Food Control 43 (2014) 88–97

Contents lists available at ScienceDirect

Food Control

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

Quantification of egg yolk contamination in egg white using UV/Vis spectroscopy: Prediction model development and analysis

Linxing Yao [a], Wen Zhou [b], Tong Wang [a,*], Muhua Liu [c,1], Chenxu Yu [c]

[a] Department of Food Science and Human Nutrition, Iowa State University, 2312 Food Sciences Building, Ames, IA 50011, USA
[b] Department of Statistics, Iowa State University, Ames, IA 50011, USA
[c] Department of Agricultural and Biosystems Engineering, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 18 November 2013
Received in revised form
20 February 2014
Accepted 22 February 2014

Keywords:
Egg yolk contamination in egg white
Foaming ability
Statistical prediction modeling
Sensitivity coefficient
Detection limit
UV/Vis

ABSTRACT

The objective of this study was to develop a statistical model for predicting egg yolk contamination level in egg white using a spectroscopic method. Eggs that were stored for 0, 0.5, 1.0, 1.5, 2.5, and 6.5 wk at 4 °C were manually processed to cleanly separate the yolk from the white. Egg white samples containing 0–0.5% (w/w) of yolk were prepared by adding yolk to the white and further diluting with the pure egg white. Transmission spectra of samples were acquired at 500 nm wavelength. The optical absorbance of the "contaminated" egg white samples positively correlated with the yolk concentration, and its intensity was affected by the freshness of eggs, egg variety, and measuring temperature. A nonlinear prediction model, or a detection function, was developed using 182 measurements to predict yolk concentration with a known storage time. This highly sensitive method was validated using 102 separate measurements.

1. Introduction

Eggs from the hen are a high quality food and have the lowest price per unit weight compared to other animal proteins (United States Department of Agriculture, 2012). It consists of a cuticle layer, pore canals, a hard shell, two shell membranes, egg white or albumen, yolk, and vitelline (yolk) membrane (United States Department of Agriculture, 2000). Because of their unique functionalities, egg yolk and white have different markets. Many foods such as meringues, angel cakes, and other bakery products, are prepared using egg white because of its foaming ability. Egg yolk, being an excellent emulsifier, is also used extensively in the food industry, such as mayonnaise, salad dressing, and sauces. In commercial egg breaker plants, shell eggs are broken and the edible portions are separated into yolk and white. Of the 223.7 million cases of shell eggs (or 80.5 billion eggs) produced in U.S. in 2012, approximately 32% underwent breaking for further processing (American Egg Board, 2013). Accidental contamination of egg white by yolk during the breaking operation often occurs, which decreases egg white's functionalities and performance and results in economic loss. Yolk has a detrimental effect on the foaming properties of egg white. The presence of even small quantities of yolk (0.02–0.1%) caused significant reduction in egg white foaming capacity (Kobayashi, Kato, Ohmiya, & Shimizu, 1980; St. John & Flor, 1931; Wang & Wang, 2009). The possible reasons as given in previous studies were that lipids in yolk compete with proteins to be adsorbed to the air–water interface, resulting in a less stable viscoelastic film (Alleoni & Antunes, 2004), and/or the lipids form a complex with ovomucin, one of the major proteins in albumen, and thus inhibit the foaming functionality (Lomakina & Míková, 2006; Stadelman & Cotterill, 1995).

Automatic egg breaking machine is widely used and the breaking speed has increased up to 400 cases (or 144,000 eggs) per hour (Moba, 2013). The problem of yolk contamination is becoming more critical with the high speed of egg breaking and high demand for product consistency. Because it is practically impossible to produce completely yolk-free white on a commercial production scale, it is important to develop a rapid and sensitive in-line detection method to provide timely feedback to the breaker operator in order to improve the processing efficiency and product

* Corresponding author. Tel.: +1 515 294 5488; fax: +1 515 294 8181.
E-mail address: tongwang@iastate.edu (T. Wang).
[1] Present address: Jiangxi Agricultural University, 1225 Zhimin Ave, Qingshanhu, Nanchang, Jiangxi, 330029, China.

http://dx.doi.org/10.1016/j.foodcont.2014.02.037
0956-7135/

quality. Chemical methods for yolk detection are usually tedious and time-consuming. Nielsen (2000) used chemical method to determine yolk contamination in the range of 0.005–0.4% by colorimetric quantification of cholesterol or ester following lipid extraction. Some advanced egg breakers are equipped with an electronic yolk scanner, which catches the color differential between yolk and white under special LED light to detect yolk contamination. The yolk scanner was reported to be able to detect 0.03% fat (equivalent to 0.1% yolk contamination) in albumen (Moba, 2013). UV/VIS/NIR transmission light has been increasingly used to predict egg freshness and quality as well as for many other agricultural products because it is rapid and non-invasive (Birth, Dull, Renfore, & Kays, 1985; Slaughter & Crisosto, 1998). Attempts to predict egg freshness were made using transmission spectral data obtained from 411–1729 (Abdel-Nour, Ngadi, Prasher, & Karimi, 2011), 500–900 (Kemps et al., 2007), and 400–600 nm (Liu, Ying, Ouyang, & Li, 2007). Recently, we reported the feasibility of detecting trace amount of yolk contamination in egg white using UV/VIS spectroscopy, and the initial attempt to select an optimal wavelength for the prediction using partial least square regression (Liu, Yao, Wang, Li, & Yu, 2014).

Chemical and physical properties of egg white change with storage time, temperature and humidity. The pH of the albumen of freshly laid eggs is usually between 7.6 and 8.5 (Macherey, 2007, pp. 3–17). With the loss of CO_2 during storage, the pH of egg white in shell egg rises to a maximum of 9.5 in the first few days. The increase of pH results in the gel-like thick albumen losing its viscosity, a phenomenon referred as egg white thinning (Omana, Liang, Kav, & Wu, 2011). Egg white thinning is an important indicator for egg freshness. As an egg ages, the yolk membrane strength decreases and this makes it even more difficult to cleanly separate yolk from the white (Stadelman & Cotterill, 1995). Egg quality is also affected by genetics, feed and environmental factors (Lomakina & Míková, 2006; Samli, Agma, & Senkoylu, 2005).

The objective of this study was to demonstrate the development of a prediction model for quantifying egg yolk contamination in egg white using UV/Vis spectroscopy. Effects of storage time, pH, egg variety, and measuring temperature on the spectroscopic detection of yolk in egg white were evaluated.

2. Materials and methods

2.1. Materials

Fresh eggs from Hy-Line W-36 laying hens were obtained from a commercial production facility (Stuart, IA). Fresh eggs from three different breeds of hens, Fayoumis-38 (denoted as M), Broilers-67 (B), and Leghorns-W36 (L) raised for poultry breeding research, were obtained from the Iowa State University Poultry Farm (Department of Animal Science, ISU, IA) in order to study the effect of egg variety on spectral signal (Section 2.7).

2.2. Preparation of egg yolk-contaminated egg white

The eggs were processed immediately or stored at 4 °C upon collection. A sufficient number (~30) of eggs were manually broken to obtain pure egg white. The chalazae were manually removed. The adhering albumen on the yolk was removed by rolling the yolk on a filter paper. The pooled egg white (about 800 mL) was homogenized with a Bamix brand hand-held blender (Switzerland) set at the "low" speed for 5 s with one pulse per second, after which an apparent decrease in viscosity of albumen was observed. Then the egg white was mixed for another 15 min using a magnetic stirring bar set at 60 rpm at ambient temperature. Three yolk-contaminated egg white samples with a yolk concentration of 0.1% w/w were prepared by adding 0.1 g of yolk to 100 g of egg white, and the sample was mixed with a magnetic stirring bar at 150 rpm for 15 min. Each of the 0.1% yolk in white samples was diluted using pure egg white to produce a series of 10-g samples with yolk concentrations of 0.075, 0.05, 0.025, 0.02, 0.015, 0.01, 0.0075, 0.005, 0.0025, and 0.001% w/w. A contamination level of 0.5% was prepared by adding 0.05 g yolk to 9.95 g egg white. The 10-g yolk-white samples were mixed with a magnetic stirring bar at 200 rpm for 7 min. As a result, three replicates were obtained for each contamination level. All the sample preparation was done at an ambient temperature and using an analytical balance (Model A 250, Fisher Scientific, Pittsburgh, PA) with 0.1 mg precision. Sample containers (glass beakers) were covered with aluminum foil during the preparation to minimize the moisture loss, and final samples for spectroscopic measurements were stored in screw capped glass vials. The prepared yolk-white mixture and the pure egg white control were stored at 4 °C for 1 h then immediately measured except for those samples used for testing the effect of measuring temperature on signal intensity.

