In ovo sexing of eggs from brown breeds with a gender-specific color using visible-near-infrared spectroscopy: effect of incubation day and measurement configuration

Matthias Corion ,* Janos Keresztes,[1] Bart De Ketelaere , and Wouters Saeys

KU Leuven Department of Biosystems, MeBioS, 3001 Leuven, Belgium

ABSTRACT The culling of day-old male chicks is an animal welfare issue within the laying hen industry that raises substantial ethical concern. Alternative methods are sought to pre-select males during embryonic development. This method is called in ovo sexing and allows more humane male culling. A robust and non-invasive in ovo color sexing technique was developed and validated in this research. To this end, visible-near-infrared (**vis-NIR**) point spectroscopy was used, which has advantages over state-of-the-art hyperspectral imaging in terms of accuracy and cost. Two independent experiments were each conducted on a batch of 600 *Isa Brown* eggs. These eggs were individually illuminated on d 8 to 14, and d 18 of incubation by a halogen lamp and the signal was measured in the vis-NIR range from 300 to 1,145 nm. Next, optimal preprocessing strategies were applied and partial least squares discriminant analysis (**PLS-DA**) models were built and further optimized after performing a forward interval partial least squares (**FiPLS**) for variable selection. Results demonstrated that d 12 is too early for vis-NIR in ovo sexing, resulting in a prediction accuracy of 86.49%. However, after 13 d of incubation, an accuracy of 97.78% was obtained, increasing to 99.52% on d 14. Furthermore, these accuracies were higher than earlier reported percentages from hyperspectral imaging and successful sexing was expedited from d 14 to d 13. Moreover, prediction improvement up to 99.05% was obtained on d 13 by correcting for the variability in eggshell properties using d 0 eggshell corrections. Applying the method on d 18 resulted in a lower accuracy of 94.62% due to stronger light attenuation by the growing embryos. Finally, a reduced spectral range of 749 to 861 nm was found to be sufficient for correct classification of 98.46% of the eggs. This paves the way for high-throughput and cost-efficient usage of smaller and cheaper spectrophotometers in commercial hatcheries.

Key words: in ovo sexing, visible-near-infrared spectroscopy, brown chicken egg, chick down color sexing

2022 Poultry Science 101:101782
https://doi.org/10.1016/j.psj.2022.101782

INTRODUCTION

The culling of millions of day-old male chicks in the laying hen industry is heavily under pressure due to ethical concerns. As table-egg laying stocks are genetically inefficient for meat production, they are separated from the female chicks and euthanized. A total of 372 million laying hens was held on European farms in 2020, implicating that an equal number of male chicks were euthanized (European Commission, 2021). These cullings are expected to rise along with the increasing consumption of eggs. In 2018, 6.9 million tons of consumption eggs were produced in the European Union and this is forecasted to increase to 7.7 million tons by 2030 (European Commission, 2018).

Alternatives to this practice are to raise these male layer chicks, to breed dual-purpose chickens or to determine the sex in the egg and prevent the hatching of males (in ovo sexing). While rearing the males for slaughter is often considered the most desirable option, both approaches following that path face economic and sustainability challenges since the cockerels require a longer rearing period and they eat more feed per kilogram of meat produced than broilers (Giersberg and Kemper, 2018; Siekmann et al., 2018). On the other hand, in ovo sexing prevents the hatch of male layers and opts for continuing the breeding of specialized commercial lines for table eggs or meat production. Therefore, it is expected that this concept will become the substitute in western countries (Reithmayer and Mußhoff, 2019).

Considering the in ovo male euthanasia, this should be performed as early as possible and preferably before

Received December 23, 2021.
Accepted February 7, 2022.

[1]Present address: Apple Inc., 1 Apple Park way, 95014 Cupertino, CA, United States

*Corresponding author matthias.corion@kuleuven.be

the embryos can experience pain. However, there is no consensus on which incubation day the embryos could feel pain. A gray zone has been defined from d 7 till d 15 in which embryos might or might not feel pain (Deutscher Bundestag, 2017). For sure, it is accepted that embryos are not able to feel pain before d 7 since the multisynaptic reflex arches have not yet closed and movements to external stimuli such as mechanical skin stimulation, heat, and cold are not yet observed (Rosenbruch, 1997). Later, these nociception reflexes can be observed, although it is not certain whether they would be perceived as pain. As the first electroencephalogram signals are observed on d 15, it is expected that embryos might be sensitive to pain from that day on (Mellor and Diesch, 2007). One potential approach to cope with this uncertainty would be to stun the embryos by applying electrical anesthesia prior to euthanasia (Zumbrink et al., 2020).

Besides the day on which in ovo sexing should be applicable regarding pain perception, other characteristics such as prediction accuracy, analysis time, cost efficiency, and obtaining good hatchability are important as well. A desirable accuracy would be 98.5% and the rate should range between 20,000 and 30,000 sexed eggs per hour (Phelps et al., 2003). Concerning the cost, incubation costs would be reduced by 50% when only female eggs would have to be incubated from d 0 to 21. When only females would be incubated for the last 3 d in the hatcher, the cost would be reduced by 14% (Seemann, 2003). Logically, the earlier the sexing can be applied, the more costs can be saved. Furthermore, the day-old chick sexing costs are estimated to range from 0.6 to 3.8 cents per animal depending on the applied sexing method (Seemann, 2003).

