



2010-01-01

# Use of Fourier Transform Infrared Spectroscopy and Chemometric Data Analysis To Evaluate Damage and Age in Mushrooms (*Agaricus bisporus*) Grown in Ireland

Aoife O'Gorman

*Technological University Dublin, aoife.ogorman@dit.ie*

Gerard Downey

*Teagasc*

Aoife Gowen

*University College Dublin, aoife.gowen@ucd.ie*

Catherine Barry-Ryan

*Dublin Institute of Technology*

Jesus Maria Frias

*Dublin Institute of Technology, Jesus.Frias@dit.ie*

Follow this and additional works at: <https://arrow.dit.ie/schfsehart>

 Part of the [Food Chemistry Commons](#)

## Recommended Citation

O'Gorman, A. (2010): Use of Fourier Transform Infrared Spectroscopy and Chemometric Data Analysis To Evaluate Damage and Age in Mushrooms (*Agaricus bisporus*) Grown in Ireland. *Journal of Agriculture and Food Chemistry*, 58 (13), pp.7770–7776.  
doi:10.1021/jf101123a

This Article is brought to you for free and open access by the School of Food Science and Environmental Health at ARROW@TU Dublin. It has been accepted for inclusion in Articles by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@dit.ie, arrow.admin@dit.ie, brian.widdis@dit.ie.



**Use of Fourier-transform infrared spectroscopy and chemometric data analysis to evaluate damage and age in mushrooms (*Agaricus bisporus*) grown in Ireland**

**Aoife O’Gorman<sup>a</sup>, Gerard Downey<sup>b</sup>, Aoife A. Gowen<sup>c</sup>, Catherine Barry-Ryan<sup>a</sup>, and Jesus M. Frias<sup>a\*</sup>**

<sup>a</sup> School of Food Science & Environmental Health, Dublin Institute of Technology, Cathal Brugha Street, Dublin 1, Ireland

<sup>b</sup> Teagasc, Ashtown Food Research Centre, Ashtown, Dublin 15, Ireland

<sup>c</sup> Biosystems Engineering, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Dublin 4, Ireland

\* Corresponding author. Tel +353 1 402 4459, Fax +353 1 402 4495  
E-mail: [jesus.frias@dit.ie](mailto:jesus.frias@dit.ie)

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

**Abstract**

The aim of this research was to investigate whether the chemical changes induced by mechanical damage and aging of mushrooms can be (a) detected in the mid-infrared absorption region and (b) identified using chemometric data analysis. Mushrooms grown under controlled conditions were bruise-damaged by vibration to simulate damage during normal transportation. Damaged and non-damaged mushrooms were stored for up to 7 days post-harvest. Principal component analysis of FTIR spectra showed evidence that physical damage had an effect on tissue structure and the aging process. Random forest classification models were used to predict damage in mushrooms producing models with error rates of 5.9 and 9.8% with specific wavenumbers identified as important variables for identifying damage, PLS models were developed producing models with low levels of misclassification. Modeling post-harvest age in mushrooms using random forests and PLS resulted with high error rates and misclassification; however, random forest models had the ability to correctly classify 82% of day zero samples, which may be a useful tool in discriminating between ‘fresh’ and old mushrooms. This study highlights the usefulness of FTIR spectroscopy coupled with chemometric data analysis in particular for evaluating damage in mushrooms and with the possibility of developing a monitoring system for damaged mushrooms using the FTIR ‘fingerprint’ region.

**Key words:** FTIR spectroscopy; chemometrics; mushrooms; aging; damage

29

30

## 31 INTRODUCTION

32

33 Mushroom cultivation is a worldwide business with the global market valued at over

34 \$45 billion in 2005 (1). In Ireland more than 60, 000 tons of button mushrooms

35 (*Agaricus bisporus*) are produced annually, making them one of the most important

36 horticultural crops grown (2). Mushrooms are one of the most perishable food

37 products with a maximum shelf-life of 3-4 days at ambient temperature (3) mainly

38 because they have no cuticle to offer protection from physical damage, microbial

39 attack or water loss (4). They may be bruised easily by physical stress during

40 harvesting, handling and transportation. This mechanical damage triggers a browning

41 process which is the major cause of loss of value in the market (5, 6). A second

42 significant factor determining mushroom quality is time elapsed between harvesting

43 and delivery to the marketplace. Post-harvest age is particularly important for any

44 mushroom exporting country (i.e. Ireland) for which access to the food markets in

45 larger, neighboring countries within Europe is vital. There is a need for a method

46 which would allow objective evaluation of mushroom quality to ensure that only high

47 quality produce reaches the retail market and that is able to produce information on

48 the metabolites in mushrooms affected by senescence and damage (7).

49         Fourier-transform infrared spectroscopy is an analytical technique that enables

50 the rapid, reagentless and high-throughput analysis of a diverse range of samples (8).

51 Its importance lies in its ability to allow rapid and simultaneous characterization of

52 different functional groups such as lipids, proteins, nucleic acids and polysaccharides

53 (9-12) in biological molecules and complex structures. FTIR spectroscopy is an

54 important tool used for quality control and process monitoring in the food industry

55 because it is less expensive, has better performance and is easier to use than other

56 methods (13). In the same way, FTIR spectroscopy has been used as a fingerprinting  
57 tool to study response of cells to various stressing situations (14-16).

58 A key to the successful operation of this technique is the availability of  
59 mathematical tools for the interrogation and mining of large spectral data sets.  
60 Principal component analysis (PCA), partial least squares (PLS) regression and  
61 random forests (RF) are chemometric tools that have been successfully used to extract  
62 information from FTIR data (17, 18).

63 The objective of this study was to investigate the damage and aging of mushrooms  
64 grown in Ireland using FTIR spectroscopy in order to (a) differentiate between  
65 damaged and undamaged mushrooms and (b) to determine mushroom post-harvest  
66 age. The ability to develop a tool that could detect physical damage before browning  
67 becomes visible would be of importance to the mushroom industry and could reduce  
68 economic losses.

## 69 MATERIALS AND METHODS

70

71 **Mushrooms.** Second flush mushrooms were grown at the Teagasc Research  
72 Centre Kinsealy (Dublin, Ireland), harvested damage-free. A set of 160 closed cap,  
73 defect-free *Agaricus bisporus* strain Sylvan A15 (Sylvan Spawn Ltd., Peterborough,  
74 United Kingdom) mushrooms (3-5 cm cap diameter) were selected for this study and  
75 immediately transported by road to the testing laboratory. Special trays were  
76 designed to hold mushrooms by the stem using a metal grid to avoid contact between  
77 (a) mushrooms and (b) between the top of mushroom caps and the tray lid during  
78 transportation. Mushrooms arrived at the laboratory premises within 1 h after  
79 harvesting and were either damaged for the specified time length or remained  
80 damage free and then stored at 4°C until required for analysis.