2.3. UV/Vis/NIR spectroscopic analysis

LAMBDA 750 UV/Vis/NIR Spectrophotometer (PerkinElmer, Inc., Waltham, MA) was used for the acquisition of the transmission spectra of the samples. The sample was scanned from 200 nm to 860 nm with 2 nm interval using Deuterium lamp. Data acquisition was controlled by UV/Vis Winlab software (PerkinElmer, Inc., Waltham, MA).

2.4. Effect of storage time on the intensity of spectral signal

Eggs (Hy-Line W-36) were stored at 4 °C and sampled at 0, 0.5, 1.5, 2.5, and 6.5 wk. About 30 eggs were processed (as described in Section 2.2) at each sampling time to study the effect of storage time on yolk detection.

2.5. Effect of pH on the intensity of spectral signal

Egg white obtained from 30 fresh eggs (Hy-Line W-36, pH 8.1) was pooled and its pH was adjusted to 9.0 using 1 N NaOH. Then a series of samples with yolk concentration from 0.001 to 0.1% was prepared from pH adjusted and non-adjusted egg white as described in Section 2.2.

2.6. Effect of measuring temperature on the intensity of spectral signal

Because eggs in the commercial plant may be stored in cold room before processing, or they may be left at ambient temperature before entering the breaking room, the actual temperature of the sample during breaking and measurement may vary. We examined the effect of measurement temperature (4 °C, 15 °C and 25 °C) on the spectral characteristics. Samples prepared from the fresh eggs (Hy-Line W-36) were used. Each 10-g yolk-white sample was divided into 3 subsamples and stored separately at 4 °C, 15 °C, and 25 °C for 1 h before spectroscopic measurements. Three replicates for each temperature and concentration combination were used. The storage temperature was monitored using a thermometer and was consistent throughout the experiments. The spectrophotometer did not have temperature control, but because each scan was finished in less than 5 s, the deviation of the measuring temperature from the storage temperature was considered negligible.

2.7. Effect of egg variety on the intensity of spectral signal

Fresh eggs from Hy-Line W-36 laying hens (or commercial eggs, abbreviated as H) were compared with three research farm eggs (i.e. L, M, and B) for the effect of egg variety on the UV/Vis detection signal.

2.8. Statistical prediction model based on absorbance values and storage time

For the prediction model development, we used 182 data points whose absorbance values were obtained at 500 nm from the study described in Section 2.4. The relationships between absorbance values and contamination levels, subjected to the different storage times, were explored geometrically. The relationship was modeled by an absorbance function $\mathcal{A}(c,t)$, which was fitted to the data using the generalized least squares (McCullagh & Nelder, 1989), where the form of the weighed matrix was of compounded symmetry based on the experimental design. The 95% Wald confidence intervals for the fitted parameters were obtained for inference.

The prediction model of contamination levels using absorbance values and storage times was derived by the inverse absorbance function, $\mathcal{A}^{-1}(c,t)$. The prediction intervals were derived using the Delta Method (Serfling, 1980). Independent data set of 102 samples (all fresh eggs with storage time of 2 h) was used for validation and examination of the method, and comparison of the prediction performances of the proposed model and other data driven methods from statistical learning were reported as well.

Defined by the fitted parameters of the proposed prediction model, sensitivity coefficient showed that the model was able to boost weak signal for fresh eggs to detect low contamination levels. The limit of detection (LOD) for contamination in fresh eggs was identified based on the prediction intervals.

In addition, a two-way factorial analysis was performed on the log-transformed data to test the significance of the differences among egg variety across yolk contamination, and measuring temperature across yolk contamination.

The computation and statistical analysis were conducted using SAS (Version 9.3, SAS Institute Inc. Cary, NC), R (version 2.15.3) and Matlab (version R2012a).

3. Results and discussion

3.1. Wavelength selection for model development

Statistical learning methodologies such as partial least square regressions (PLSR), ridge regressions and principal component regressions (PCR) were applied for optimal wavelength selection based on an implicit linear model assumption (Abdel-Nour et al., 2011; Osborner, Jordan, & Kunnemeyer, 1997). Based on a method similar to that of PLSR, optimal wavelengths 378 and 430 nm were selected for the prediction of yolk contamination in our previous report (Liu et al., 2013). Recently, Bayesian variable selection was proposed to accommodate the high dimensionality and lack of independence for optimal wavelength selections (Brown, Fearn, & Vannucci, 1999; Brown, Vannucci, & Fearn, 1998; Tadesse, Vannucci, & Lio, 2004). For such work, one may adopt methods such as Bayesian variable selection and high dimensional hypothesis testing to refine the optimal wavelength selection without assuming linearity and target on prediction powers. In this study, we used 500 nm for all the analysis to demonstrate the development of a statistical prediction model. Based on our experimental observation (Sections 3.2 and 3.3), it is highly likely that the difference in the absorbance of different level of yolk contamination was resulted from the turbidity change of egg white, so that 500 nm, a wavelength from the visible wavelength range, was appropriate to use as used by many other emulsion turbidity studies. Our demonstrated prediction model is compatible with other wavelengths as well.

3.2. Effect of yolk concentration on the optical absorbance of egg white

Fig. 1 shows that the absorption spectra of egg white (with or without yolk) had maximum absorbance peaks at the region of 280–310 nm, which might be attributed to the aromatic amino acids in the egg white protein. Within the range of visible light (400–700 nm), the absorbance of egg white decreased with increasing wavelength, as typically observed in colloids when the particle size is smaller than the wavelength, and the so-called Rayleigh scattering occurs (Pickering, 1992). Liquid egg white is a colloidal solution of water (84–89%) and protein (10–11%) (Li-Chan

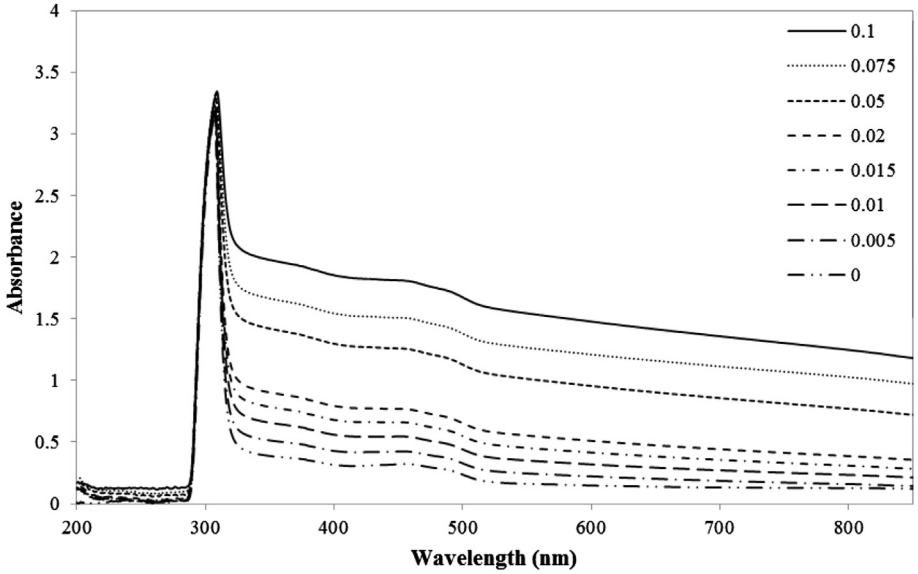

Fig. 1. The transmission spectra (200–860 nm) of egg white samples containing 0–0.1% egg yolk. The values in the legend were the yolk concentration (%). The scans of 0.5, 0.025, 0.0075, 0.005, and 0.001% were not included to improve the graph clarity, but their curves were in the same shape and order. Measuring temperature was 4 °C.