The two approaches that are the furthest developed toward application are extraction and subsequent analysis of allantoic fluid, and the acquisition and classification of a spectroscopic signal. Regarding the allantoic fluid analysis, an immunoassay towards the female estrone sulfate exists which is higher in females and distinguishable on d 9 (Weissmann et al., 2013). Moreover, a mass spectrometric analysis has been described of metabolites such as glucose, choline, valine, and an amino compound named (3-[(2-aminoethyl)sulfanyl] butanoic acid). The valine concentration appears to be higher in males whereas the other metabolites appear to be lower from d 9 on (Bruins, 2014; Bruins, 2017). Finally, a polymerase chain reaction analysis exists of genomic DNA for the presence of a W-specific sequence (Weigel et al., 2017). All three techniques claim to obtain good results and are applicable around d 9 of incubation. However, the most important hurdle remaining is the analysis time, as the desired screening speed of 30,000 eggs per hour during the transfer from the setter to the hatcher cannot yet be reached. On the other hand, the two proposed spectroscopic methods differ largely in their moment of applicability. Successful gender prediction around d 3.5 of the incubation has been obtained based on a Raman spectrum and fluorescence signal of an excited blood vessel (Galli et al., 2018). More specifically, the Raman bands correspond to differences in blood biochemistry between the genders and the fluorescence signal was higher for males due to the higher hemoglobin content (Galli et al., 2016). While this early gender prediction is highly desirable, the level of invasivity of this method impacts the analysis time. Concerning the analysis procedure, shells need to be windowed using a CO_2 laser, a blood vessel with a diameter larger than 100 μm needs to be brought in focus and an acquisition time of 40 s is required. Furthermore, the 91% prediction accuracy is not yet close enough to the desired 98.5% at which day-old chicks are sexed. Finally, the authors reported that hatchability was not affected.

In contrast, visible-Near Infrared (**vis-NIR**) transmittance measurements through brown eggs from layer lines with gender-specific down feather color only deliver information on the gender-related color on d 14 of incubation (Göhler et al., 2016). Day 14 is situated in the late grey zone of pain perception and the technique is only applicable on eggs from brown breeds with a gender-specific color. However, it is the only technique that is completely non-invasive since the egg is left intact when a light source illuminates the embryo and the transmitted signal is captured by a hyperspectral camera. Moreover, the high prediction accuracy of 97% and the acquisition time below one second favor a robust and fast application in commercial hatcheries (Göhler et al., 2016).

Besides spectral information, hyperspectral imaging also delivers spatial information wherein each pixel in the image consists of a broader spectrum. For in ovo sexing, a region of pixels from a single egg has to be selected to calculate an average spectrum (Göhler et al., 2016). Given the fact that only the color information of the embryo is used and light is significantly scattered within an egg, the benefit of using this spatial information is not exploited. In contrast, point spectrometers only acquire an integrated signal but have the potential to measure this with a higher signal-to-noise ratio (**SNR**) than a hyperspectral camera. Therefore, it is hypothesized here that higher prediction accuracy could be obtained at a lower cost and higher speed by using point spectroscopy instead of hyperspectral imaging.

Consequently, the objective of this study was to develop and optimize a vis-NIR point spectroscopy method for in ovo sexing of brown down colored embryos. In comparison with hyperspectral sexing, the improved point spectroscopic SNR will be further maximized to increase the sexing accuracy and to potentially expedite the day at which embryos can be successfully sexed. After applying different combinations of preprocessing techniques, the best-performing models will be built to distinguish the predominantly brown female embryos from the more yellow males. Moreover, a variable selection will be performed to select the most relevant wavelengths for discriminating the genders and narrow the spectral range, allowing the use of smaller and cheaper spectrophotometers in commercial settings. Subsequently, the best-performing model with a reduced wavelength range will be tested on a fully independent

egg batch. This type of external validation will provide an extra verification of the classification performance that can be expected in practice. Next to a transmission setup, a reflection setup will be tested to look for an additional benefit in collecting a higher signal. Furthermore, eggs were measured on d 0 to correct the acquired spectra on later days for individual egg properties such as color, shape, and structure differences, and potentially improve prediction accuracy. Finally, in ovo sexing will also be tested on d 18 of incubation, because this would allow to implement gender sorting during the transfer from setter to hatcher and to limit in ovo vaccination to the female chicks.

MATERIALS AND METHODS

Egg Handling

For each of the 2 experiments, a total of 600 *Isa Brown* incubation eggs were obtained from Vepymo n.v. (Belgium). The parent flocks were respectively 44 and 49 wk old for experiments 1 and 2 and produced chicks with gender-linked down color. The 600 eggs were numbered and incubated at 37.7 °C and 55% relative humidity (**RH**) in a *SmartCombi 41* incubator (Pas Reform, the Netherlands). Four trays with a capacity of 150 eggs each were tilted every hour over 90°. Starting from d 8, the trays were alternately taken out of the incubator daily until d 14 in experiment 1. In experiment 2, the eggs were taken out on d 11 to 14 and on d 18. On these days, the egg trays were sequentially positioned in a thermally insulated box covered with a heating blanket at 35 °C. Meanwhile, the vis-NIR spectroscopy measurements were performed on the individual eggs over these days to follow up on the discriminating signals originating from the developing down feathers and their sex-specific color (see further).

On incubation d 18, candled eggs containing living embryos were divided over individually labeled cages and were further incubated at 37.7 °C and 65% RH. After 21 d and 21 h, the hatched chicks were color sexed and were numerically labeled with a *micropore* tape around a leg. The total incubation time was slightly extended until the early morning for practical reasons since the manipulations otherwise had to take place during the night. This prolongation was not a consequence of the experiment itself since the eggs were heated during the measurements and cooling down was limited. Next, predominantly brown down feather colored chicks were identified as females while predominantly yellow chicks were identified as males. An independent vent sexing was performed as a double check by an expert (Hobo C&S bvba, Belgium). Data was only included from chicks that were confirmed for both color- and vent sexing. Both experiments were approved by the Animal Ethics Committee of the KU Leuven with project number ECD 134-2016.

