Computers and Electronics in Agriculture 145 (2018) 1–10

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques

Julian Coronel-Reyes[a], Ivan Ramirez-Morales[a,b,*], Enrique Fernandez-Blanco[b], Daniel Rivero[b], Alejandro Pazos[b]

[a] Universidad Técnica de Machala, Faculty of Agricultural & Livestock Sciences, 5.5 km Pan-American Av, Machala, El Oro, Ecuador
[b] Universidade A Coruña, Department of Computer Science, 15071 A Coruña (03082), A Coruña, Spain

ARTICLE INFO

Keywords:
Non-destructive
Chemometrics
Freshness
Poultry
Neural networks

ABSTRACT

Currently, consumers are more concerned about freshness and quality of food. Poultry egg storage time is a freshness and quality indicator in industrial and consumer applications, even though egg marking is not always required outside the European Union.

Other authors have already published works using expensive laboratory equipment in order to determine the storage time and freshness of eggs. This paper presents a novel alternative method based on low-cost devices for the rapid and non-destructive prediction of egg storage time at room temperature (23 ± 1 °C).

H&N brown flock with 49-week-old hens were used as a source for the sampled eggs. Samples were scanned for a period of 22 days beginning from the time the egg was laid. The spectral acquisition was performed using a low-cost near-infrared reflectance (NIR) spectrometer which has a wavelength range between 740 nm and 1070 nm. The resulting dataset of 660 samples was randomly split according to a 10-fold cross-validation in order to be used in a contrast and optimization process of two machine learning algorithms. During the optimization, several models were tested to develop a robust calibration model.

The best model used a Savitzky Golay pre-processing technique with a third derivative order and an artificial neural network with ten neurons in one hidden layer. Regressing the storage time of the eggs, tests achieved a coefficient of determination (*R-squared*) of 0.8319 ± 0.0377 and a root mean squared error in cross-validation test set (*RMSECV*) of 1.97 days.

Although further work is needed, this technique shows industrial potential and consumer utility to determine an egg's freshness using a low-cost spectrometer connected to a smartphone.

1. Introduction

Many people find eggs as an affordable source of nutrients. However, the freshness and quality of that source are highly influenced by the storage time and conditions declining along the time. Variability in freshness might be perceived by consumers as lack of quality. Additionally, degradation can reach a point where the egg is unfit for human consumption. For this reason, it is very important to develop methods to better monitor egg storage time (Abdel-Nour et al., 2011; Akter et al., 2014; Akyurek and Okur, 2009; Mathew et al., 2016).

Important and complex changes occur in eggs during storage. Predicting these changes is critical in order to monitor egg freshness. These changes include the thinning of albumen, weakening of the vitelline membrane and an increase in the water content of the yolk. The foaming and emulsifying properties of the albumen and yolk, respectively, are affected by the protein concentration, pH and ionic strength (Karoui et al., 2006).

Storage time, temperature, humidity, air quality, and handling are external factors which can contribute to the degradation of eggs. In particular, the storage time is related to two major issues: the reduction of the nutritional value of eggs (Stadelman et al., 1995) and the decrease of freshness in a logarithmic relation (Silversides and Scott, 2001).

Akter et al. (2014) demonstrated that egg weight, pH, oxidation and Haugh Units are also adversely affected with increasing storage time at room temperature. In the same work, the authors propose a maximum storage time of 14 days at room temperature (28–31 °C).

The freshness can be assessed by physical, biochemical, microbial and sensory parameters. The Haugh Unit (HU) method is a widely used but destructive method to measure egg quality (Haugh, 1937).

* Corresponding author at: Universidad Técnica de Machala, Faculty of Agricultural & Livestock Sciences, 5.5 km Pan-American Av, Machala, El Oro, Ecuador.
E-mail address: iramirez@utmachala.edu.ec (I. Ramirez-Morales).

https://doi.org/10.1016/j.compag.2017.12.030
Received 19 September 2017; Received in revised form 7 December 2017; Accepted 18 December 2017

However, quality measurements based on HU are biased by the strain and age of the hen (Silversides and Scott, 2001). Liu et al. (2007) demonstrated a high correlation between HU and storage time with an R-squared value of 0.9868.

Sensor technologies are an attractive strategy for non-destructive determination of freshness of the egg, either at the production plant or at food industry sites (Galiş et al., 2012; Karoui et al., 2006).

In recent years, non-destructive techniques for assessing freshness and storage time at room temperature have emerged. These techniques include electronic nose (Yongwei et al., 2009), ultrasound (Aboonajmi et al., 2014), ultraviolet-visible spectroscopy (Liu et al., 2007), hyperspectral imaging (Suktanarak and Teerachaichayut, 2017), and near-infrared spectroscopy (Abdel-Nour et al., 2011; Aboonajmi et al., 2015; Lin et al., 2015; Zhao et al., 2010).

The food industry has used NIR spectroscopy for a long time (Stark, 1996) because it is an accurate, rapid, and non-destructive quality analysis technique (Kumaravelu and Gopal, 2015). Recent works have been published using NIR to predict storage time associated with freshness in Atlantic salmon (Kimiya et al., 2013), large yellow croaker (Gangying et al., 2015), snow crab (Lorentzen et al., 2016), pork (Chen et al., 2011), apples (Liu and Tang, 2015), Valerianella locusta (Giovenzana et al., 2014), and eggs (Abdel-Nour et al., 2011; Aboonajmi et al., 2015; Lin et al., 2015; Zhao et al., 2010).