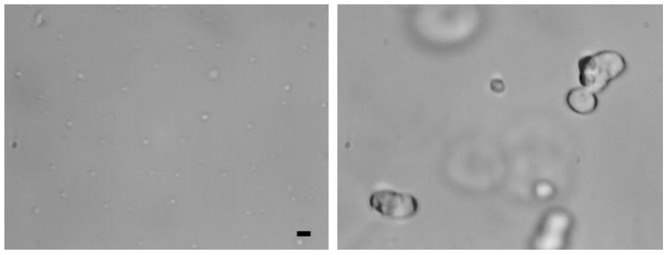

Fig. 2. Images of egg white proteins observed under optical microscope with 100× magnification. Left: fresh egg white; right, fresh egg white contains 0.5% yolk. The measuring bar represents 10 μm.

& Nakai, 1989). Egg white proteins are mainly glycoproteins and water soluble. Ovalbumin is the most abundant protein in egg white, constituting about 54% of total egg white protein, followed by ovotransferrin, ovomucoid, ovomucin, lysozyme, globulin and other minor proteins. Ovomucin is the main contributor for its gel-like structure, and globulin is important for the foaming properties of egg white (Li-Chan, Powrie, & Nakai, 1995). Egg white is a translucent and viscous colloid.

The major constituents in egg yolk are water (49%), lipid (32–36%) and protein (16–17%) (Li-Chan & Nakai, 1989). The yolk lipids are present in the lipoprotein particles. Adding egg yolk to egg white meant adding less dispersible components to the albumen, increasing the turbidity of the egg white as observed visually. It was noticed that egg yolk also caused random aggregation of egg white's globular proteins as shown in Fig. 2. The complexes and aggregates formed by yolk component and ovalbumin change the folding structure of the protein (Lomakina & Míková, 2006). Both the insoluble lipoproteins and induced aggregates altered the scattering of the transmitting light and thus resulted in a higher absorbance value. Therefore, we observed a consistent increase of absorbance with yolk concentration at a fixed wavelength as illustrated in Figs. 1 and 3.

3.3. Effect of storage time on optical change of egg white contaminated with yolk

Egg storage time affected the degree to which yolk contamination increased absorbance (Fig. 4). The slope of the curve decreased with longer storage time, which means that for aged eggs, the difference in absorbance among different yolk contamination levels became smaller. In general, at a given yolk concentration, aged egg gave lower absorbance than fresh one. The effect of storage on egg white quality has been well studied (Lomakina & Míková, 2006; Stadelman & Cotterill, 1995). The change of pH and viscosity of egg white, and protein degradation during storage may have resulted in the interaction and relationship between yolk concentration and light transmittance as shown in Fig. 4. The main effect of egg storage time was found to be significant ($p < 0.05$). The greatest changes of absorbance occurred in the first two weeks.

In commercial egg production, eggs from in-line systems are washed and processed within a day after they are laid. However, eggs from off-line systems may be stored up to 7–10 days under refrigeration before they are processed (American Egg Board, 2014). The difference in the egg production and processing management systems can result in variations in hen's age, egg storage time, and temperature of eggs after washing and breaking. The liquid egg from the refrigerated eggs is expected to be cooler than that from the in-line eggs. These can all have great impact on the spectrophotometric measurement.

3.4. Effect of pH on optical change of egg white contaminated with yolk

One of the most pronounced change during egg storage is the increase of egg white pH due to the loss of CO_2 through the pores of the shell. The high pH leads to proteolysis, and along with the enzyme catalyzed protein hydrolysis, the thinning of egg white occurs (Omana et al., 2011). Light scattering is expected to decrease with increasing amount of soluble peptides. When yolk concentration was less than 0.05%, no apparent difference was detected between the pH adjusted (pH 9.0) and control (pH 8.1) samples (Fig. 5). However, at higher yolk concentrations 0.075 and 0.1%, the pH adjusted samples gave much lower absorbance than the non-adjusted sample. It should be noted that samples with adjusted pH will not have experienced chemical degradation. Any changes in these samples will be due to physical interactions; for example, protein dispersion (Zayas, 1997, p. 33), leading to lower absorption rates than those observed with low pH samples.

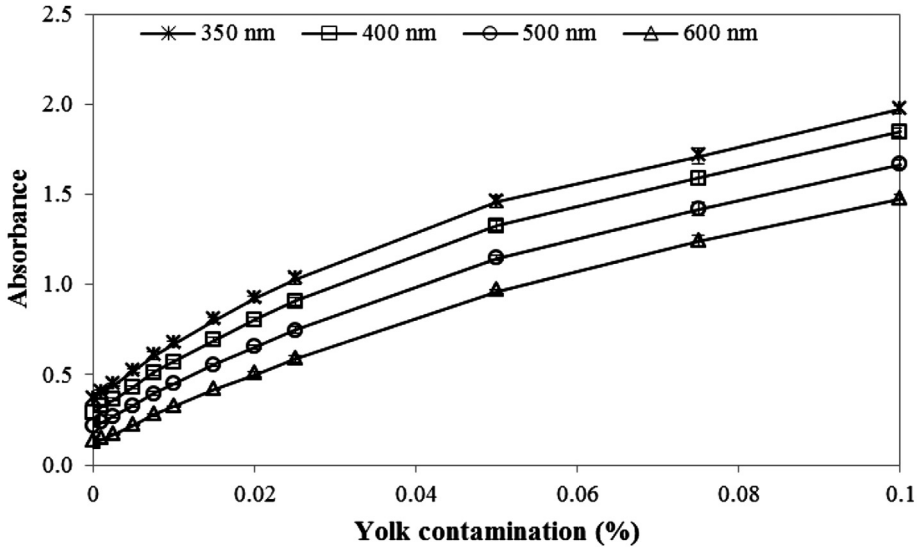

Fig. 3. The change of absorbance with yolk concentration at various wavelengths. The data were from one batch of fresh eggs. The error bars represent the standard deviation of three samples.

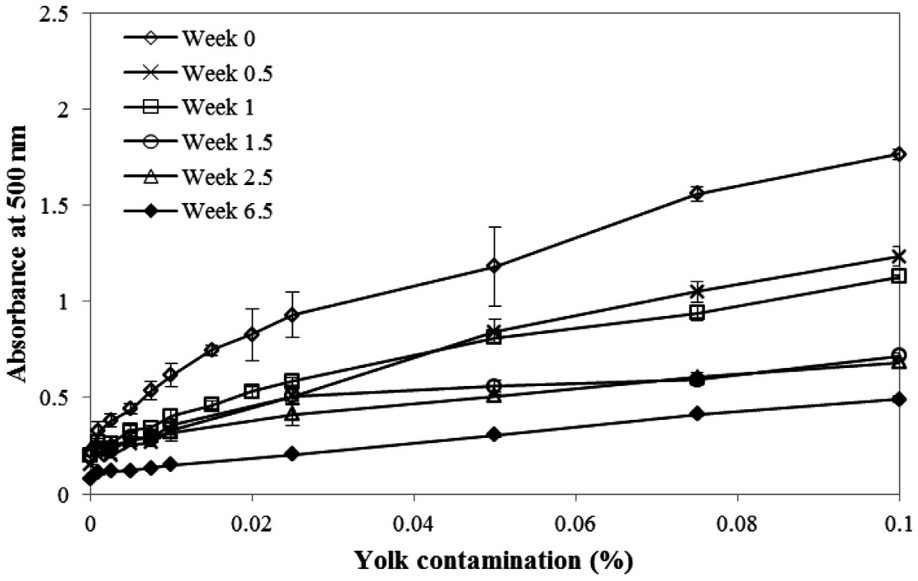

Fig. 4. Effect of egg storage time on the absorbance of egg white with various levels of yolk contamination. Error bars represent the standard deviation of three samples. The measuring temperature was 4 °C.