81 **Mushroom treatments.** Mushrooms (n=160) were harvested in the conventional

82 manner on a single occasion. On the day of harvest, a subset (n=80) was subjected to  
83 physical damage using a mechanical shaker (Gyrotory G2, New Brunswick Scientific  
84 Co. USA) set at 300 rpm (rotations per minute) for 20 minutes; these samples were  
85 labeled as damaged (D). The remaining 80 mushrooms were untreated and labeled  
86 undamaged (UD). Ten (10) damaged and 10 undamaged mushrooms were selected at  
87 random from their respective sub-sets on the day of harvesting and prepared for  
88 spectroscopic analysis (see below); these are referred to as day 0 samples. The  
89 remainder of the mushrooms (70 each of damaged and undamaged) were placed in  
90 plastic punnets (six mushrooms per punnet) and stored as separate batches at 4°C in a  
91 controlled temperature facility. On each of 7 consecutive days of such storage, a set  
92 of 10 damaged and 10 undamaged mushrooms was randomly selected, removed from  
93 storage and prepared for FTIR analysis.

94 **FTIR spectroscopy.** Sample preparation involved the manual dissection of each  
95 mushroom into its three main tissue types (cap, gills and stalk) before freezing  
96 overnight at -70°C in a cryogenic refrigerator (Polar 340V: Angelantoni Industrie  
97 spA, Massa Martana, Italy) followed by freeze-drying (Micro-modulyo, EC  
98 Apparatus Inc, New York, USA) for 24 h. Freeze-dried samples were manually  
99 ground into fine particles using a pestle and mortar. Then, 9 mg (3% w/w) of each  
100 sample was mixed with 291 mg (97% w/w) KBr (Sigma Aldrich, Dublin, Ireland).  
101 KBr pellets were prepared by exerting pressure of 100 kg/cm<sup>2</sup> (1200 psi) for  
102 approximately 2 min in a pellet press (Specac, UK). To eliminate any interference  
103 which might be caused by variation in pellet thickness different pellets were prepared  
104 from the same sample and their infrared spectra compared. These samples were  
105 identical with their average spectra used for analysis (19).

106 Spectra were collected using a Nicolet Avatar 360 FTIR E.S.P (Thermo Scientific,  
107 Waltham, MA, USA) over the frequency range 4000-400  $\text{cm}^{-1}$ . One hundred scans of  
108 each pellet were collected at 4  $\text{cm}^{-1}$  resolution at room temperature using OMNIC  
109 software (version ESP 5.1). The average of the 100 scans was used for further data  
110 analysis. FTIR spectral data were discretized resulting in spectra containing 1868  
111 individual points (discretised every 2  $\text{cm}^{-1}$ ) for chemometric analysis.

112 **Chemometric data analyses.** Multivariate models for damage and age prediction  
113 in mushrooms using both raw (i.e. unmodified) and pre-treated spectral data were  
114 developed; the pre-treatment used was standard normal variate [SNV] and was  
115 intended to reduce scatter-induced effects in the spectra (20). The frequency region  
116 studied was 2000-400  $\text{cm}^{-1}$  (fingerprint region); this spectral range encompasses  
117 absorptions from most of the chemical species present and attenuation of the dataset  
118 in this way avoids spectral regions which have low information content and may  
119 therefore interfere with effective model development.

120 Random forest modeling achieves a classification by constructing a series of  
121 decision trees (21) and takes input variables down all trees in order to optimize  
122 classification. Each tree is constructed using a different bootstrap sample from the  
123 original data, about one-third of the cases are left out of the bootstrap sample and are  
124 not used in the construction of the k-th tree. These sets of unseen samples are called  
125 out-of-bag (OOB) sets. RF makes use of these OOB sets in many ways, in particular  
126 to give an unbiased estimate of the prediction error on unseen cases (22).

127 Random forest models were built to (a) discriminate between damaged and  
128 undamaged mushrooms and (b) to predict mushroom ages. The number of trees fitted  
129 to build the random forest was 1000, the number of random wavenumbers tried at  
130 every node of the tree was set at 500 after optimization and the random forest model

131 trained was made using a stratified random sampling strategy of the sample spectra  
132 that would take the same number of samples from each of the tissues. Principal  
133 component analysis (PCA) was used to identify patterns in data in a way which  
134 emphasizes differences and similarities. It is used to indicate relationships among  
135 groups of variables in a data set and show relationships that might exist between  
136 objects (23).

137 Partial least squares (PLS) regression was applied to the spectral data sets to  
138 develop a quantitative model for prediction of the age of damaged mushrooms. A  
139 common problem in development of multivariate prediction models is selection of the  
140 optimum number of PLS loadings; often, this selection is based on an examination of  
141 the RMSECV but identification of a minimum is not always possible or unambiguous  
142 and sub-optimal models incur a significant risk of overfitting. Experience has shown  
143 that this can be a problem when parameters which are of practical relevance, such as  
144 post-harvest age or damage, but have unclear molecular basis are being modeled. In  
145 order to avoid overfitting, model cross validation was employed as follows:

- 146 1. Samples were randomly-designated from each tissue/damage status/time  
147 grouping as calibration (60%) or validation (40%) samples. The validation  
148 subset was left completely out during the optimization of model based on the  
149 calibration set.
- 150 2. The model optimization step was carried out in order to estimate the optimal  
151 dimensionality of the PLS model built on the calibration set. The method  
152 employed for this was based on the observation that an indication of  
153 overfitting is the appearance of noise in regression vectors; this takes the form  
154 of a reduction in apparent structure and the presence of sharp peaks with a  
155 high degree of directional oscillation. A simple method (24) for objectivity



156 quantifying the shape of a regression vector, combined with the root mean  
157 square error of cross-validation (RMSECV) for the calibration set was applied  
158 in this study.

159 3. The random sample designation, model development and evaluation were  
160 performed 100 times. At the end of this cycle, models were initially examined  
161 on the basis of the number of latent variables selected, the most common  
162 number was then chosen as the optimum.

163

164 Mushroom discrimination (damaged *versus* undamaged) were performed using  
165 partial least squares discrimination analysis (PLS-DA). For PLS-DA, a dummy Y-  
166 variable was assigned to each mushroom tissue sample, 1 for damaged and 0 for  
167 undamaged. PLS-DA calibration models were developed and assessed using 100  
168 randomly-populated calibration and validation sample sets.

