (*Drosophila suzukii*; Diptera: Drosophilidae) in Cherries and Raspberries. *Appl. Environ. Microbiol.* **2012**, *78*, 4869–4873. [CrossRef] [PubMed]
11. Bolda, M.P.; Goodhue, R.E.; Zalom, F.G. Spotted Wing Drosophila: Potential Economic Impact of a Newly Established Pest—Giannini Foundation of Agricultural Economics. *ARE Update* **2010**, *13*, 5–8.
12. De Ros, G.; Anfora, G.; Grassi, A.; Ioriatti, C. The Potential Economic Impact of *Drosophila suzukii* on Small Fruits Production in Trentino (Italy). *IOBC-WPRS Bull* **2023**, *91*, 317–321.
13. Mazzi, D.; Bravin, E.; Meraner, M.; Finger, R.; Kuske, S. Economic Impact of the Introduction and Establishment of *Drosophila suzukii* on Sweet Cherry Production in Switzerland. *Insects* **2017**, *8*, 18. [CrossRef]
14. Ekramirad, N.; Adedeji, A.; Alimardani, R. A Review of Non-Destructive Methods for Detection of Insect Infestation in Fruits and Vegetables. *Innov. Food Res.* **2016**, *2*, 6–12.
15. Jamshidi, B.; Mohajerani, E.; Farazmand, H.; Mahmoudi, A.; Hemmati, A. Pattern Recognition-Based Optical Technique for Non-Destructive Detection of Ectomyelois Ceratoniae Infestation in Pomegranates during Hidden Activity of the Larvae. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2019**, *206*, 552–557. [CrossRef] [PubMed]
16. Mostafa, M.; Amor, A.I.; Admane, N.; Anfora, G.; Bubici, G.; Verrastro, V.; Scarano, L.; El Moujabber, M.; Baser, N. Reduction of Post-Harvest Injuries Caused by *Drosophila suzukii* in Some Cultivars of Sweet Cherries Using a High Carbon Dioxide Level and Cold Storage. *Insects* **2021**, *12*, 1009. [CrossRef] [PubMed]
17. Saranwong, S.; Haff, R.P.; Thanapase, W.; Janhiran, A.; Kasemsumran, S.; Kawano, S. A Feasibility Study Using Simplified near Infrared Imaging to Detect Fruit Fly Larvae in Intact Fruit. *J. Near Infrared Spectrosc.* **2011**, *19*, 55–60. [CrossRef]
18. Sudarjat; Kusumiyati; Hasanuddin; Munawar, A.A. Rapid and Non-Destructive Detection of Insect Infestations on Intact Mango by Means of near Infrared Spectroscopy. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *365*, 012037. [CrossRef]
19. Cen, H.; He, Y. Theory and Application of near Infrared Reflectance Spectroscopy in Determination of Food Quality. *Trends Food Sci. Technol.* **2007**, *18*, 72–83. [CrossRef]
20. Xia, Y.; Huang, W.; Fan, S.; Li, J.; Chen, L. Effect of Spectral Measurement Orientation on Online Prediction of Soluble Solids Content of Apple Using Vis/NIR Diffuse Reflectance. *Infrared Phys. Technol.* **2019**, *97*, 467–477. [CrossRef]
21. Alander, J.T.; Bochko, V.; Martinkauppi, B.; Saranwong, S.; Mantere, T. A Review of Optical Nondestructive Visual and Near-Infrared Methods for Food Quality and Safety. *Int. J. Spectrosc.* **2013**, *2013*, 1–36. [CrossRef]
22. Matera, A.; Altieri, G.; Genovese, F.; Di Renzo, G.C. Improved Spectrophotometric Models and Methods for the Non-Destructive and Effective Foodstuff Parameters Forecasting. *Acta Hortic.* **2021**, *1311*, 395–402. [CrossRef]