3.5. Effect of egg variety on optical change of egg white contaminated with yolk

Among the four types of eggs there were differences in size and shell color. The genetic lines of research farm laying hens were unique to the ISU research farm due to their genetic selection program. Regardless of the yolk contamination level, the commercial eggs H (Hy-Line W-36) and the research farm egg M gave significantly higher absorbance than the other two research farm eggs L and B ($p < 0.05$). There was no significant difference detected between H and M, or B and L (Fig. 6). Genotype and age of laying hens (can be up to 130-wk after two moltings) affected the egg quality as reported by Scott and Silversides (2000) and Silversides and Scott (2001). Genetic factors seemed to only have a minor effect on egg white-yolk mixture's absorbance in this study. With the use of uniform breed, commercial production methods and feeding practices in compliance with U.S. egg industry regulations, we do not expect significant inherent egg quality differences.

3.6. Effect of measuring temperature on optical change of egg white contaminated with yolk

Measuring temperature affected the absorbance (Fig. 7). Samples measured at 4 °C gave significantly higher absorbance than those at 15 and 25 °C ($p < 0.0001$). At low concentration (<0.025%), the effect of temperature on the absorbance was greater. There was no significant difference between 15 and 25 °C ($p > 0.05$). The absorbance obtained from pure egg white as measured at the three temperatures was the same. It suggests that the impact of measuring temperature resulted from the egg yolk. At 4 °C, some of the yolk lipids might have crystallized or the lipoprotein particles formed larger aggregates and thus altered the light scattering.

The Food and Drug Administration's (FDA) 2009 Shell Egg Safety Rule requires egg producers to refrigerate shell eggs at 7.2 °C if it will take longer than 36 h to reach the processing facility or during storage and transportation to prevent *Salmonella* Enteritidis contamination (FDA, 2009). However, egg breaking typically occurs at an ambient temperature. Therefore, the actual temperature during UV/Vis measurement may vary depending on the egg's "stand-by" time. Good manufacture practice for maintaining a consistent temperature during breaking should be enforced so that temperature is not a random variable during breaking and product monitoring. If the measuring temperature is determined as an influencing factor, it should be considered to ensure accurate yolk quantification.

Although the prediction model developed in Section 3.7 used data from one temperature (4 °C), the proposed modeling approach can be used for any new set of data at any temperature. Therefore, in the future designing of an in-line detector, the temperature measured during breaking can be built into the prediction model as a calibration tool for an individual breaker plant.

3.7. Development of the prediction model for quantifying yolk contamination level

Existing prediction models naturally can be divided into two groups: the data driven models like support vector machine, random forest, gradient boosting machine or principal component regression (Hastie, Tbshirani, & Friedman, 2009); and the mechanism driven models like structure equation models and mathematical models. Data driven models have the advantages for data with high dimensionality and complex interactions, but lack of interpretability in general. In contrast, for data with relatively small number of predictors, mechanism driven models can provide reasonably accurate predictions and preserve interpretations of underlying mechanisms. To predict the yolk contamination level with absorbance values, we therefore utilized mechanism driven model for its interpretability and robust performance.

We focused on the relationship among the contamination levels, absorbance values and storage times for modeling.

3.7.1. Modeling development

We performed log-transformation to the data (absorbance values) for analysis and modeling to accommodate the fact that the margin distribution of data was right skewed. Data presented later are log-transformed data unless otherwise specified.

The experiment of testing the effect of storage time on the absorbance of egg white with various yolk contamination levels followed a split-plot design, which can be modeled by a linear mixed model (Morris, 2010):

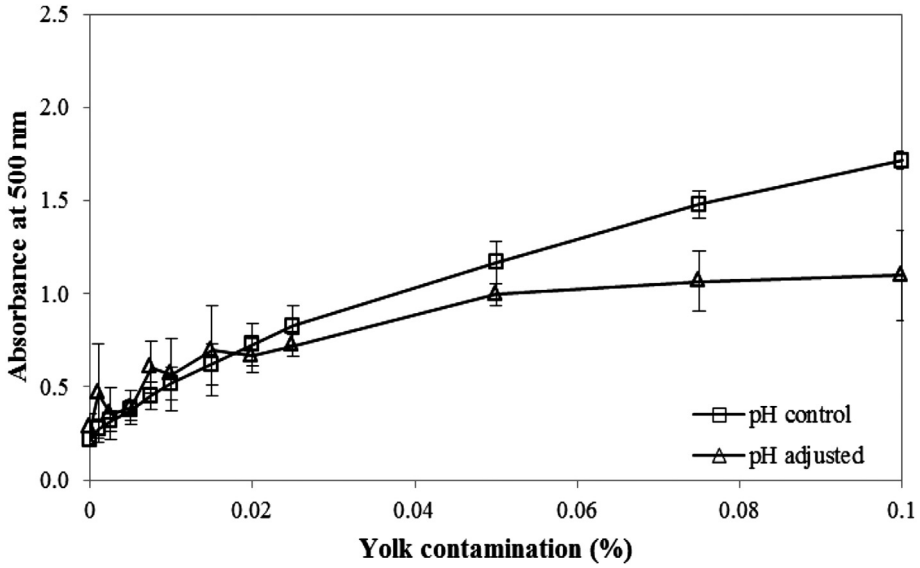

Fig. 5. Effect of pH on the absorbance of egg white with various levels of yolk contamination. Data were the average of three measuring temperatures of 4, 15, and 25 °C (each temperature has three replicates). Error bars represent the standard deviations. The pH values are 9.0 and 8.1 for the pH adjusted and pH control samples.

$$Y_{ijk} = \mu + \alpha_i + \beta_{k(i)} + (\alpha\beta)_{ik(i)} + \eta_{ij} + \varepsilon_{ijk(i)},$$

where α_i ($i = 1,\cdots,6$) denotes the storage time, $\beta_{k(i)}$ ($k = 1,\cdots n_i$) denotes the contamination level, and $(\alpha\beta)_{ik(i)}$ denotes the interaction, whose existence is suggested by Fig. 4. Independent random components $\eta_{ij} \sim N(0, \sigma_g^2)$ and $\varepsilon_{ijk(i)} \sim N(0, \sigma_s^2)(j = 1, 2, 3)$ model the sample's variation and measurement error, respectively. Significant effects of storage time, contamination level and their interactions were detected (p-values of F-test were all less than 0.0001). The estimates of random components of $\widehat{\sigma}_g^2 = 0.00191$ and $\widehat{\sigma}_s^2 = 0.0034$ were both significant (p-values are 0.0189 and less than 0.0001, respectively). Hence, the correlations between subsamples within each replicate were incorporated for model development.