Vis-NIR Spectroscopic Data Acquisition

Both a reflectance and a transmittance setup were tested in experiment 1 (Figure 1). In the reflectance setup, a light source was positioned perpendicular to a detector and an egg was individually placed in-between on its side. In the transmittance setup, the light source was positioned straight on the opposite side of the detector with an upright egg in-between with the broad end on top. This light source was a gold-coated 150 W halogen lamp operating at 15 V (model 64635 HLX, Osram, Germany), mounted in a holder with a diameter of 50 mm. Herein, the lamp's heat was blocked by a 50 mm heat-absorbing short pass filter glass of 3-mm thick (model Schott KG1, Edmund Optics Ltd., Barrington, NJ). The heat from the filter glass was removed by a cooling fan. At the end of the lamp holder, the light was concentrated by a 50-mm aspheric condenser lens of focal length 40 mm (model ACL5040U-DG15, Thorlabs Inc., Newton, NJ). From this condenser lens, light was directed toward the egg. The reflected or transmitted light was collected into an optical fiber with an inner core diameter of 650 μm using another aspheric condenser lens. Then, the fibers from both setups were, together with a third fiber for the dark signal, connected using a *Leoni M optical FiberSwitch* (Leoni Fiber Optics GmbH, Germany). From this switch, a 650-μm outlet fiber was fed to an *MCS 600* spectrometer (Zeiss, Germany). The signal was measured over the 300 to 1,145 nm wavelength range with a total of 256 wavebands and a step of 3.30 nm. A computer with a custom *LabVIEW* program (National Instruments, Austin, TX) was used to acquire the measurements. In experiment 2, only the transmittance setup was used together with a thicker inner core fiber with a 1 mm diameter to obtain a higher signal.

Dark and white reference measurements were preceding every small batch of 75 eggs (i.e., approximately every 30 min). The dark reference was measured after covering the detector (i.e., at zero light transmittance). The white reference was measured with a 25 mm fused silica metallic neutral density filter of 0.01% transmission (model NDUV40B, Thorlabs Inc.) between the light

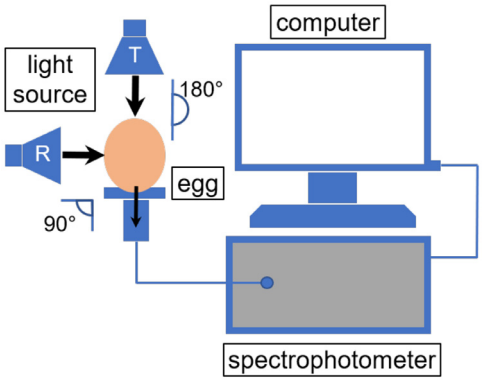


Figure 1. Visible-Near-Infrared (vis-NIR) setup wherein a light source illuminates an egg over 90° (reflectance, R) or 180° (transmittance, T). The signal is sent via an optical fiber to a spectrophotometer and a computer processes the data. For the reflectance measurements, the egg was positioned on its side.

source and the condenser lens towards the detector. An average signal over 5 cycles was captured per sample with a specific exposure time. This exposure time ranged from 500 to 3,000 ms and was daily optimized (i.e., maximized without causing saturation) depending on the egg translucence which decreased over time because of embryonic growth. Underdeveloped embryos or infertile eggs were not taken into account when optimizing the exposure time. An overview of the used exposure times is presented in Figure 2. In experiment 1, detector saturation took place around 650 to 675 ms for the white reference. In experiment 2, the white reference was measured with a shorter exposure time of 100 ms due to early saturation of the detector caused by the thicker fiber diameter. Afterward, the white reference signal was linearly multiplied by a factor corresponding to the exposure time used on the eggs on that specific day. Finally, the eggs in experiment 1 were measured from d 8 till 14. In experiment 2, the eggs were measured on d 0, d 11 to 14, and d 18.

Data Analysis

The individual raw signal acquired for the ith egg in the jth batch on day D is referred to as $X_{egg_{i,j},\ D}$. The index i was ranging from egg number 1 to 600, whereas index j was ranging from batch 1 to 6 with 75 eggs per batch. The signal was normalized based on the signals for the corresponding dark ($X_{dark_j,D}$) and white ($X_{white_j,D}$) reference measurements. The calculation for this corrected signal $X^c_{egg_{i,j},D}$ is as follows (Equation 1):

$$X^c_{egg_{i,j},D} = \frac{X_{egg_{i,j},D} - X_{dark_j,D}}{X_{white_j,D} - X_{dark_j,D}} \quad (1)$$

The SNR was calculated using the white reference measurements to check for signal quality and to quantify the effect of the setup improvements from experiment 1 to experiment 2. This was calculated by dividing the average white reference per day ($\overline{X}_{white,D}$) by its standard deviation per day ($\sigma_{white,D}$; Equation 2):

$$SNR = \frac{\overline{X}_{white,D}}{\sigma_{white,D}} \quad (2)$$

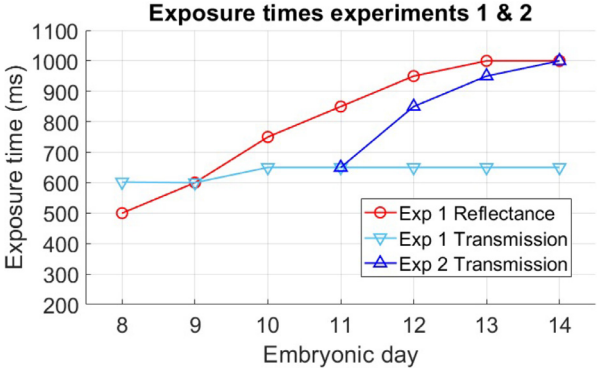


Figure 2. Exposure times used in experiments 1 and 2 as a function of embryonic day. The exposure time on embryonic d 18 in experiment 2 was 3,000 ms (not shown in the figure).

Next, the signals were cropped to a wavelength range from 402 to 1,101 nm since the information outside this range was not useful and showed low SNR. Finally, the dataset comprising the spectral data from all measurement days was randomized and split into a calibration set (60%) and a prediction set (40%).