169 Principal component analysis (PCA) and partial least squares (PLS) regression were  
170 performed using MATLAB and The Unscrambler software (v.9.7; Camo A/S, Oslo,  
171 Norway). The routine for selection of the optimum number of PLS loadings was also  
172 performed in MATLAB. Random forest modeling was performed using R 2.8.0 (25).

173

## 174 **RESULTS AND DISCUSSION**

175

176 **Spectral data.** Average raw spectra of each of the three tissue types collected  
177 from all the damaged and undamaged samples (day 0-7 in each case) are shown in  
178 Figure 1(a, b and c). A number of observations may be made on these spectra.  
179 Firstly, the major feature is a vertical offset from one average plot to another; this  
180 offset originates in light scatter effects and may be a complication in further data  
181 analysis. Average spectra of the three tissue types also bear a close resemblance to

182 each other; there is little visible difference in peak minima locations in Figure 1. In  
183 terms of minima locations, there are major bands at 1650, 1090, 1020, 935  $\text{cm}^{-1}$ ;  
184 minor minima may be seen at 1560, 1150 and 1050  $\text{cm}^{-1}$  (Figure 2). Unambiguous  
185 identification of the molecular source of features in mid-infrared spectra of biological  
186 material is difficult but the peak at 1650 may be attributed to an amide I group while  
187 at 1560  $\text{cm}^{-1}$  may be identified as resulting from amide II groups (26, 27). Both major  
188 absorbance peaks at 1090 and 1020  $\text{cm}^{-1}$  have been attributed as structures in chitin, a  
189 major structural polysaccharide in mushrooms; absorbance at 1090  $\text{cm}^{-1}$  may also  
190 arise from secondary alcohols. Smaller features at 1150 and 1050  $\text{cm}^{-1}$  have been  
191 attributed to tertiary and primary alcohol structures (28). Minima at 935, 890 and 874  
192  $\text{cm}^{-1}$  bands correspond to  $\alpha$ - or  $\beta$ - anomer  $\text{C}_1\text{-H}$  deformations. The bands at 935 and  
193 890  $\text{cm}^{-1}$  are attributed to glucan bands, while the band at 874  $\text{cm}^{-1}$  is assigned to a  
194 mannan band (29-31). An inability to attribute all spectral features is a common  
195 feature of spectroscopy of complex biological matrices but the presence of such  
196 spectral detail implies the detection of a significant quantity of information which  
197 may be usefully interrogated by multivariate mathematical methods.

198 **Principal component analysis (PCA).** Undamaged samples were studied  
199 separately on the basis of their tissue type i.e. caps, gills and stalks. Initial PCA of the  
200 mushroom caps data revealed a single sample (day 7) which lay anomalously at some  
201 significant distance from the others; this was deleted and the resulting score plot is  
202 shown in Figure 3 for PC1 vs PC2; these first two principal components accounted for  
203 97 and 2% respectively of the total variance in the spectral dataset and some sample  
204 clustering on the basis of storage time is readily apparent. As a general observation, it  
205 may be stated that the majority of the day 0 mushroom caps have a score value on  
206 PC1 greater than zero and are therefore located on the right-hand-side of Figure 3a.

207 While there are indications that samples of different storage time cluster together, the  
208 spread of these clusters is quite large and it is not possible to readily discern any trend  
209 relating plot position and storage time in the plots. There is a suggestion that the  
210 dispersion of the samples decreases as the length of storage time increases. With  
211 regard to undamaged gill tissue, observations similar to those made above in relation  
212 to undamaged caps may be made although the distribution patterns are somewhat  
213 different.

214 In the case of damaged mushroom tissues, a different pattern was found. It is  
215 clear from Figure 3c, d and e that day 0 samples clustered together but separately  
216 from those of day 1 to day 7 samples irrespective of tissue type. This strongly  
217 suggested that physical damage had a significant effect on tissue structure and the  
218 subsequent aging process. Some implications regarding the rate of change of  
219 mushroom tissue composition with aging may be garnered from the observation that  
220 separation of day 0 from all other subsequent days accounts for the most variation in  
221 the spectral collection of damaged mushroom caps, gills and stalks.

222 Examination of PC loadings may provide information on the absorbing species  
223 which are involved in separations observed on a PC scores plot; however, meaningful  
224 interpretation of loadings arising from this dataset (data not shown) was not possible.

225 **Detection of damage (random forests).** The first random forest model  
226 developed attempted to identify which wavenumbers could be used to predict damage  
227 specifically. The model tried to predict damage in mushrooms using the IR spectra, a  
228 variable indicating the tissue from which the spectra originated (cap, gill or stalk) and  
229 the age of the mushroom (in days from 0-7) as explanatory variables. This resulted in  
230 good classification between damaged and undamaged samples with an out-of-bag  
231 error rate (OOB) of 5.9%, sensitivity of 93.3% and specificity of 95%.

232 In random forests there are two measures of importance to indicate how  
233 informative a particular variable (a wavenumber in our case) is, the mean decrease in  
234 accuracy and the Gini index. The decrease in Gini index is not as reliable as the  
235 marginal decrease in accuracy (32, 33) and for that reason the latter was analysed.  
236 The variables containing the most importance for predicting damage in the model are  
237 shown in Figure 4a. The most important variable for predicting damage was the age  
238 of the mushrooms followed by the wavenumbers 1868, 1870 and 1845  $\text{cm}^{-1}$ .

239 Induced damage in mushrooms leads to an enzymic response which is followed by  
240 brown discoloration. The enzymes involved in this response, tyrosinase, or  
241 polyphenol oxidases, catalyse the oxidation of phenols, which in turn promote the  
242 formation of melanin like compounds. This reaction is found not only in damaged  
243 mushrooms, but is also part of the natural aging process, with color in mushrooms  
244 becoming darker and less firm during storage (34). The three wavenumbers identified  
245 have the ability to differentiate between the chemical changes that are induced by the  
246 mechanical damage and are independent of those that take place due solely to aging.  
247 The three wavenumbers identified above are unassigned peaks.

248 By removing the variable age from the model a second model was built which  
249 would take IR spectra of mushrooms (independently of their age) and try to predict  
250 whether there is damage or not. This random forest could be used as a classifier of  
251 mushroom damage and gave a very good prediction model with an OOB error rate of  
252 9.8%, sensitivity of 89.2% and specificity of 91.2%. Even receiving mushrooms  
253 whose storage time after harvest was unknown the model would still classify damaged  
254 and undamaged mushroom samples with a very good classification rate. The  
255 variables of importance involved in this classification model are shown in Figure 4b.

