Journal of Food Engineering 93 (2009) 502–512

Contents lists available at ScienceDirect

Journal of Food Engineering

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

Study on non-destructive evaluation methods for defect pods for green soybean processing by near-infrared spectroscopy

Panmanas Sirisomboon [a], Yuki Hashimoto [b], Munehiro Tanaka [b,*]

[a] Department of Agricultural Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Bangkok 10520, Thailand
[b] Laboratory of Agricultural Production Engineering, Faculty of Agriculture, Saga University, Saga 840-8502, Japan

ARTICLE INFO

Article history:
Received 24 June 2008
Received in revised form 27 January 2009
Accepted 17 February 2009
Available online 3 March 2009

Keywords:
Green soybean
Defect
Principle component analysis
SIMCA
PLS-DA
Near-infrared spectroscopy

ABSTRACT

Reflectance spectroscopy ranging from visible light to the near infrared region (600–1100 nm) was investigated for the detection of outer and inner defects of fresh green soybean pods. The outer defect parameters, defined by the processing factory which exports frozen green soybean, were hangnail, thinness, brown spot, insect-eaten, rotten, and worm-eaten, and the inner parameters were those caused by disease, i.e. downy mildew and anthracnose, or by worm inside the pod. Using 802 samples, classification models for each group were identified, based on principle component analysis (PCA). Models were developed using primary spectra and second derivative spectra. The PCA score plots could classify clearly the group affected by downy mildew, those with worm inside, and the worm-eaten, rotten, and thin groups from the good pod group. The primary spectra with or without some kind of pretreatment showed a higher classification performance than the second derivative spectra. The good pod model created by primary spectra correctly classified 77.2% of samples as good pods or defective pods. SIMCA showed obviously better performance than PLS-DA in classification of green soybean pods. The good pod model by SIMCA could 100% self-prediction, though it showed low performance in predicting the other groups. This study provided the information by using NIR spectroscopy in the green soybean grading process in order for the appropriate sorting instruments to be developed.

1. Introduction

Green vegetable soybean (*Glycine max*) or Edamame is a nutrient food crop, which is cultivated extensively throughout Asia and the United States (Sciarappa, 2005). It is a popular food in Japan and increases in consumption of edible soybeans in the US have been attributed to their health benefits and flavour (Gaskell, 2001; Iwata et al., 1982). According to the quality standard of the processing factories, especially those performing freezing of pods for export, the green soybean pods are harvested and classified as good pods or defective pods. Physically, the good pods which are accepted by the Japanese market had the following characteristics (Sirisomboon et al., 2007): big pod, which is not less than 4.5–5 cm in length and contains 2–3 seeds; 500 g of green soybean sample must contain not less than 175–180 pods; the pod should be bright green without defect or other color spots; the hair on the pod should be white or gray (Benziger and Shanmugasundaram, 1995; Sitatani and Vasi, 1992). The defective or imperfect pods are separated into small, thin, irregular shape, hangnail, insect eaten, worm eaten, rotten and/or with brown spot and/or diseased, i.e. downy mildew and anthracnose, groups.

After harvest, the green soybean pods are manually classified at the farm and again at the processing factory to separate good quality pods from defective pods. This classification needs a lot of labour and experts. The classifying processes are conducted at first by many farmers in the fields, then the pods are brought to a consolidating point and washed. After that they are transported to the factory. Large loads of these pods are delivered to the factories where the workers stand close to each other along both sides of a slow moving conveyor belt to do the classification. There is also a vibrating tray to separate thin pods from thick pods (more perfect pods) as in the principle described by Sirisomboon et al. (2007). This method, however, still has the problem of pods sticking along the opening slots causing incomplete separation. Quite a long time is needed for the classifying process thus reducing the freshness and nutrient quality of soybeans. Another critical defect is the present of insects or worms inside pods, which is impossible to detect manually by conventional classification methods.

Near infrared (NIR) spectroscopy offers a rapid, non-destructive, and inexpensive method of analysis. It was used to classify insect-infested and sound seeds of a tropical multipurpose tree, *Cordia africana* Lam. by Tigabu and Odén (2002). The calibration model derived by partial least squares regression of orthogonal signal corrected spectra resulted in a 100% classification rate. They concluded that the difference in spectrum and in partial least

* Corresponding author. Tel./fax: +81 952 288798.
E-mail address: mune@cc.saga-u.ac.jp (M. Tanaka).

0260-8774/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jfoodeng.2009.02.019

squares weight indicated that absorbance differences between insect-infested and sound seeds might have been due to differences in composition of chitin and cuticular lipid components as well as moisture content. Tigabu et al. (2007) evaluated the potential of visible (VIS) and NIR reflectance spectroscopy for sorting sound and insect-damaged seeds of *Juniperus procera*. Sound and damaged seeds were distinguished with 90% accuracy by VIS and NIR spectroscopy. They indicated that VIS and NIR spectroscopy have demonstrated a great potential for sorting damaged seeds, thereby upgrading seed lot purity. Zhang et al. (2003) studied tomatoes infected with late blight to illustrate the capability of applying hyperspectral remote sensing to monitor crop disease, and concluded that the NIR reflectance spectra, especially 0.7–1.3 μm, were much more valuable than the VIS range in detecting crop disease. In case of grain, several studies have been performed by Neethirajan et al. (2007) and Elizabeth et al. (2002) who determined that the NIR system is the best method to detect single kernels of wheat that contain live or dead internal rice weevils at various life stages. They also classified sound kernels and kernels containing live pupae, large larvae, medium-sized larvae, and small larvae with an accuracy of 94%, 92%, 84%, and 62%, respectively. The NIR system used to detect insects in kernels can scan 1000 kernels per second (Dowell et al., 1999). A NIR system is rapid (<1 min/sample) and does not require sample preparation (Neethirajan et al., 2007). Baker et al. (1999) indicated that individual kernels of wheat containing immature rice weevils, *Sitophilus oryzae* (L.) (Coleoptera: Curculionidae) parasitized by *Anisopteromalus calandrae* (Howard) (Hymenoptera: Pteromalidae) could be separated from uninfected kernels and unparasitized weevils by using NIR spectroscopy.