Furthermore, the egg spectra were preprocessed to remove structured noise from the data and to accentuate small structural differences between nearly identical spectra. The methods considered in this work were: second-order Savitzky-Golay smoothing with a window of 9 or 15 wavebands (i.e., an interval of 29.7 or 49.5 nm), first and second-order derivatives, standard normal variate (**SNV**), autoscaling to a zero mean and unit variance, mean centering, subtraction of a baseline offset (detrending), multiplicative scatter correction with a mean or median as reference spectrum and normalization to the absolute value of the area under the curve or the maximum value in the row (spectrum). More information on these methods can be found in (Saeys et al., 2019). These methods were individually tested or in combination with each other up to 3 different preprocessing steps in *Matlab* (Matlab version 2018b, The Mathworks, Inc., Natick, MA) using the *PLS toolbox* (version 8.7 2019, Eigenvector Research, Inc., Wenatchee, WA). An overview of the different combinations can be found in Table S1 in the supplementary data A contiguous blocks cross-validation method was used with 10 data splits and partial least squares discriminant analysis (**PLS-DA**) models were built to check the prediction accuracy in cross-validation.

Next, the best three preprocessing methods per day were selected based on their prediction accuracy. For the preprocessing methods using Savitzky-Golay smoothing and derivatives, the data was further cropped to the range from 569 to 868 nm to remove noise from the spectral edges. The PLS-DA models were manually built using the PLS toolbox's graphical user interface. For these models, the number of latent variables (**LVs**) was selected based on the relative variance explained by the components in a scree plot and the minimum after which the improvement became marginal in the root mean square error plot of the cross-validation set (**RMSECV**).

Before continuing to model testing, an outlier analysis was performed by evaluating the Q residuals for checking the remaining variance after projection onto the model and the Hotelling T^2 values for checking the variation in each sample within the model. Toward variable selection and model improvement, a forward interval partial least squares (**FiPLS**) was applied on the cropped spectral range. Here, the number of intervals was chosen automatically for a minimal RMSECV and the wavebands per interval were set between 2 and 8. Outliers that were visually distinct from other observations in the preprocessed data plot and on the Q residuals vs. Hotelling T^2 plot were excluded from the model (i.e., these chicks were not included for calculating the in ovo sex determination accuracy).

The performance of the models was validated on the internal prediction set and the models with the highest

prediction accuracy were selected. Finally, the selectivity ratio was plotted to check for the most relevant wavelength ranges in the model. The latter is the ratio between explained and residual variance of a variable and thus indicates its importance in the classification (Rajalahti et al., 2009).

This procedure was followed for experiments 1 and 2. During experiment 2, we also investigated the added value of a preincubation measurement of each egg on d 0 to serve as a correction factor. This signal is denoted by $X^{c}_{egg_{i,j},0}$. We then proceeded by correcting each signal by first smoothing it using a Savitzky-Golay filter (**SavGol**) with a window of 9 wavebands and apolynomial order of 2 before calculating the normalized spectrum as follows (Equation 3):

$$X^{c,D_0}_{egg_{i,j},D} = \frac{SavGol\left(X^{c}_{egg_{i,j},D}\right)}{SavGol\left(X^{c}_{egg_{i,j},0}\right)} \quad (3)$$

Since the noise of the d 0-corrected data increased at the edges of the spectra, they were cropped to a narrower range from 536 to 966 nm.

RESULTS AND DISCUSSION

Hatching Results

From the 600 eggs that were incubated in each experiment, experiment 1 had a total hatchability of 86.3% with 2.8% infertile eggs and 10.8% dead embryos. Subsequently, the hatchability of living embryos was 88.9%. In experiment 2, the hatchability was 86.2% with 2.2% infertile eggs and 11.7% dead embryos. The resulting hatchability of living embryos was 88.1%. Due to a mismatch in color- and vent sexing, a few chicks were excluded from the dataset and a total of respectively 514 and 511 animals were used for the data analysis in experiments 1 and 2. With an average hatchability of living embryos of more than 88%, it was assumed that the small incubation disturbances during measurements had minor effects on the hatchery results.

Preprocessing Approach

Generally, it was found that a 3-fold preprocessing in which a standard normal variate was performed, followed by a second-order Savitzky-Golay smoothing with a window of 9 wavebands and a second derivative, finished by a mean centering provided the best results over almost all days. This 3-fold preprocessing method was applied to both reflectance and transmittance spectra to a cropped wavelength range from 569 to 868 nm to remove the noisy wavelengths at the edges. Better performances of other methods on a specific day were marginal and the use of method 39 for all days allowed for a more uniform comparison. It is hypothesized that the superiority of this preprocessing approach follows from the removal of baseline shifts due to light scattering by the SNV and an accentuation of small structural differences by the smoothing in combination with the second-order derivatives. Finally, the cropping of the spectra to this specific range improved the performance of the PLS-DA models in combination with the FiPLS to find relevant variables and obtain higher prediction accuracies.

In Figure 3A, normalized spectra of transmittance data on day 14 in experiment 1 are visualized and it is already noticed that females generally absorb more light than males. When the preprocessing is performed, distinguishable grouping structures around certain wavelengths are revealed (Figure 3B).

Incubation Experiment 1

Reflectance vs. Transmittance First, a comparison was made between the reflectance and transmittance setups in experiment 1 by comparing the best PLS-DA models on d 8 to 14. In Figure 4, the preprocessed signals of both setups are presented for d 8 and 14. Concerning the day differences, gender groupings are not visually separable on d 8, while this is the case on d 14. Furthermore, the reflectance spectra carry more unstructured noise below 700 nm than the transmittance spectra on d 14. This is observed in Figure 4B where there is no visual