256 The most important variable for predicting damage according to the mean decrease  
257 accuracy plot is tissue used in the analysis followed by the wavenumbers 1868, 1870  
258 and  $1560\text{ cm}^{-1}$ . The peak at  $1560\text{ cm}^{-1}$  is attributed to amide II vibrations of proteins  
259 (29). Amide II bands along with amide I bands are major regions of the protein  
260 infrared spectrum. Amide II bands are associated with an out-of-phase combination  
261 of in-plane C-N stretching and N-H bending of amide groups (35). Absorption of this  
262 band was found to be higher in damaged samples and therefore an important variable  
263 for detecting damage in mushroom samples. The wavenumbers 1868 and  $1870\text{ cm}^{-1}$   
264 are unassigned.

265 **Detection of damage (PLS).** PLS-DA models were developed to discriminate  
266 between undamaged and damaged mushrooms of all tissue types separately. A  
267 summary of the average and dispersion of the results obtained on a percentage basis  
268 for each tissue is shown in Table 2; it is apparent that misclassification errors  
269 associated with all models were low, especially so in the case of gills and stalks. In  
270 terms of numbers of samples misclassified, these percentages translate to 1 or 2 only  
271 in each case. These results indicate that FTIR of freeze-dried mushroom tissues  
272 (especially gills and stalks) may be used to discriminate between damaged and  
273 undamaged mushrooms aged post-harvest from 0 to 7 days with almost complete  
274 confidence.

275 Modeling damage in mushrooms has been reported in literature in 2008 by Gowen  
276 and colleagues and in 2009 by Esquerre et al. (36, 37). Gowen and colleagues  
277 investigated the use of hyperspectral imaging and principal components analysis  
278 (PCA) to develop models to predict damage on mushroom caps with correct  
279 classification ranging from 79-100%. Using near infra-red spectroscopy and partial  
280 least squares (PLS) regression, Esquerre and colleagues were able to correctly classify

281 undamaged mushrooms from damaged ones with an overall correct classification  
282 model with 99% accuracy. The models for predicting damage using FTIR and  
283 random forests correctly classified 94 and 90% of samples respectively, whilst the  
284 PLS predictive models correctly classified 92-99% of undamaged samples from  
285 damaged ones. These results highlight the usefulness of FTIR and chemometrics for  
286 detecting physical damage in mushrooms with the possibility of developing a  
287 classification system for the industry.

288 **Predicting post-harvest age (random forests).** Initial random forest models were  
289 built to try and predict the mushroom age from day zero to day seven (0-7) using the  
290 IR spectra from the tissues and knowing whether they had been subjected to damage  
291 or not with the aim to identify specific wavenumbers associated with aging. The  
292 random forest model produced an OOB error rate of 32% i.e. 68% of samples were  
293 correctly classified. The results of the model fit are shown in Table 3.  
294 Misclassification of samples was seen for all mushroom ages particularly days 4, 5  
295 and 7. Classification of day zero samples performed quite well in the model with  
296 82% of samples correctly classified, which leads to the possibility of using IR  
297 spectroscopy as a tool to discriminate fresh mushrooms (D0) from mushrooms that  
298 have been subjected to refrigeration. This type of tool could enable packers and  
299 producers to avoid fraud and ‘recycling’ of product, supporting the evidence from  
300 visual inspection. The variables of importance identified by the mean decrease  
301 accuracy plot were damage, tissue type and the wavenumbers 399, 952 and 1508  $\text{cm}^{-1}$ .

302 A second model was developed to predict age using the same approach as above but  
303 removing the damage variable from the model. The model performed much the same  
304 as above with an OOB error rate of 33%; again misclassification within all sample  
305 ages was seen. The model correctly classified 79% of day zero models. The

306 important variables identified to predict age were tissue type and the wavenumbers  
307 399, 952 and 1508  $\text{cm}^{-1}$ . The peak at 952  $\text{cm}^{-1}$  is a glucan band ( $\beta$ -anomer C-H  
308 deformation) (29), glucans play many different roles in the physiology of fungi, some  
309 accumulate in the cytoplasm as storage, however most are present in the cell wall  
310 structure (38). This suggests that the ability to model aging in mushrooms may  
311 depend on the affect of glucan levels changing in the cell wall due to natural  
312 senescence. The wavenumbers at 399 and 1508  $\text{cm}^{-1}$  are unassigned. The OOB  
313 errors produced to model aging were quite large >33% which may be due to the low  
314 sample numbers.

315 **Predicting post-harvest age (PLS).** PLS regression was applied separately to the  
316 caps, gills and stalks datasets in an attempt to develop separate quantitative models for  
317 prediction of the age of mushrooms, both damaged and undamaged. Selection of the  
318 appropriate number of latent variables for each model was assessed on the basis of the  
319 frequency of their occurrence. As shown in Figure 5, this was a clear and  
320 unambiguous choice. A summary of the results obtained using mushrooms from day  
321 0 to day 7 inclusive is shown in Table 3. In the case of undamaged mushrooms, root  
322 mean squared error of cross validation (RMSECV) values achieved were relatively  
323 high, only permitting the prediction of post-harvest age of damaged mushrooms to  
324 within  $\pm 2$  to 3 days approximately (95% confidence limit) depending on tissue type.  
325 The practical utility of such accuracy levels may be gauged by examination of the  
326 SD/RMSECV ratio, all but one of which are below 3.0, the generally accepted  
327 minimum value for a model to be of practical utility. With regard to damaged  
328 mushrooms, model predictive accuracies were similar for caps and stalks with  
329 RMSECV (and RER) values of 1.3 (1.9) and 1.2 (2.) respectively. In the case of gill  
330 tissue, better predictive accuracy was achieved with RMSECV and RER values equal

331 to 0.8 and 3.1 respectively. The number of latent variables associated with these  
332 models was low and similar in all cases, with a variation between 6 and 8 only. The  
333 application of an objective indicator of the optimum number of PLS loadings to  
334 include in any model contributed to their stable performance.

335 The results presented for modeling age in mushrooms using FTIR and  
336 chemometrics had misclassification errors of over 30% (random forests) yielding  
337 relatively unsuccessful results. However, random forest models were able to classify  
338 day zero samples reasonably well with correct classifications of 82 and 79% which  
339 leads to the possibility of using IR spectroscopy in detecting fresh mushrooms from  
340 old mushrooms and could be used within the sector for detecting fraud and ‘recycling’  
341 of product. The time required for freeze-dried sample preparation is in the order of  
342 hours, thus this approach would be applicable for research and quality control  
343 purposes.