A large quantity of raw green soybean pods is fed to the processing factories each day in the harvesting season where the classification parameters are very complex and delicate. It is obviously essential to reduce the time, labour, and experts required in the classification of green soybean pods before processing. Therefore, there is a need for the development of a rapid and non-destructive sorting system to detect the quality of raw material after harvest. The objective of this study was to evaluate the potential of visible and near infrared spectroscopy in classifying the green soybean pods according to the requirements of factory processing.

2. Materials and methods

2.1. Materials

'Ryokkoh 75' green soybean pods were harvested in the early morning from a cultivation field in Mae Suay, Chiangrai Province in the North of Thailand, then immediately brought to the laboratory in the factory in Chiangmai Province for the experiment. Before the scanning of samples, all samples were classified into 10 groups of pods: good pods, and defective pods which were hangnail, thin, brown spotted, insect eaten, worm eaten, rotten, worm inside, downy mildew, and anthracnose. Fig. 1 shows the detail of each type.

(a) *Good pod:* This is a pod with bright green color that has no defects. It contains 2–3 seeds. The pod thickness is more than 7 mm.
(b) *Hangnail:* It is caused by mis-plucking of the pod from the plant, tearing out the pod rib side. The taste of the seed is still good.
(c) *Thin pod:* Small, thin pods whose thickness is less than 7 mm.
(d) *Rotten pod:* The pod becomes rotten due to bacteria or fungus at high temperature. In the case of bacteria, it is wet and has a bad odour, whereas in the case of fungus, it is dry.
(e) *Downy mildew:* There are brown pads on the outer surface of the pod at the seed position. The inner surface of the pod in the same position is also brown. The seed has an unpleasant flavour. This defect is caused by infection with a fungus named "*Pseudoperonospora cubensis*".
(f) *Anthracnose:* This defect is caused by a fungus named "*Collectotrichum sp.*" There are 3–4 very short lines like scratches close to each other on the pod.
(g) *Brown spot:* There are brown spots scattered on the outer surface of the pod. This defect is caused by the effect of high temperature (>30 °C). If a pod is kept for a long time, the spots will appear and become darker. It contains good seed, however, and the taste is good.
(h) *Damage from animal:* There are two kinds, insect-eaten and mouse-eaten. There are holes on the pod. Holes made by insects are small and those made by mice are large.
(i) *Worm inside:* On the outer surface, there is a very small black spot and the pod contains a worm.
(j) *Worm-eaten:* On the outer surface of the pod, there is a very small black hole but inside there is damage caused by a worm. The worm has already eaten the seed and only droppings remain.

2.1.1. Acquisition of VIS and NIR spectra

VIS and NIR reflectance spectra, expressed in the form of log (1/Reflectance), were collected from single pods with FQA–NIR Gun (FANTEC inc., Japan) from 600 to 1100 nm with a resolution of 2 nm using interactance mode. Individual pods were placed on a gold-coated sample cell that had a half hole to integrate light energy. Fig. 2 shows the schematic diagram indicating how to measure the NIR spectrum. To block ambient light, the cell was placed in a small box to enclose the light source, the detector, and the sample, which allowed the collection of radiation reflected from the pod. The scanning duration was 15 s. A polystyrene pack was scanned first as a reference spectrum for measurement.

The spectrum of each pod was scanned and 814 spectra were collected. After that, the abnormal spectra, i.e. the spectra that had an irregular shape, were erased and 802 spectra remained (Table 1). The statistical analysis was performed by ANOVA to test the hypothesis that the means of absorbance at 960 nm of different pod groups of green soybean were equal, and Duncan's multiple range test ($p < 0.05$) was performed to detect differences among the means. The absorbance at only 960 nm which was assigned to water in primary spectra was selected for the test because the NIR optical data at different wavelengths are collinear.

2.2. Spectral data processing

Spectral data were analyzed using the Unscrambler 7.01 software (Camo ASA, Oslo, Norway) and pretreated by second derivatives of the Savitzky Goley method with two different gap sizes, 10 and 20 nm, and 5 different methods, mean normalization, multiplicative scatter correction (MSC), both full MSC and common offset MSC, and baseline correction. After the data pretreatment, the spectra were randomly separated into two sets, the calibration set and the prediction set, by assigning the first three samples into the calibration set, and the following two into the prediction set until all samples had been allocated. The principal component analysis (PCA) was performed to build the classification models for every group, except the groups that had samples of fewer than 20 pods (Worm inside group and anthracnose group) (Table 1). The full cross validation method was applied to PCA. The PCA model of good pods was applied to the prediction set of each group to examine the classification ability of the model. The PCA models were developed for the good pod group with each other group of pods,