The interaction of contamination level and storage time was considered for the prediction model. We denoted the log-absorbance value by $\mathcal{A}$ and considered it as a function of contamination concentration c and storage time t, i.e. $\mathcal{A} := \mathcal{A}(c, t)$. Function $\mathcal{A}$ is thus the **absorbance function**. Fig. 4 suggests that $\mathcal{A}(c, \cdot)$ is a family of concave functions with respect to c. In addition, the growth rate of the absorbance function with respect to the concentration, $\partial\mathcal{A}(c, t)/\partial c$, is nonnegative for a small enough concentration $(\partial\mathcal{A}(c, t)/\partial c)(t_2) < (\partial\mathcal{A}(c, t)/\partial c)(t_1)$ with $t_2 > t_1$. Plotting absorbance values against the storage time for different concentrations (figure not included) revealed that with the increasing of storage time t, the absorbance function $\mathcal{A}(\cdot, t)$ was decreasing in a linear fashion and $\mathcal{A}(\cdot, t)$'s dependence on storage time became clearer as the contamination level increased. In addition, plotting contamination levels against the absorbance values suggests that the relationship between them can be characterized by an exponential function, whose intercepts change with storage time.

Based on the above observations, the contamination level can be described by the absorbance values by

$$c(\mathcal{A}, t) = e^{(\beta_0 + \beta_2 t^\alpha) + (\beta_1 + \beta_3 t^\alpha)\mathcal{A}} \quad (1)$$

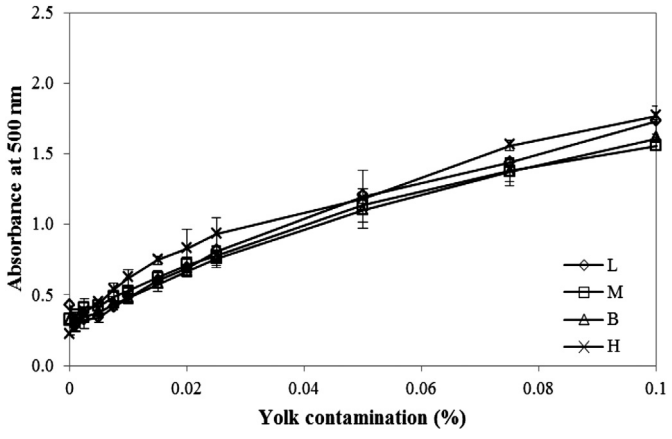

Fig. 6. Effect of egg variety on the absorbance of egg white with various levels of yolk contamination. H stands for Hy-Line W-36, M for Fayoumis-38, B for Broilers-67, and L for Leghorns-W36. Error bars represent the standard deviation of three samples. The measuring temperature was 4 °C.

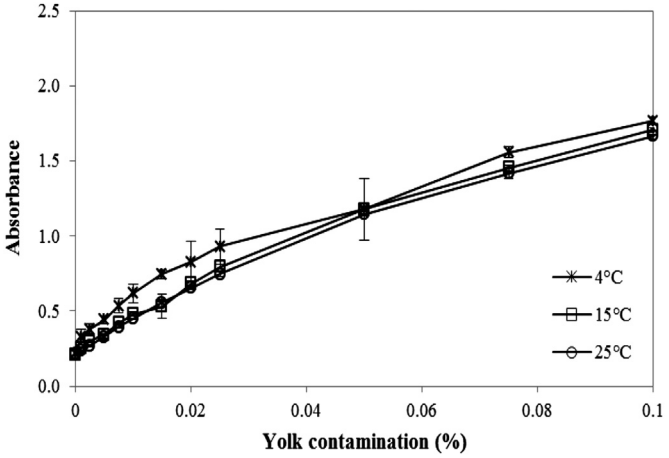

Fig. 7. Effect of measuring temperature on the absorbance of egg white with various levels of yolk contamination. Fresh eggs of Hy-Line W-36 were used. Error bars represent the standard deviation of three samples.

Table 1
Point estimates and confidence intervals of parameters in Equation (2).

Parameter	Point estimates	95% Wald confidence interval
$\widehat{\beta}_0$	−3.434	(−3.740, −3.128)
$\widehat{\beta}_1$	2.318	(2.150, 2.487)
$\widehat{\beta}_2$	1.396	(1.014, 1.779)
$\widehat{\beta}_3$	0.290	(0.158, 0.423)
$\widehat{\alpha}$	0.583	(0.431, 0.734)

so that the absorbance function is

$$\mathcal{A}(c,t) = \frac{\ln(c) - \beta_0 - \beta_2 t^{\alpha}}{\beta_1 + \beta_3 t^{\alpha}} \tag{2}$$

where $\beta_3 > 0$ and $\alpha > 0$ model the effect of storage time on the growing rate of absorbance value with respect to contamination level. Contamination level c, theoretically, might be zero so that in practice Equation (2) can be approximated by a perturbed version

$$\mathcal{A}^{\varepsilon}(c,t) = \frac{\ln(c+\varepsilon) - \beta_0 - \beta_2 t^{\alpha}}{\beta_1 + \beta_3 t^{\alpha}}$$

for any $\varepsilon > 0$ (a small numerical value to avoid computational difficulty). Theoretically, $\mathcal{A}^{\varepsilon} \to \mathcal{A}$ as $\varepsilon \to 0$ by the continuity of the absorbance function with respect to contamination level. As the contamination level approaches to limit, the quantity $-(\beta_0+\beta_2 t^{\alpha})/(\beta_1+\beta_3 t^{\alpha})$ defines the limiting absorbance values with respect to different storage times.

Equation (1) is termed the **detection function**, which is automatically nonnegative for predicting contamination levels.

3.7.2. Model fitting and inference

We fitted the absorbance function to the 182 data points using the generalized least squares with compound symmetric covariance. The maximum likelihood estimates and asymptotic confidence intervals of parameters are reported in Table 1. Likelihood ratio test of the fitted model comparing to the linear mixed model had p-value of 0.99, which shows that the proposed model in Equation (2) is sufficient to represent the data.

All the parameters are significantly different from zero from the inference based on confidence intervals. Estimates $\widehat{\beta}_3 > 0, \widehat{\alpha} > 0$ imply that the growth rate of absorbance values with the increasing contamination levels will be compromised by long storage time. This reflects that the detection power of contamination is reduced as storage time increases. In addition, the estimated ceiling of absorbance values (when the contamination level approaches to the limit) $-\frac{\widehat{\beta}_0+\widehat{\beta}_2 t^{\widehat{\alpha}}}{\widehat{\beta}_1+\widehat{\beta}_3 t^{\widehat{\alpha}}}$ decreases with respect to increasing storage times. In summary, longer storage time reduces the detection power of contamination level.

Fig. 8 demonstrates the fitted detection function, from which we observe that for the same absorbance values, it requests or predicts higher contamination level for longer storage time, which is consistent with the observations in Section 3.3 that storage time reduces the detection power.

3.7.3. Predictions based on detection function

To utilize the detection function in Equation (1) for predicting the contamination level based on the storage time and measured absorbance value, we only need to plug in the estimates from Table 1 into Equation (1). For example, if the measured absorbance value is $\mathcal{A}_{\text{measured}}$ (in log-scale) at storage time t_{measured}, then the predicted contamination level is

$$\widehat{c}_{\text{pred}} := \widehat{c}(\mathcal{A}_{\text{measured}}, t_{\text{measured}}) = \exp\Big(\Big(\widehat{\beta}_0 + \widehat{\beta}_2 t_{\text{measured}}^{\widehat{\alpha}}\Big) + \Big(\widehat{\beta}_1 + \widehat{\beta}_3 t_{\text{measured}}^{\widehat{\alpha}}\Big)\mathcal{A}_{\text{measured}}\Big).$$

It remains to derive the variance of prediction $\widehat{c}(\mathcal{A}_{\text{measured}}, t_{\text{measured}})$ to complete the statistical prediction. As the detection function, Equation (1), is differentiable on the parameter space, we have the prediction confidence interval derived asymptotically using the Delta method (Serfling, 1980). The derivation is presented in Appendix.