344

#### 345 **Acknowledgments:**

346

347 The authors acknowledge financial support from the Irish Department of Agriculture  
348 and Food under the Food Institutional Research Measure (FIRM), supported through  
349 EU and national funds. Thanks are due to Dr. Helen Grogan and Ted Cormican,  
350 Teagasc, Kinsealy Research Centre, Dublin, Ireland for the supply of mushrooms and  
351 background information.

352

353

354

355

356



357 **LITERATURE CITED**

- 358 1. Chang, S. T., The world mushroom industry: Trends and technological  
359 development. *International Journal of Medicinal Mushrooms* **2006**, 8, 297-317.
- 360 2. Teagasc, The Irish Agriculture and Food Development Authority. *Teagasc*  
361 *Mushroom Newsletter* **2007**, 29.
- 362 3. Brennan, M.; Le Port, G.; Gormley, R., Post-harvest Treatment with Citric  
363 Acid or Hydrogen Peroxide to Extend the Shelf Life of Fresh Sliced Mushrooms.  
364 *Lebensmittel-Wissenschaft und-Technologie* **2000**, 33, (4), 285-289.
- 365 4. Burton, K. S.; Noble, R., The influence of flush number, bruising and storage  
366 temperature on mushroom quality. *Postharvest Biology and Technology* **1993**, 3, (1),  
367 39-47.
- 368 5. Burton, K. S., Cultural factors affecting mushroom quality - cause and control  
369 of bruising. *Mushroom Science* **2004**, 16, 397-402.
- 370 6. Jolivet, S.; Arpin, N.; Wicher, H. J.; Pellon, G., *Agaricus bisporus* browning:  
371 A review. *Mycological Research* **1998**, 102, 1459-1483.
- 372 7. Heinemann, P. H.; Hughes, R.; Morrow, C. T.; Sommer, H. J. I.; Beelman, R.;  
373 Wuest, P. J., Grading of mushrooms using a machine vision system. *Transactions of*  
374 *the American Society of Agricultural Engineers* **1994**, 37, (5), 1671-1677.
- 375 8. Harrigan, G. C.; Goodacre, R., *Metabolic profiling: its role in biomarker*  
376 *discovery and gene function analysis*. Kluwer Academic Publishers: Boston, 2003.
- 377 9. Melin, A.-M.; Allery, A.; Perromat, A.; Bébéar, C.; Déléris, G.; de Barbeyrac,  
378 B., Fourier transform infrared spectroscopy as a new tool for characterization of  
379 mollicutes. *Journal of Microbiological Methods* **2004**, 56, (1), 73-82.
- 380 10. Bozkurt, O.; Dincer Bilgin, M.; Severcan, F., The effects of diabetes mellitus  
381 on rat skeletal extensor digitorum longus muscle tissue: An FTIR study. *Spectroscopy*  
382 **2007**, 21, 151-160.
- 383 11. Dogan, A.; Siyakus, G.; Severcan, F., FTIR spectroscopic characterization of  
384 irradiated hazelnut (*Corylus avellana* L.). *Food Chemistry* **2007**, 100, (3), 1106-1114.
- 385 12. Toyran, N.; Turan, B.; Severcan, F., Selenium alters the lipid content and  
386 protein profile of rat heart: An FTIR microspectroscopic study. *Archives of*  
387 *Biochemistry and Biophysics* **2007**, 458, (2), 184-193.
- 388 13. Van de Voort, F. R.; Sedman, J.; Ismail, A. A., A rapid FTIR quality-control  
389 method for determining fat and moisture in high-fat products. *Food Chemistry* **1993**,  
390 48, 213-221.
- 391 14. Gidman, E.; Goodacre, R.; Emmett, B.; Smith, A. R.; Gwynn-Jones, D.,  
392 Investigating plant-plant interference by metabolic fingerprinting. *Phytochemistry*  
393 **2003**, 63, (6), 705-710.
- 394 15. Corte, L.; Rellini, P.; Roscini, L.; Fatichenti, F.; Cardinali, G., Development of  
395 a novel, FTIR (Fourier transform infrared spectroscopy) based, yeast bioassay for  
396 toxicity testing and stress response study. *Analytica Chimica Acta* **2010**, 659, (1-2),  
397 258-265.
- 398 16. Santos, C.; Fraga, M. E.; Kozakiewicz, Z.; Lima, N., Fourier transform  
399 infrared as a powerful technique for the identification and characterization of  
400 filamentous fungi and yeasts. *Research in Microbiology* In Press, Corrected Proof.
- 401 17. Ammor, M. S.; Argyri, A.; Nychas, G.-J. E., Rapid monitoring of the spoilage  
402 of minced beef stored under conventionally and active packaging conditions using  
403 Fourier transform infrared spectroscopy in tandem with chemometrics. *Meat Science*  
404 **2009**, 81, (3), 507-514.