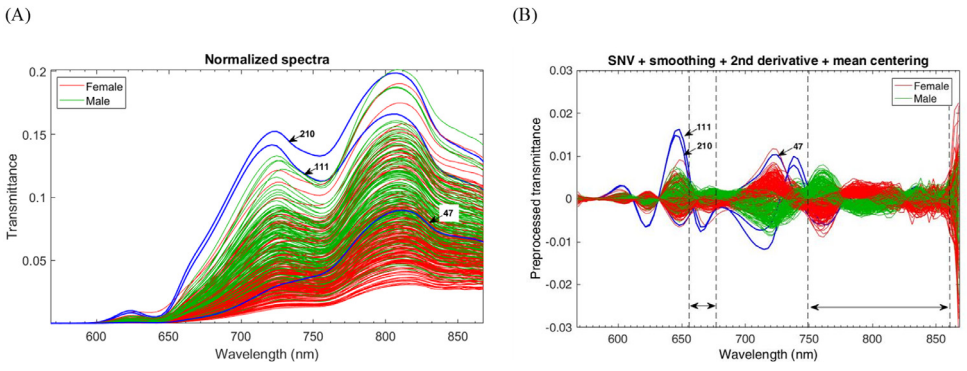


Figure 3. Transmission spectra of the calibration set on d 14 of incubation in experiment 1 in which three male outliers based on the Hotelling T^2 and Q residuals plot are indicated in blue and females and males are respectively colored in red and green. (A) Transmittance spectra normalized for the white and the dark reference. (B) Preprocessed transmittance signals of (A) already indicating a gender-specific grouping on both sides of the horizontal axis. Dashed lines represent the two Forward interval Partial Least Squares selected wavelength intervals.

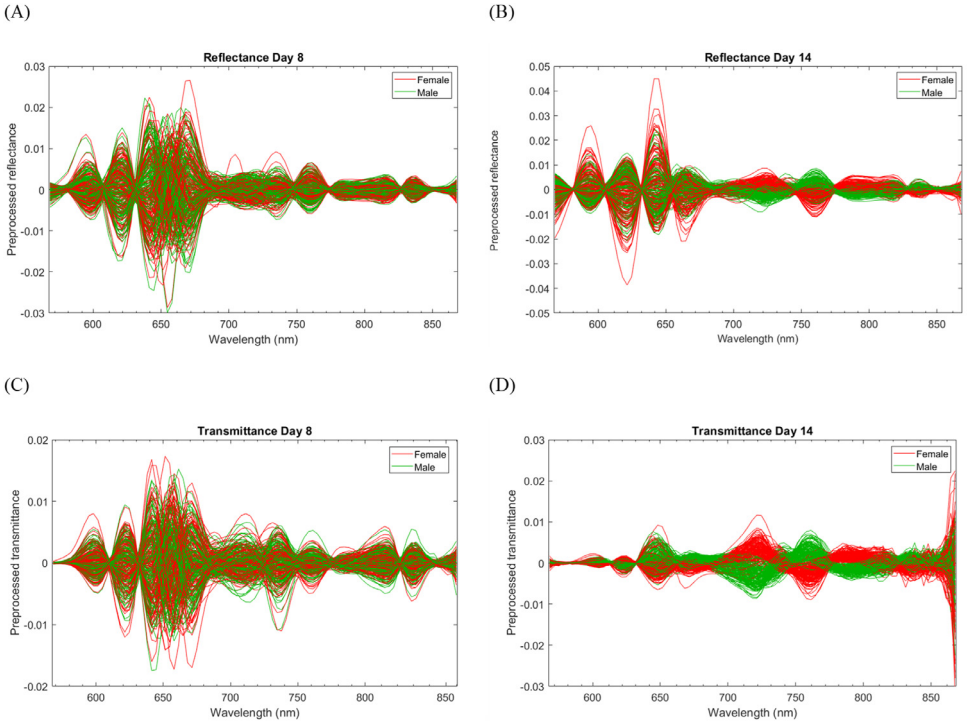


Figure 4. Reflectance and transmittance spectra of the calibration set in experiment 1 preprocessed after applying a Standard Normal Variate (SNV) followed by a second-order Savitzky-Golay smoothing with a window of 9 wavebands and a second derivative, finished by a mean centering. (A, B) Reflectance on respectively d 8 and 14 whereby a clearer gender-specific grouping around the horizontal axis is observed on d 14. (C, D) Transmittance on respectively d 8 and 14 whereby a gender-specific grouping is visible for d 14 and signals show less noise in the shorter wavelength range than the preprocessed reflectance spectra.

gender separation in the groupings below 700 nm. On the other hand, the groupings in Figure 4D are more subtle in the transmittance data in which distinguishable groupings appear between 650 and 700 nm.

An overview of the different PLS-DA models from d 8 to 14 is presented in the supplementary data in Table S2. Concerning the prediction accuracies, the best results were obtained on d 14. Subsequently, PLS-DA models were also built separately on the reflectance and the transmittance spectra and were compared with the combined approach. On d 14, a prediction accuracy of 98.33% was obtained for the 2 setups combined, while 96.36% was obtained for reflection only and 99.01% was obtained for transmission only. Noticeably, the FiPLS selected only variables from the transmission spectra in the combined reflectance and transmittance setup. Presumably, the reflectance data contained less relevant information than the transmittance data. Furthermore, it may be assumed that the signal passage through the egg for the transmission measurement carried more information than the internal volume reflection of the egg. Considering this, it was concluded that the transmittance setup captured the most relevant information from the embryos and carried the potential for single use. Subsequently, only the transmission data were used for further model optimizations in both experiments 1 and 2. Furthermore, transmission measurement has more potential for direct integration in existing hatchery equipment since the light source can be positioned above the trays and the detector below or vice versa.