In the past ten years, the evolution of small, hand-held instruments has seen considerable growth (Barton, 2016; Haughey et al., 2014). Recently, some low-cost NIR devices have appeared in the market, making NIRS applications affordable and, therefore, more accessible to the wider public (Haughey et al., 2014).

NIR spectra are the result of vibrational transitions associated with chemical bonds present in most organic compounds (dos Santos et al., 2013; Kumaravelu and Gopal, 2015; Teye et al., 2013). The resulting spectrum is a consequence of the modifications made simultaneously in all the properties of the sample, making the calibration process more complicated (Florkowski et al., 2009; Martens and Naes, 1992).

Chemometrics has become an essential technique aimed at developing NIR calibration models. Using this technique, it is possible to process numerous samples in a short time (Moros et al., 2010).

Multivariate analysis techniques are commonly used to process spectral data, and techniques such as principal component analysis (PCA) and partial least squares (PLS) have been widely used (Kumaravelu and Gopal, 2015). Recently, some machine learning techniques are being presented as good alternatives to the classic techniques because they are based on pattern recognition (Brereton, 2015).

The aim of this study was to assess the potential of a low-cost NIR spectrometer as a non-destructive and rapid technique for egg storage time assessment. A more specific objective was to develop and evaluate a chemometric NIR calibration model based on machine learning techniques for the determination of egg storage time at room temperature.

2. Materials and methods

The overall methodology, as seen in Fig. 1, consists of three segments: the acquisition of the data (Section 2.1), data partition using a cross-validation technique (Section 2.2) and the optimization of the chemometric model (Section 2.3). Each segment consists of several steps. In the following subsections, the methodology is described in detail.

2.1. Spectral acquisition of eggs

A SCiO^{TM} handheld NIR Spectrometer was used to collect the sample, which has a spectral range between 740 nm and 1070 nm (Goldring et al., 2016). Its low price and reduced dimensions allow the researchers to perform rapid tests which could be developed and integrated for example into a smartphone (Cartwright, 2016; Das et al., 2015; Pügner et al., 2016). Previous works have already pointed to the high potential of this low-cost device (Haughey et al., 2014; Schulte et al., 2015).

Samples were collected from intact shell eggs, which were scanned twice from the top in the blunt end using the SCiO^{TM} shade accessory. The use of that accessory helps to avoid the influence of external light when the spectra were sampled and also it also helps to keep the same 10-mm distance in all the 660 collected spectral signals.

While the aim is to develop a non-destructive method to measure the freshness of the eggs, the sampled spectra represent the information of the shell and, consequently, no information of the inner parts are included, e.g., the yolk. A scheme of scanning method is shown in Fig. 2a, while Fig. 2b shows a raw signal of a sample along with the dark and white reference values taken with the spectrometer.

A 1-nm resolution between 740 nm and 1070 nm was used in the spectral data recording, which was lately stored in a cloud-based dataset with their corresponding reference values for time of storage. Using a research license of SCiO Lab, egg spectral signals were downloaded and imported into Matlab (The MathWorks Inc., Natick, MA) in order to develop and optimize the chemometric models.

Those 660 spectral curves are the result of 22 days of continuous monitoring of 30 shell-intact brown poultry eggs with weights between 55 g and 65 g (size M). Eggs used in the study were picked up from a flock of 20.000H&N strain hens between 49 and 52 weeks old. Those hens were housed in a stacked cage system and were fed with a standard ration without the use of egg-laying promoters.

The spectral data used for experimentation were obtained by averaging two duplicate measurements taken successively at the blunt end. Eggs were scanned in the poultry house immediately after being laid (day 0) and then transported to the laboratory in a thermally insulated container. Measurements from day 1 to day 21 were obtained in laboratory conditions monitored hourly at 23 ± 1 °C and a relative humidity of 90 ± 2%. The interval between each measurement was exactly 24 h, and the employed procedure is simple with a very short time required to perform the measurements. A non-destructive technique was used in this experiment because the intention was to understand how the spectrum is modified in each of the eggs over time.

2.2. Data partition

The raw dataset was downloaded and then partitioned using a repeated 10-fold cross-validation technique in order to have training, validation and test subsets for optimization of the calibration model.

The model performance measures should be evaluated in a set of new data which have not been used for the training model. A good model should be able to make accurate estimations on this test data (Mucherino et al., 2009).

Cross-validation is one common technique applied in machine learning to maximize the use of available data. In this technique, the dataset is randomly divided into multiple subsets for training and testing the model. Cross-validation is used to avoid overfitting of the model. (Kuhn and Johnson, 2013; Refaeilzadeh et al., 2009).

In this work, spectral data were divided into training (calibration), validation and test subsets using a variation known as repeated cross-validation (Garcia and Filzmoser, 2015; Kuhn and Johnson, 2013). A repeated 10-fold cross-validation technique was chosen. Therefore, data were split into 10 groups in which 9 are used as calibration/validation sets, and the remaining one is used as a test set. This process was repeated 50 times.

The training and test set were changed until all folds had been tested. Data partition for each fold randomly divided the dataset, allotting 462 samples (70%) for training, 132 samples (20%) for validation and 66 samples (10%) for testing.