- 405 18. Rossel, R. A. V.; Behrens, T., Using data mining to model and interpret soil  
406 diffuse reflectance spectra. *Geoderma* In Press, Corrected Proof.
- 407 19. Garip, S.; Gozen, A. C.; Severcan, F., Use of Fourier transform infrared  
408 spectroscopy for rapid comparative analysis of Bacillus and Micrococcus isolates.  
409 *Food Chemistry* **2009**, 113, (4), 1301-1307.
- 410 20. Preisner, O.; Lopes, J. A.; Menezes, J. C., Uncertainty assessment in FT-IR  
411 spectroscopy based bacteria classification models. *Chemometrics and Intelligent*  
412 *Laboratory Systems* **2008**, 94, (1), 33-42.
- 413 21. Kim, H. T.; Kim, S. S.; Kim, S. J. In *IEEE Engineering in Medicine and*  
414 *Biology, 27th Annual Conference*, Shanghai, China, 2005; Shanghai, China, 2005.
- 415 22. Granitto, P. M.; Gasperi, F.; Biasioli, F.; Trainotti, E.; Furlanello, C., Modern  
416 data mining tools in descriptive sensory analysis: A case study with a Random forest  
417 approach. *Food Quality and Preference* **2007**, 18, (4), 681-689.
- 418 23. Zheng, L.; Watson, D. G.; Johnston, B. F.; Clark, R. L.; Edrada-Ebel, R.;  
419 Elseheri, W., A chemometric study of chromatograms of tea extracts by correlation  
420 optimization warping in conjunction with PCA, support vector machines and random  
421 forest data modeling. *Analytica Chimica Acta* **2009**, 642, (1-2), 257-265.
- 422 24. Gowen, A. A.; Downey, G.; Esquerre, C.; O'Donnell, C. P., A simple method  
423 for estimation of PLS model dimensionality using regression vectors. *Journal of*  
424 *Chemometrics* **2010**, in press.
- 425 25. R\_Development\_Core\_Team *R: a language and environment for statistical*  
426 *computing*, UK, 2007.
- 427 26. Belton, P. S.; Colquhoun, I. J.; Grant, A.; Wellner, N.; Field, J. M.; Shewry, P.  
428 R.; Tatham, A. S., FTIR and NMR studies on the hydration of a high-Mr subunit of  
429 glutenin. *International Journal of Biological Macromolecules* **1995**, 17, (2), 74-80.
- 430 27. Di Mario, F.; Rapanà, P.; Tomati, U.; Galli, E., Chitin and chitosan from  
431 Basidiomycetes. *International Journal of Biological Macromolecules* **2008**, 43, (1), 8-  
432 12.
- 433 28. Workman, J., *Methods and Interpretations in "Handbook of Organic*  
434 *Compounds"*. Academic Press: London, 2001; p 209-242.
- 435 29. Mohacek-Grosev, V.; Bozac, R.; Puppels, G. J., Vibrational spectroscopic  
436 characterization of wild growing mushrooms and toadstools. *Spectrochimica Acta*  
437 *Part A: Molecular and Biomolecular Spectroscopy* **2001**, 57, (14), 2815-2829.
- 438 30. Pierce, J. A.; Rast, D. M., A comparison of native and synthetic mushroom  
439 melanins by fourier-transform infrared spectroscopy. *Phytochemistry* **1995**, 39, (1),  
440 49-55.
- 441 31. Sandula, J.; Kogan, G.; Kacurakova, M.; Machova, E., Microbial (1->3)-beta-  
442 d-glucans, their preparation, physico-chemical characterization and  
443 immunomodulatory activity. *Carbohydrate Polymers* **1999**, 38, 247-253.
- 444 32. Pang, H. L.; Lin, A.; Holford, M.; Enerson, B. E.; Lu, B.; Lawton, M. P.;  
445 Floyd, E.; Zhao, H., Pathway analysis using random forests classification and  
446 regression. *Bioinformatics* **2006**, 16, (22), 2028-2036.
- 447 33. Breiman, L., Random Forests. *Machine Learning* **2001**, 45, 5-32.
- 448 34. Eastwood, D.; Burton, K., Mushrooms - a matter of choice and spoiling  
449 oneself. *Microbiology Today* **2002**, 29, 18-19.
- 450 35. Militello, V.; Casarino, C.; Emanuele, A.; Giostra, A.; Pullara, F.; Leone, M.,  
451 Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and  
452 light scattering. *Biophysical Chemistry* **2004**, 107, (2), 175-187.
- 453 36. Gowen, A. A.; O'Donnell, C.; Taghizadeh, M.; Cullen, P. J.; Frias, J.;  
454 Downey, G., Hyperspectral imaging combined with principle component analysis for

455 bruise damage detection on white mushrooms (*Agaricus bisporus*). *Journal of*  
456 *Chemometrics* **2008**, 22, 259-267.  
457 37. Esquerre, C.; Gowen, A. A.; O'Donnell, C.; Downey, G., Initial studies on the  
458 quantitation of bruise damage and freshness in mushrooms using visible-near infrared  
459 spectroscopy. *Journal of Agricultural and Food Chemistry* **2009**, 57, 1903-1907.  
460 38. Ruiz-Herrera, J., *Fungal cell wall, structure, synthesis and assembly*. CRC  
461 Press: Florida, 1992; p 65.  
462  
463

## FIGURE CAPTIONS

Figure 1 FTIR transmittance spectra of all mushroom tissues in (a) 400-1800 cm<sup>-1</sup> (b) 2800-3050 cm<sup>-1</sup>, and (c) 3050 – 4000 cm<sup>-1</sup> wavenumber ranges

Figure 2. Average undamaged caps spectrum (raw data)

Figure 3. PC1vs PC2 score plots of undamaged mushroom tissue (a) caps; (b) gills (c) stalks and damaged tissue (d) caps; (e) gills and (f) stalks

Figure 4(a) Relative importance plot of variables that are important in the random forest model for predicting damage/undamaged samples. The variable age being the most important followed by the wavenumbers 1868, 1870 and 1845 cm<sup>-1</sup>. 4(b) Relative importance plot of variables that are important in the random forest model for predicting damaged/undamaged samples when age is not a variable. The most important variables are tissue type followed by the wavenumbers 1868, 1870 and 1560 cm<sup>-1</sup>

Figure 5 Frequency of generation of PLS regression models for mushroom post-harvest age on the basis of the number of latent variables selected. (a) undamaged caps, (b) un damaged gills, (c) undamaged stalks, (d) damaged caps, (e) damaged gills and (f) damaged stalks. Abscissa – no. of latent variables in model; ordinate – number of occurrences

## TABLES

Table.1 Summary of results for mushroom discrimination on the basis of damage

	#Samples	#Loadings	% undamaged misclassified mean (std. deviation)	% damaged misclassified mean (std. deviation)
Caps	160	7	4.1 (4.3)	7.6 (4.0)
Gills	160	9	2.1 (3.0)	0.8 (1.7)
Stalks	160	12	1.7 (2.1)	0.6 (1.5)

Table 2 Confusion matrix and the error rate for the prediction of mushroom age. The OOB error rate: 32%. The highlighted numbers are correctly classified samples

	0	1	2	3	4	5	6	7	Error rate
0	<b>49</b>	3	0	3	2	0	3	2	0.18
1	1	<b>42</b>	2	4	0	1	4	6	0.30
2	4	5	<b>43</b>	2	3	0	0	3	0.28
3	1	3	5	<b>47</b>	2	1	0	1	0.22
4	3	0	3	3	<b>32</b>	2	8	9	0.47
5	0	0	3	12	3	<b>29</b>	4	8	0.51
6	1	0	6	0	2	0	<b>48</b>	3	0.20
7	2	1	5	2	2	6	8	<b>34</b>	0.43