Fig. 9 demonstrates the performance of prediction using Equation (1) for 102 independent validation samples. The overall performance of prediction is satisfactory and as expected, the prediction variability increased with increasing predicted contamination level. One great advantage of this prediction model is demonstrated by its good prediction performance at low concentration range, which would be practically critical for a powerful yolk contamination detector. In practice, precise measurements are expected to provide less prediction variability. Table 2 reports the comparison of prediction performances between the proposed detection function and other widely used data driven prediction

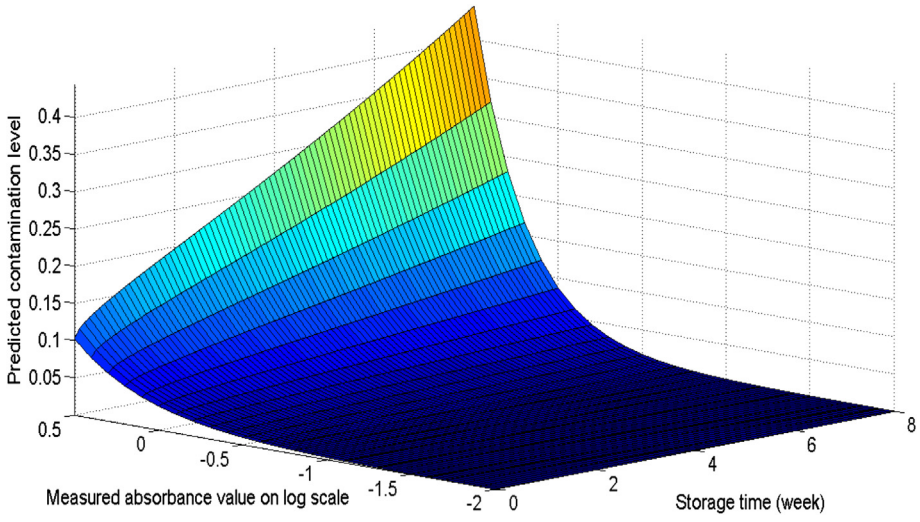

Fig. 8. Fitted detection function $c(\mathcal{A}, t)$, the detection surface, defined in Equation (1).

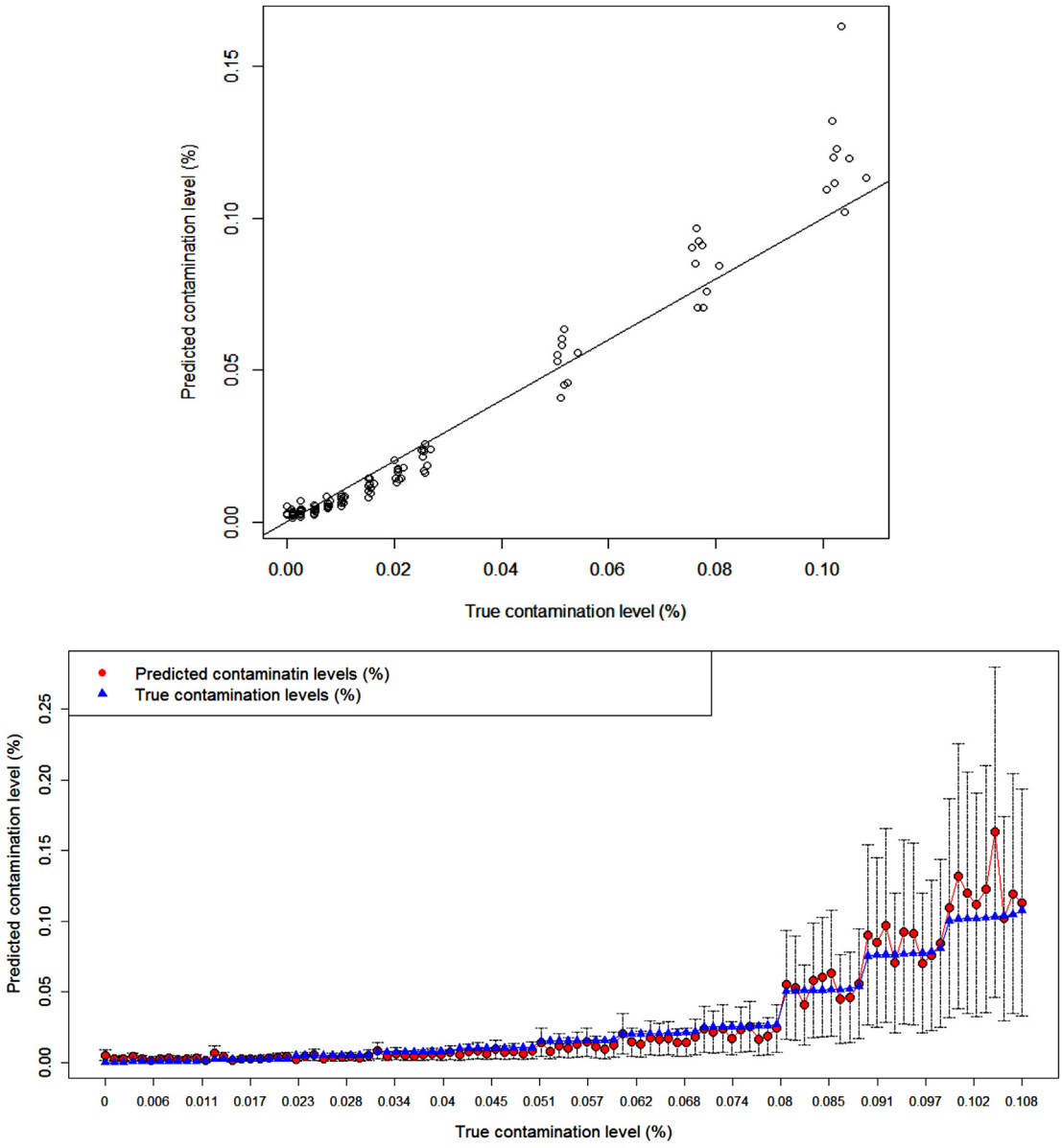

Fig. 9. Validation of the detection function, Equation (1), using 102 independent data points collected with storage time equal to 0.02 wk. Upper: The straight line indicates the exact match. Bottom: The vertical bars indicate 95% prediction intervals.

models: support vector machine, random forest, and gradient boosting machine. The detection function performed reasonably well in comparison to other models.

3.7.4. Sensitivity of the method

Sensitivity is defined as the magnitude of change of the log-absorbance values with respect to changes in yolk contaminations. Using Taylor expansion, it is easy to see that from Equation (1)

$$\Delta\mathcal{A} := \mathcal{A}_2 - \mathcal{A}_1 \approx (c_2 - c_1)\frac{2-(c_2-c_1)}{2c_1(\beta_1+\beta_3 t^{\alpha})}.$$

We focused on low levels of contamination so that the increment $\Delta c := c_2 - c_1 \sim o(1)$, and therefore we can define the **sensitivity coefficient** to be

$$s(t) := \frac{1}{c_1(\beta_1+\beta_3 t^{\alpha})}. \tag{3}$$

Given the results in Table 1, for fresh eggs (i.e. $t \to 0$) the sensitivity coefficient is approximately

$$\widehat{s} \approx \frac{1}{c_1\widehat{\beta}_1} = \frac{1}{2.318c_1} = \frac{0.43}{c_1}$$

with 95% confidence interval being $((1/2.487c_1), (1/2.15c_1)) = ((0.402/c_1), (0.465/c_1))$. The quantity c_1, in practice, represents almost zero contamination and therefore is very small (i.e. $c_1 \sim o(1)$). Hence, for fresh egg with very low level of contamination, the proposed detection function is able to boost small signals for detection. For example, if the true contamination level has c_2=0.001% and the control level has c_1=0.0001%, then the detection function (Equation (1)) can magnify the contamination signal by 4330 times. Thus, the proposed detection function is powerful for detecting small contaminations of fresh eggs.