23. Nicolaï, B.M.; Beullens, K.; Bobelyn, E.; Peirs, A.; Saeys, W.; Theron, K.I.; Lammertyn, J. Nondestructive Measurement of Fruit and Vegetable Quality by Means of NIR Spectroscopy: A Review. *Postharvest Biol. Technol.* **2007**, *46*, 99–118. [CrossRef]
24. Qu, J.-H.; Liu, D.; Cheng, J.-H.; Sun, D.-W.; Ma, J.; Pu, H.; Zeng, X.-A. Applications of Near-Infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. *Crit. Rev. Food Sci. Nutr.* **2015**, *55*, 1939–1954. [CrossRef] [PubMed]
25. Pu, Y.-Y.; O'Donnell, C.; Tobin, J.T.; O'Shea, N. Review of Near-Infrared Spectroscopy as a Process Analytical Technology for Real-Time Product Monitoring in Dairy Processing. *Int. Dairy J.* **2020**, *103*, 104623. [CrossRef]
26. Jamshidi, B. Non-Destructive Safety Assessment of Agricultural Products Using Vis/NIR Spectroscopy. *NIR News* **2017**, *28*, 4–8. [CrossRef]
27. Jamshidi, B.; Mohajerani, E.; Jamshidi, J.; Minaei, S.; Sharifi, A. Non-Destructive Detection of Pesticide Residues in Cucumber Using Visible/near-Infrared Spectroscopy. *Food Addit. Contam. Part A* **2015**, *32*, 857–863. [CrossRef] [PubMed]
28. Jamshidi, B. Ability of Near-Infrared Spectroscopy for Non-Destructive Detection of Internal Insect Infestation in Fruits: Meta-Analysis of Spectral Ranges and Optical Measurement Modes. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2020**, *225*, 117479. [CrossRef] [PubMed]
29. Peshlov, B.N.; Dowelt, F.E.; Drummond, F.A.; Donahue, D.W. Comparison of Three near Infrared Spectrophotometers for Infestation Detection in Wild Blueberries Using Multivariate Calibration Models. *J. Near Infrared Spectrosc.* **2009**, *17*, 203–212. [CrossRef]
30. Xing, J.; Guyer, D.; Ariana, D.; Lu, R. Determining Optimal Wavebands Using Genetic Algorithm for Detection of Internal Insect Infestation in Tart Cherry. *Sens. Instrum. Food Qual. Saf.* **2008**, *2*, 161–167. [CrossRef]
31. Moscetti, R.; Haff, R.P.; Saranwong, S.; Monarca, D.; Cecchini, M.; Massantini, R. Nondestructive Detection of Insect Infested Chestnuts Based on NIR Spectroscopy. *Postharvest Biol. Technol.* **2014**, *87*, 88–94. [CrossRef]
32. Xing, J.; Guyer, D. Comparison of Transmittance and Reflectance to Detect Insect Infestation in Montmorency Tart Cherry. *Comput. Electron. Agric.* **2008**, *64*, 194–201. [CrossRef]
33. Li, D.; Long, J.; Tang, Z.; Han, L.; Gong, Z.; Wen, L.; Peng, H.; Wen, T. Detection and Classification of Citrus Fruit Infestation by Bactrocera Dorsalis (Hendel) Using a Multi-Path Vis/NIR Spectroscopy System. *Agriculture* **2023**, *13*, 1642. [CrossRef]
34. Zheng, Y.; Zhou, Y.; Liu, P.; Zheng, Y.; Wei, Z.; Li, Z.; Xie, L. Improving Discrimination Accuracy of Pest-Infested Crabapples Using Vis/NIR Spectral Morphological Features. *J. Food Meas. Charact.* **2024**, *18*, 8755–8766. [CrossRef]
35. Cao, N. Calibration Optimization and Efficiency in Near Infrared Spectroscopy. Ph.D. Thesis, Iowa State University, Ames, IA, USA, 2013.