0-7: Sample age in days from day zero to day seven

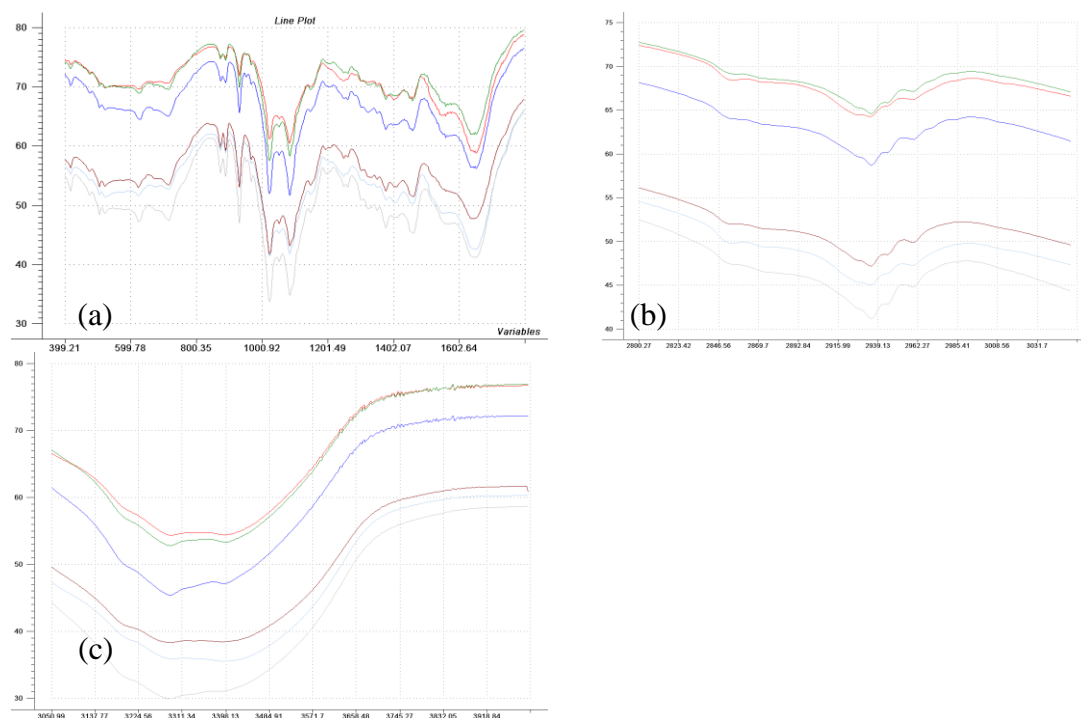
Error rate: The % misclassification for each sample age

**Table 3 Summary of PLS regression results for the prediction of post-harvest age (day 0-7 inclusive) in undamaged and damaged mushrooms**

<i>Treatment</i>	<i>Tissue</i>	<i>#Samples</i>	<i>#Loadings</i>	<i>RMSECV*</i>	<i>RER**</i>
Undamaged	Caps	80	7	1.2	2.0
	Gills	80	7	1.5	1.6
	Stalks	80	7	1.2	1.9
Damaged	Caps	80	7	1.3	1.9
	Gills	80	8	0.8	3.1
	Stalks	80	6	1.2	2.2

\*RMSECV= root mean square error of cross-validation (mean of 100 runs); \*\*RER = SD/RMSECV

## FIGURES



**Figure 1 FTIR transmittance spectra of all mushroom tissues in (a) 400-1800  $\text{cm}^{-1}$  (b) 2800-3050  $\text{cm}^{-1}$ , and (c) 3050 – 4000  $\text{cm}^{-1}$  wavenumber ranges**

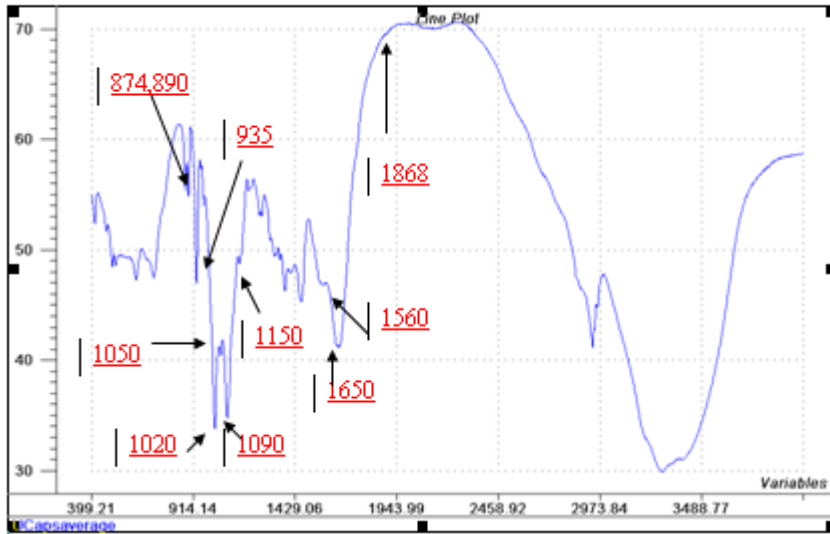
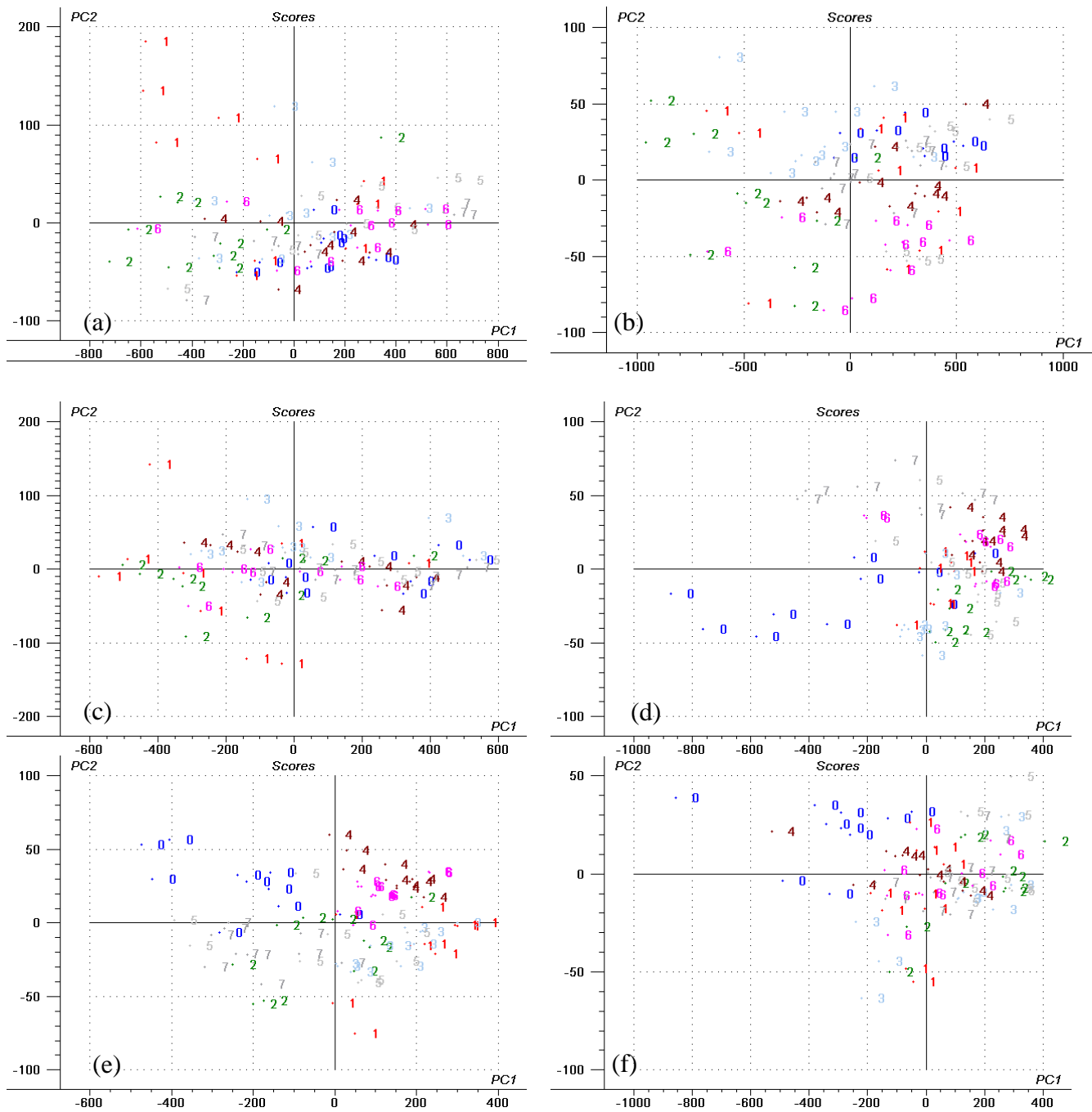