Also, it is noticed that $s(t)$ does decrease as storage time increases, i.e. storage time will compromise the detection power,

Table 2
Comparison of prediction performances for the 102 validation data points.

	Detection function, Equation (1)	Gaussian-SVM	Random forest	Polynomial-SVM	GBM
MSE	0.0090	0.0097	0.0129	0.0098	0.0103

Note: The criterion is root mean squared error (RMSE). SVM stands for support vector machine. GBM stands for gradient boosting machine.

which is consistent with observations and discussion in previous sections.

3.7.5. Limit of detection (LOD)

Statistically, a 95% prediction interval with zero included indicates inability to detect the signal or detect the contamination level, which is an analogy to the classical hypothesis testing (Lehmann & Romano, 2008; Vardeman & Jobe, 2000). Hence, the limit of detectable contamination level, c_{LOD} is the level of true contamination at which the lower bound of the prediction interval is less than zero. Based on the derivation in Appendix, c_{LOD} is the solution to the following stochastic algebraic inequality

$$c^2 e^{2\left(\widehat{\beta}_1+\widehat{\beta}_3 t^{\widehat{\alpha}}\right)\varepsilon}\left(1-Z_{\frac{\alpha}{2}}^2\sigma^2\left(\widehat{\beta}_1+\widehat{\beta}_3 t^{\widehat{\alpha}}\right)\right) \le Z_{\frac{\alpha}{2}}^2 \widehat{\nabla c}^{\mathrm{T}}(c)\widehat{\Sigma}\widehat{\nabla c}(c) \qquad (4)$$

where $\widehat{\nabla c}(c)^{\mathrm{T}}\widehat{\Sigma}\widehat{\nabla c}(c)$ is defined in Appendix, parameters are in Table 1, $Z_{\alpha/2}$ is $(1-(\alpha/2))$ quantile of the standard normal distribution. Random variable $\varepsilon \sim N(0,\sigma^2)$ models the measurement error of the process collecting log-absorbance values, where σ^2 models the measurement variance.

Using the Monte Carlo method (Robert & Casella, 2005), the stochastic inequality (Equation (4)) is solved for c given fixed σ^2 and small storage time (t=0.024 is used in the simulation study, which stands for 4 h, i.e. fresh eggs). The numerical solution suggests that for a large measurement variance (about 0.2 units) the limit of detection of contamination level is about 5.75×10^{-4}%, i.e. 5.75 ppm. In addition, simulation studies (Fig. 10) show that LOD increases as the measurement variance increases (i.e. the noise will compromise the signal).

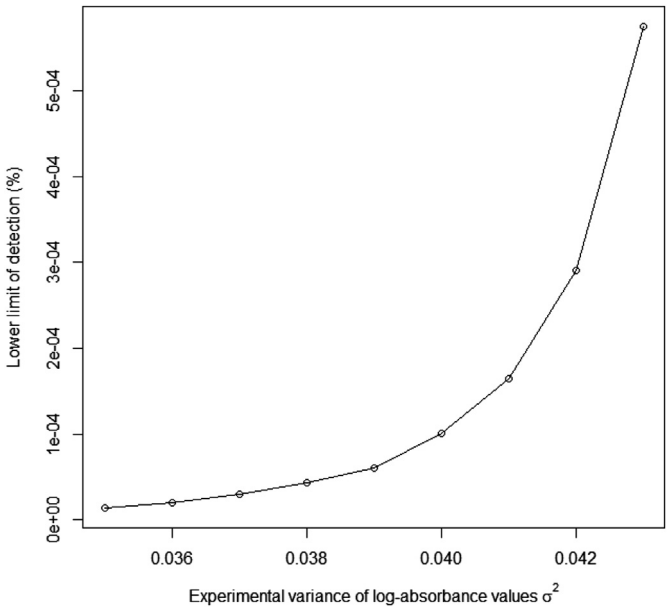

Fig. 10. Numerical simulation demonstration of the change of LOD with respect to the change of measurement variances in collecting absorbance values. Numerical time is set to be 4 h, i.e. 0.024 wk.

4. Conclusions

Egg yolk concentration in albumen can be predicted using optical absorbance and a nonlinear prediction model. The change of absorbance with yolk concentration was affected by the age of egg, egg variety, and measuring temperature. The statistical prediction model developed in this study demonstrates a methodology for the future development of an automated and in-line detector, which will include a spectroscopic measurement unit and a data processing unit employing such a model. This is an efficient and sensitive yolk detection method with the lowest detection limit ever reported.

Acknowledgment

We thank Mr. Charlwit Kulchaiyawat and Ms. Show-Ling Lee for assisting in sample preparation, Professor Hongwei Xin for advices and coordinating egg collections, and William R. Larson for providing non-commercial eggs.

Appendix. Derivation of prediction interval

In this appendix, we derive the asymptotic prediction interval for $\widehat{c}_{\mathrm{pred}}$ using the Delta method. Denote $(\widehat{\Sigma})$ the sample covariance matrix of $(\widehat{\beta}_0,\widehat{\beta}_1,\widehat{\beta}_2,\widehat{\beta}_3,\widehat{\alpha})$, we have

$$\widehat{\sigma}_c^2 = \widehat{\nabla c}^{\mathrm{T}}\widehat{\Sigma}\widehat{\nabla c} + \widehat{\sigma}^2_{\mathcal{A}_{\mathrm{measured}}}\widehat{c}^2_{\mathrm{pred}}\left(\widehat{\beta}_1+\widehat{\beta}_3 t^{\widehat{\alpha}}_{\mathrm{measured}}\right)^2 \qquad (A1)$$

where

$$\widehat{\nabla c} = \widehat{c}_{pred}\begin{pmatrix} 1 \\ \mathcal{A}_{\mathrm{measured}} \\ t^{\widehat{\alpha}}_{\mathrm{measured}} \\ \mathcal{A}_{\mathrm{measured}} t^{\widehat{\alpha}}_{\mathrm{measured}} \\ \left(\mathcal{A}_{\mathrm{measured}}\widehat{\beta}_3+\widehat{\beta}_2\right)t^{\widehat{\alpha}}_{\mathrm{measured}}\ln t \end{pmatrix} \qquad (A2)$$

Quantity $\widehat{\sigma}^2_{\mathcal{A}_{\mathrm{measured}}} \ge 0$ is the estimate of the measurement variation of newly collected absorbance values at storage time t_{measured}. The $(1-\alpha)$% prediction interval is therefore, asymptotically,

$$\widehat{c}(\mathcal{A}_{\mathrm{measured}}, t_{\mathrm{measured}}) \mp Z_{\alpha/2}\widehat{\sigma_c}$$

In practice, the procedure to predict unknown contamination level is:

1. Obtain absorbance values to estimate $\mathcal{A}_{\mathrm{measured}}$ and its variance $\widehat{\sigma}^2_{\mathcal{A}_{\mathrm{measured}}}$;
2. Plug $\mathcal{A}_{\mathrm{measured}}$, t_{measured} into Equation (1) to get $\widehat{c}(\mathcal{A}_{\mathrm{measured}}, t_{\mathrm{measured}})$ using Table 1;
3. Plug $\mathcal{A}_{\mathrm{measured}}$, t_{measured}, and estimates from Table 1 into Equation (A.2) to get $\widehat{\nabla c}$;
4. Obtain $\widehat{\sigma}_c^2$ using $\widehat{\nabla c}$, $\widehat{\Sigma}$ (reported in appendix) and $\widehat{\sigma}^2_{\mathcal{A}_{\mathrm{measured}}}$;
5. Report $\widehat{c}(\mathcal{A}_{\mathrm{measured}}, t_{\mathrm{measured}}) \mp Z_{\alpha/2}\widehat{\sigma_c}$ as the prediction interval, and $\widehat{c}(\mathcal{A}_{\mathrm{measured}}, t_{\mathrm{measured}})$ as the prediction value.