Optimization of the Prediction Accuracies Per Day Using Transmittance Spectra Following the conclusion that the transmittance setup captured the most relevant information, PLS-DA models were built on d 8 to 14. The prediction accuracies are presented in Figure 7 and reference is made to Table S2 in the supplementary data for more details. On d 8, 9, and 10, the best prediction accuracies were respectively 52.04, 52.97, and 61.90%. These results demonstrated that the vis-NIR technique is not applicable for gender-determining embryos at these stages, as the prediction accuracy obtained by randomly guessing the gender would be 50%. Furthermore, the highest accuracies on d 11 and 12 were respectively 64.41 and 86.49%. On d 13, more desirable results were obtained with a prediction accuracy of 97.78% for a model involving 8 LVs based on the wavelength ranges 572 to 703 nm, 732 to 835 nm, and 865 to 868 nm. In the selectivity ratio plot (results not presented) a relatively high weight was attributed to the wavelength of 782 nm. Consequently, given that lower prediction accuracies of 71.36 and 88.61% were obtained on respectively d 12 and 13 using hyperspectral imaging (Göhler et al., 2016), these results confirmed the hypothesis that higher prediction accuracies can be obtained using point spectroscopy. Moreover, successful sexing was advanced from d 14 to d 13.

Similar to the combined reflectance and transmittance models, the best results were obtained on d 14 with a prediction accuracy of 99.01%. The variable selection with FiPLS resulted in the selected wavelength ranges of 656 to 676 nm and 749 to 861 nm. These wavelength intervals are indicated in Figure 3B by dashed lines. A new PLS-DA model was finally built on these selected wavelength ranges with only 4 LVs. No misclassification

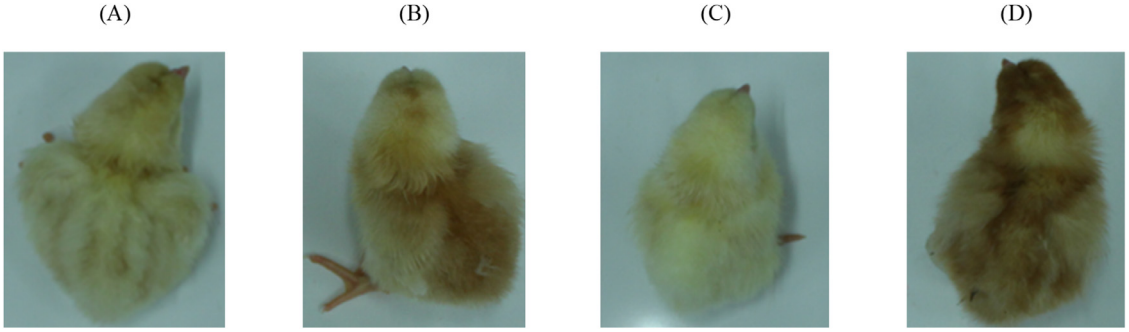


Figure 5. Pictures of hatched chicks in experiment 1 to demonstrate that the misclassified chicks had colors that resembled the down feather color of the opposite gender. (A) Misclassified male with brown down feather in the neck and stripes on the back. (B) Misclassified female with yellow down on the head and light brown on the back. (C) Example male with predominantly yellow down feathers. (D) Example female with predominantly brown down feathers

was observed in the calibration or cross-validation set. The prediction accuracy was 99.01% for the prediction set with a sensitivity and specificity for the females of respectively 98.90 and 99.10%. These sensitivity and specificity were higher than respectively the 96.72 and 97.46% by Göhler et al. (2016). From our prediction set, only 2 observations out of 206 were misclassified. The first was a male with brown down feathers in the neck and stripes on the back, while the second was a female with yellow down feathers on the head and light brown on the back. Pictures of these misclassified observations together with clearly distinguishable example chicks are presented in Figure 5. Thus, it does not seem unlikely that these 2 individuals were misclassified by the algorithm because they were phenotypically similar to the other sex.

In order to obtain a better understanding of the relevant wavelengths, the selectivity ratio for this model was plotted and is presented in Figure 6A. It was observed that no significant added value was allocated to the first wavelength range (i.e., in the visible part of the spectrum). Subsequently, a new PLS-DA model with 4 LVs was built on the second wavelength range only. Three important wavelength peaks, 769, 792, and 828 nm, are indicated with arrows in the selectivity ratio plot in Figure 6B. In comparison with the previous model, a loading shift from 759 to 769 nm was observed and wavelength 792 nm lowered in importance to a similar level of the wavelength 769 nm. This new model had a slightly lower prediction accuracy of 98.46% (one extra misclassification). In practice, this prediction accuracy could still be sufficient and offer the potential of using smaller and cheaper spectrophotometers for the narrower wavelength range of 749 to 861 nm.

For in ovo sexing on d 13, a relatively high weight was attributed to the wavelength of 782 nm which is close to the wavelength of 792 nm identified for the d 14 model. It should be noted that it is more relevant to interpret the range around the selected wavelength as important rather than the individual wavelength since a second derivative was applied to the spectra. Furthermore, more weight was attributed to the near-infrared (**NIR**) wavelengths (longer than 700 nm) in comparison to the visible wavelengths (shorter than 700 nm). It is, therefore, suggested that the females with brown pigments absorb more light in the NIR region than the predominantly yellow males. These brown pigments are called 'eumelanins' and the yellow pigments are 'phaeomelanins'. Melanins have a wide spectral absorbance since they have a relatively high degree of conjugation in the molecule. The eumelanins are darker and have a stronger absorbance towards the red part of the spectrum (Riley, 1997). Female chicks have more eumelanins than male chicks and it was suggested that the eumelanins

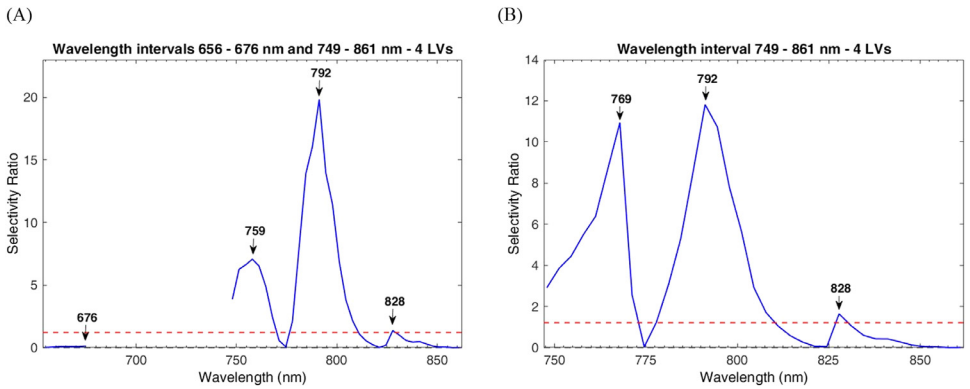


Figure 6. Selectivity ratio for determining the loading of wavelength variables of a partial least squares − discriminant analysis (PLS-DA) model with 4 latent variables (LVs). The peaks are indicated with an arrow and labeled with the corresponding wavelength. (A) Ratios of the two wavelength intervals. (B) New ratios after excluding the first interval.

caused the higher absorption in the 782 to 792 nm region in females.