36. Pearson, K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572. [CrossRef]
37. Wold, S.; Sweden, S. Personal Memories of the Early PLS Development. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 83–84. [CrossRef]
38. Wold, S.; Sjöström, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [CrossRef]
39. Bates, S.; Hastie, T.; Tibshirani, R. Cross-Validation: What Does It Estimate and How Well Does It Do It? *J. Am. Stat. Assoc.* **2021**, *119*, 1434–1445. [CrossRef] [PubMed]
40. Arlot, S.; Celisse, A. A Survey of Cross-Validation Procedures for Model Selection. *Stat. Surv.* **2010**, *4*, 40–79. [CrossRef]
41. Xu, Q.S.; Liang, Y.Z. Monte Carlo Cross Validation. *Chemom. Intell. Lab. Syst.* **2001**, *56*, 1–11. [CrossRef]
42. Xiaobo, Z.; Jiewen, Z.; Povey, M.J.W.; Holmes, M.; Hanpin, M. Variables Selection Methods in Near-Infrared Spectroscopy. *Anal. Chim. Acta* **2010**, *667*, 14–32. [CrossRef]
43. Altieri, G.; Genovese, F.; Tauriello, A.; Di Renzo, G.C. Models to Improve the Non-Destructive Analysis of Persimmon Fruit Properties by VIS/NIR Spectrometry. *J. Sci. Food Agric.* **2017**, *97*, 5302–5310. [CrossRef]
44. Altieri, G.; Matera, A.; Genovese, F.; Di Renzo, G.C. Models for the Rapid Assessment of Water and Oil Content in Olive Pomace by Near-Infrared Spectrometry. *J. Sci. Food Agric.* **2020**, *100*, 3236–3245. [CrossRef]
45. Altieri, G.; Rashvand, M.; Mammadov, O.; Matera, A.; Genovese, F.; Di Renzo, G.C. Use of Wavelength Interaction Terms to Improve near Infrared Spectroscopy Models of Donkey Milk Properties. *J. Near Infrared Spectrosc.* **2022**, *30*, 219–226. [CrossRef]
46. Sobol, I.M. *A Primer for the Monte Carlo Method*, 1st ed.; CRC Press Inc.: Boca Raton, FL, USA, 1994; ISBN 084938673X.
47. Metropolis, N.; Ulam, S. The Monte Carlo Method. *J. Am. Stat. Assoc.* **1949**, *44*, 335–341. [CrossRef] [PubMed]
48. Hansen, J.D.; Schlaman, D.W.; Haff, R.P.; Yee, W.L. Potential Postharvest Use of Radiography to Detect Internal Pests in Deciduous Tree Fruits. *J. Entomol. Sci.* **2005**, *40*, 255–262. [CrossRef]
49. Sun, X.; Subedi, P.; Walsh, K.B. Achieving Robustness to Temperature Change of a NIRS-PLSR Model for Intact Mango Fruit Dry Matter Content. *Postharvest Biol. Technol.* **2020**, *162*, 111117. [CrossRef]
50. Kays, S.E.; Archibald, D.D.; Sohn, M. Prediction of Fat in Intact Cereal Food Products Using Near-Infrared Reflectance Spectroscopy. *J. Sci. Food Agric.* **2005**, *85*, 1596–1602. [CrossRef]
51. Altieri, G.; Genovese, F.; Admane, N.; Di Renzo, G.C. On-Line Measure of Donkey's Milk Properties by near Infrared Spectrometry. *LWT-Food Sci. Technol.* **2016**, *69*, 348–357. [CrossRef]
52. Savitzky, A.; Golay, M.J.E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* **1964**, *36*, 1627–1639. [CrossRef]
53. Dardenne, P. Some Considerations about NIR Spectroscopy: Closing Speech at NIR-2009. *NIR News* **2010**, *21*, 8–14. [CrossRef]
54. Minasny, B.; McBratney, A. Why You Don't Need to Use RPD. *Pedometron* **2013**, *33*, 14–15.
55. Esteve Agelet, L.; Armstrong, P.R.; Romagosa Clariana, I.; Hurburgh, C.R. Measurement of Single Soybean Seed Attributes by Near-Infrared Technologies. A Comparative Study. *J. Agric. Food Chem.* **2012**, *60*, 8314–8322. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.