The estimated covariance matrix of parameters in Equation (A.1) is:

$$\widehat{\Sigma} = \begin{pmatrix} 0.0244 & 0.0045 & -0.0247 & -0.0048 & 0.0067 \\ 0.0045 & 0.0074 & -0.0029 & -0.0041 & -0.0002 \\ -0.0247 & -0.0029 & 0.0381 & 0.0073 & -0.0122 \\ -0.0048 & -0.0041 & 0.0073 & 0.0046 & -0.0011 \\ 0.0067 & -0.0002 & -0.0122 & -0.0011 & 0.0060 \end{pmatrix}$$

References

Abdel-Nour, N., Ngadi, M., Prasher, S., & Karimi, Y. (2011). Prediction of egg freshness and albumen quality using visible/near infrared spectroscopy. *Food Bioprocess Technology, 4*, 731–736.

Alleoni, A., & Antunes, A. J. (2004). Albumen foam stability and s-ovalbumin contents in eggs coated with whey protein concentrate. *Brazillian Journal of Poultry Science, 6*, 105–110.

American Egg Board. (2013). http://www.aeb.org/egg-industry/industry-facts/shell-egg-distribution. Accessed September 2013.

American Egg Board. (2014) http://www.aeb.org/food-manufacturers/all-about-egg-products/processing-handling-a-storage. Accessed February 2014.

Birth, G. S., Dull, G. G., Renfore, W. T., & Kays, S. J. (1985). Nondestructive spectrometric determination of dry matter in onions. *Journal of the American Society for Horticultural Science, 110*, 297–303.

Brown, P. J., Fearn, T., & Vannucci, M. (1999). The choice of variables in multivariate regression: a non-conjugate Bayesian decision theory approach. *Biometrika, 86*, 635–648.

Brown, P. J., Vannucci, M., & Fearn, T. (1998). Bayesian wavelength selection in multicomponent analysis. *Journal of Chemometrics, 12*, 173–182.

Food and Drug Administration. (2009). *Prevention of Salmonella Enteritidis in shell eggs during production, storage and transportation*. http://www.fda.gov/food/guidanceregulation/guidancedocumentsregulatoryinformation/eggs/ucm285101.htm Accessed December 2013.

Hastie, T., Tbshirani, T., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. In *Springer Series in Statistics*. New York: Springer.

Kemps, B. J., De Ketelaere, B., Bamelis, F. R., Mertens, K., Decuypere, E. M., De Baerdemaeker, J. G., et al. (2007). Albumen freshness assessment by combining visible near-infrared transmission and low-resolution proton nuclear magnetic resonance spectroscopy. *Poultry Science, 86*, 752–759.

Kobayashi, T., Kato, I., Ohmiya, K., & Shimizu, S. (1980). Recovery of foam stability of yolk-contaminated egg white by immobilized lipase. *Agricultural and Biological Chemistry, 44*, 413–418.

Lehmann, E. L., & Romano, J. P. (2008). *Testing statistical hypotheses*. In *Springer Series in Statistics*. New York: Springer.

Li-Chan, E., & Nakai, S. (1989). Biochemical basis for the properties of egg white. *Critical Review in Poultry Biology, 2*, 21–58.

Li-Chan, E. C. Y., Powrie, W. D., & Nakai, S. (1995). The chemistry of eggs and egg products. In W. J. Stadelman, & O. J. Cotterill (Eds.), *Egg science and technology* (4th ed.) (pp. 105–175). New York: Food Products Press.

Liu, M., Yao, L., Wang, T., Li, J., & Yu, C. (2014). Rapid determination of egg yolk contamination in egg white by VIS spectroscopy. *Journal of Food Engineering, 124C*, 117–121.

Liu, Y., Ying, Y., Ouyang, A., & Li, Y. (2007). Measurement of internal quality in chicken eggs using visible transmittance spectroscopy technology. *Food Control, 18*, 18–22.

Lomakina, K., & Míková, K. (2006). A Study of the factors affecting the foaming properties of egg white– a review. *Czech Journal of Food Science, 24*, 110–118.

Macherey, L. N. (2007). *Using lipase to improve the functional properties of yolk-contaminated egg whites*. Thesis. Virginia Polytechni Institute and State University.

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). In *Chapman & Hall/CRC Monographs on Statistics & Applied Probability (Book 37)* Boca Raton, Florida.

Moba. (2013). *Egg breaking systems*. In *http://www.moba.net/page/en/Processing/Moba-Egg-Breakers/EBS-Breakers Accessed September 2013.*

Morris, M. D. (2010). *Design of experiments: An introduction based on linear models*. In *Chapman & Hall/CRC Texts in statistical science*. Boca Raton, U.S.

Nielsen, H. (2000). Application of chemical methods to the determination of egg yolk contamination in commercial productions of egg white compared to enzymatic determination. *LWT – Food Science and Technology, 33*, 151–154.

Omana, D. A., Liang, Y., Kav, N. N. V., & Wu, J. (2011). Proteomic analysis of egg white proteins during storage. *Proteomics, 11*, 144–153.

Osborner, D. S., Jordan, R. B., & Kunnemeyer, R. (1997). Method of wavelength selection for partial least squares. *Analyst, 122*, 1531–1537.

Pickering, J. W. (1992). Optical property changes as a result of protein denature in albumen and yolk. *Journal of Photochemistry and Phtobiology B: Biology, 16*, 101–111.

Robert, C. P., & Casella, G. (2005). *Monte Carlo statistical methods*. In *Springer Series in Statistics*. New York: Springer.

Samli, H. E., Agma, A., & Senkoylu, N. (2005). Effects of storage time and temperature on egg quality in old laying hens. *Journal of Applied Poultry Research, 14*, 548–553.

Scott, T. A., & Silversides, F. G. (2000). The effect of storage and strain of hen on egg quality. *Poultry Science, 79*, 1725–1729.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. In *Wiley Series in Probability and Statistics*. New York.

Silversides, F. G., & Scott, T. A. (2001). Effect of storage and layer age on quality of eggs from two lines of hens. *Poultry Science, 80*, 1240–1245.

Slaughter, D. C., & Crisosto, C. H. (1998). Non-destructive internal quality assessment of kiwifruit using near-infrared spectroscopy. *Seminars in Food Analysis, 3*, 131–140.

St. John, J. L., & Flor, I. H. (1931). A study of whipping and coagulation of eggs of varying quality. *Poultry Science, 10*, 71–82.

Stadelman, W., & Cotterill, O. J. (1995). *Egg science and technology* (4th ed.). New York: Food Products Press.

Tadesse, M. G., Vannucci, M., & Lio, P. (2004). Idenfification of DNA regulatory motifs using Bayesian variable selection. *Bioinformatics, 20*, 2553–2561.

U.S. Department of Agriculture. (2000). *Egg-grading manual (AH-75)*. http://www.ams.usda.gov/AMSv1.0/getfile?dDocName=STELDEV3004502 Accessed May 2013.

U.S. Department of Agriculture Economic Research Service. (2012). *Retail data for beef, pork, poultry cuts, eggs, and dairy products*. http://www.ers.usda.gov/Data/MeatPriceSpreads/ Accessed May 2013.

Vardeman, S. B., & Jobe, J. M. (2000). *Basic engineering data collection and analysis*. Pacific Grove, U.S: Duxbury Thomson Learning.

Wang, G., & Wang, T. (2009). Effects of yolk contamination, shearing, and heating on foaming properties of fresh egg white. *Journal of Food Science, 74*, C147–C156.

Zayas, J. F. (1997). *Functionality of proteins in foods*. New York: Springer-Verlag Berlin Heidelberg.