Finally, the prediction accuracy started to increase on d 11 from 64.41 to 86.49% on d 12. However, these accuracies were significantly lower than the percentages obtained on d 13 and 14. Although the embryos were not absorbing all the light and the eggs were more translucent than on d 13 and 14, the feathers were not yet covering the full body and did not provide enough discrimination between the 2 genders. As described by Göhler et al. (2016), the feathers start to grow on d 11 and cover the body gradually during d 12 and 13. These researchers managed to make a clear distinction between embryos on d 14 based on the brown pigment differences in feather color. It was expected in this study that the signal increment at the detector in experiment 2 would provide increased discriminative power for the gender differences from d 11 to 14. These results are further elaborated in the following paragraphs. Finally, an overview of the best sexing models for d 13, 14, and 18 is presented in Table 1.

Incubation Experiment 2

Prediction Accuracy Optimization Continuing on the transmittance setup in experiment 1, a thicker optical fiber with a 1 mm diameter and a longer exposure time during measurements were used to gain more signal and obtain a higher discriminative power on d 11 to 14. Subsequently, the *SNR* in the 749 to 861 nm region was improved in experiment 2 with a factor of 1.5. For this comparison, the white references from d 11 were used since these were measured at the same exposure time in both experiments (Figure 2).

PLS-DA models were created with respectively 5 and 6 LVs for d 13 and 14 on the dataset of experiment 2 (Table 1). Similar to experiment 1, a prediction accuracy of 97.53% was obtained on d 13. Furthermore, the prediction accuracy on d 14 increased to 99.52%. Finally, the highest weight was attributed to wavelength 786 nm on both d 13 and 14 based on the selectivity ratio (results not presented). It is hypothesized that the increased SNR of the setup in experiment 2 provided more wavelength ranges with discriminating signals leading to the inclusion of more relevant wavelengths in the model. As a result, the PLS-DA models on the full cropped range provided better prediction accuracies than the models on which a FiPLS was applied. Finally, the highest possible prediction accuracies on d 11 and 12 were respectively 68.93 and 86.29% and are presented in overview Figure 7. Given that no noticeable increment in the prediction accuracy for d 12 was observed relative to the first experiment, it can be concluded that d 13 is the earliest day at which reliable gender discrimination can be obtained using vis-NIR spectroscopy.

Day 18 Prediction On d 18, measurements were performed preceding the transfer to the hatcher. The aim was to verify whether this technique works with comparable results relative to d 14. Regarding commercial application, the incorporation of an in-ovo sexing machine in existing candling and transfer systems is expected to be simpler and could provide more efficient use of hatchers and application of in ovo vaccinating machinery. However, this vision excludes the fact that pain sensation is then further developed together with the embryo (Deutscher Bundestag, 2017). Based on the measurements on d 18, a prediction accuracy of 94.62% was obtained (Table 1). Although the feathers of the chick are fully covering the body on d 18, a lower discriminative power was observed. This can be attributed to the relatively low signal reaching the detector since the almost completely developed chick was absorbing a major part of the light.

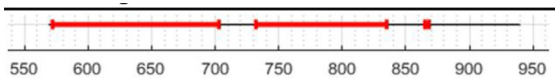
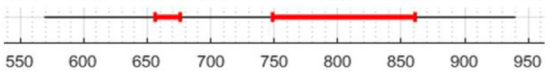
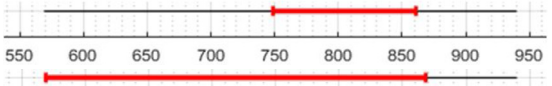
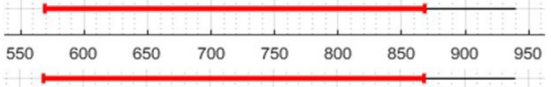
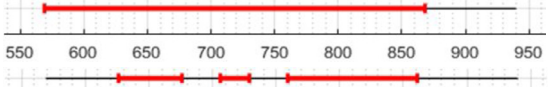
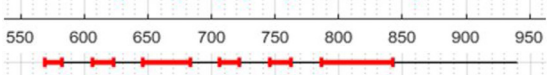
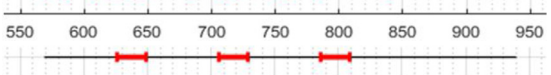
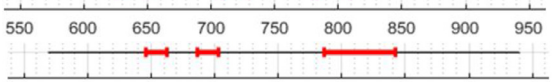
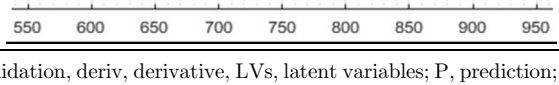
Correction for Individual Egg Variation In experiment 2, spectra were also acquired on d 0 to normalize the spectra for the large variability in eggshell properties such as color, shape, and structural differences. The most optimal preprocessing technique was to further crop the normalized spectra to the relevant range from 569 to 868 nm and to apply a mean centering. While no accuracy improvements were observed with this approach for d 11, 12 (results not shown), and 14, improvements were found for d 13 and 18. Here, the most optimal models delivered prediction accuracies of respectively 99.05 and 95.01% (Table 1), which corresponds to improvements of respectively 1.52 and 0.39%. Although the models were more complex with 8 LVs relative to 5 LVs, especially for d 13, the accuracy increased above 99% (Table 1). Regarding the application of the vis-NIR technique in practice, the d 0 measurement seems to be worth considering as it allowed improving the sexing accuracy on d 13, while the added value on d 18 is only marginal.