Figure 2. Average undamaged caps spectrum (raw data)



**Figure 3** PC1vs PC2 score plots of undamaged mushroom tissue (a) caps; (b) gills (c) stalks and damaged tissue (d) caps; (e) gills and (f) stalks ;0-7: Sample ages from zero to seven

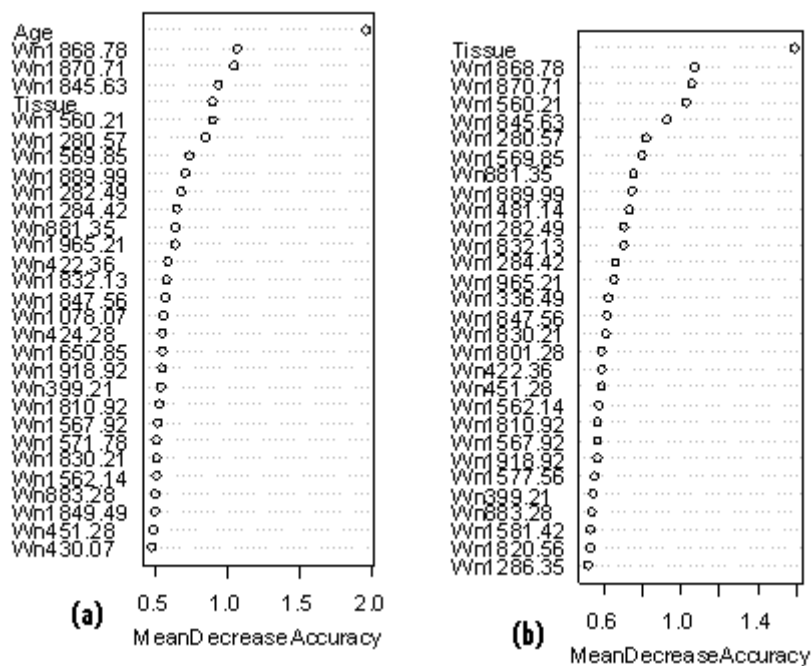
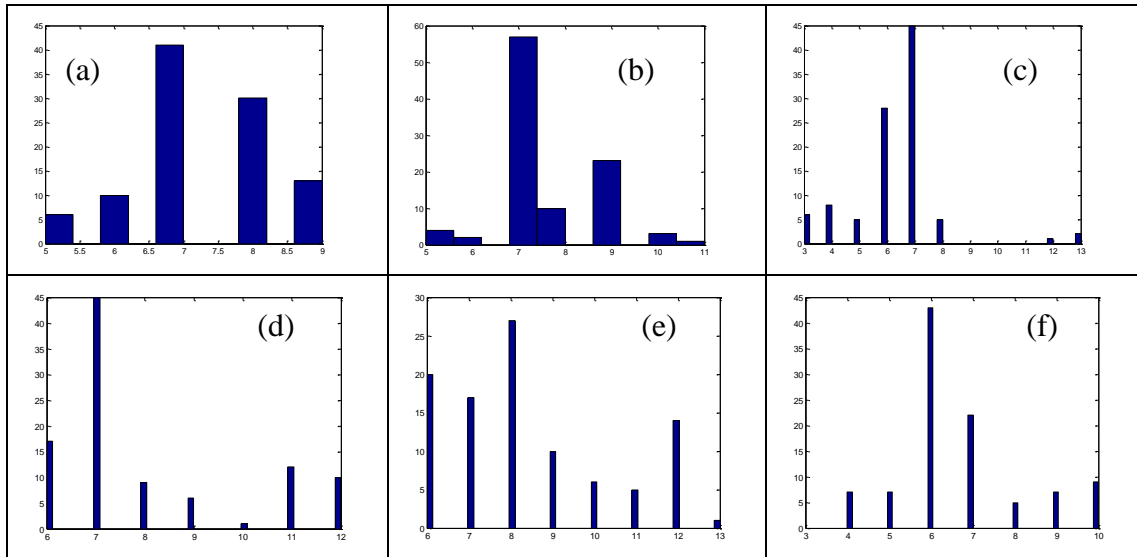


Figure 4(a) Relative importance plot of variables that are important in the random forest model for predicting damage/undamaged samples. The variable age being the most important followed by the wavenumbers 1868, 1870 and 1845  $\text{cm}^{-1}$ . 4(b) Relative importance plot of variables that are important in the random forest model for predicting damaged/undamaged samples when age is not a variable. The most important variables are tissue type followed by the wavenumbers 1868, 1870 and 1560  $\text{cm}^{-1}$ .





**Figure 5. Frequency of generation of PLS regression models for mushroom post-harvest age on the basis of the number of latent variables selected. (a) undamaged caps, (b) un damaged gills, (c) undamaged stalks, (d) damaged caps, (e) damaged gills and (f) damaged stalks. Abscissa – no. of latent variables in model; ordinate – number of occurrences**