Model Robustness Verification

The model robustness was verified by testing the d 14 PLS-DA model with the reduced second wavelength interval from experiment 1 on all the data acquired on d 14 in experiment 2. As a result, 97.50% prediction accuracy was obtained. With this model, all females in the dataset of experiment 2 were correctly classified and 5% of the males were misclassified. This skewed misclassification was caused by the usage of the calibration set's mean spectrum in the mean centering. Given the small adaptations to the measurement setup prior to experiment 2, such as the thicker fiber inner core diameter and the longer exposure times, the preprocessed spectra of the prediction set were not symmetrically located around the mean spectrum of the calibration set from experiment 1 (Figure 8A). Nevertheless, the results were still satisfying since this shift was limited thanks to the different preprocessing steps preceding the mean centering. This demonstrates the possible transferability of the preprocessing method for limited differences between setups.

Although the approach of using the mean spectrum of the calibration set is the usual way of applying a

Table 1. Overview of PLS-DA models for gender identification.

Experiment	Setup	Preprocessing steps	Day	Variable selection	Wavelength intervals	LVs	Acc_C	Acc_CV	Acc_P	Sen_C	Spe_C	Sen_CV	Spe_CV	Sen_P	Spe_P
1	T	SNV + SavGol order 2 & 2nd deriv, window 9 + mean center	13	10 intervals & 8 variables per interval		8	0.9967	0.9967	0.9778	0.9930	1.0000	0.9930	1.0000	0.9560	1.0000
			14	6 intervals & 7 variables per interval		4	1.0000	1.0000	0.9901	1.0000	1.0000	1.0000	1.0000	0.9890	0.9910
			14	Day 14 model without first interval		4	0.9934	0.9901	0.9846	0.9870	1.0000	0.9870	0.9940	0.9780	0.9910
2	T	SNV + SavGol order 2 & 2nd deriv, window 9 + mean center	13	Full cropped range		5	0.9841	0.9740	0.9753	0.9930	0.9750	0.9790	0.9690	0.9810	0.9700
			14	Full cropped range		6	0.9969	0.9969	0.9952	1.0000	0.9940	1.0000	0.9940	0.9900	1.0000
			18	7 intervals & 7 variables per interval		5	0.9283	0.9251	0.9462	0.9320	0.9250	0.9320	0.9180	0.9430	0.9490
2	EC_T	Mean centering	13	9 intervals & 6 variables per interval		8	0.9903	0.9903	0.9905	0.9930	0.9880	0.9930	0.9880	0.9810	1.0000
			14	3 intervals & 8 variables per interval		8	0.9968	0.9968	0.9905	1.0000	0.9940	1.0000	0.9940	0.9810	1.0000
			18	FiPLS 8 intervals & 8 variables interval size		8	0.9381	0.9381	0.9501	0.9590	0.9170	0.9590	0.9170	0.9810	0.9190

Abbreviations: Acc, accuracy; C, calibration; CV, cross-validation, deriv, derivative, LVs, latent variables; P, prediction; SavGol, Savitzky-Golay smoothing; SNV, Standard Normal Variate, Sen; sensitivity for females; Spe, specificity for females; T, transmission.

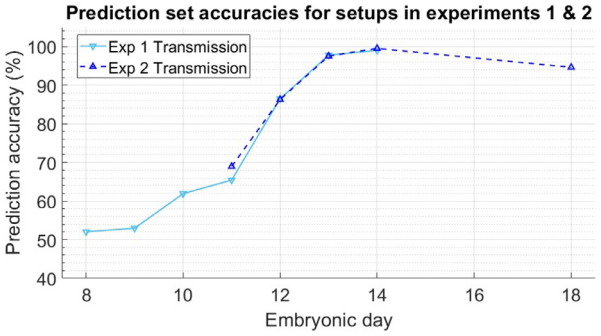


Figure 7. Prediction set accuracies for transmission setups in experiments 1 and 2. Reproducible accuracies were obtained in follow-up experiment 2 with a slightly higher percentage on d 11.

model on the prediction set, the unique situation of in ovo sexing offers another possibility to remove inter batch variation or correct for slight modifications in the setup. Despite numerous studies, the most accepted conclusion is that an equal sex ratio exists in domestic birds (Battelier et al., 2004). Conventional egg incubation trays contain 150 eggs per tray. When it is assumed that every tray of eggs contains approximately 50% of each gender, the mean spectrum of each tray can be used for the mean centering step instead of using the mean from a calibration set. In this way, the mean can be sequentially updated per scanned batch. As a final demonstration of this possibility, the mean centering was performed with the mean of the prediction set instead of using the mean spectrum calculated from the spectra included in the calibration set. Doing so, the prediction accuracy increased from 97.50 to 98.63% with a symmetrical misclassification for the genders (Figure 8B).

CONCLUSIONS

In ovo sexing using vis-NIR point spectroscopy has a high potential for non-invasive, fast, and accurate sexing of brown chicken eggs from layer lines with gender-specific down feather color. This was demonstrated in this study by performing 2 experiments in which hundreds of eggs were one-by-one placed in front of a light source and the signal was captured by a spectrophotometer. It was concluded that a transmission setup with a 180° angle between light source and detector had an optimal and reliable performance and the reflection setup was redundant.

The approach allows for a reliable gender determination on incubation days 13 and 14 with prediction accuracies of respectively 97.78 to 99.52%. This point spectroscopic approach has the advantage over a hyperspectral approach because of the increased SNR ratio due to whole egg signal integration without having resolution loss. This benefit was demonstrated by the higher prediction accuracies and the advancement of successful sexing to d 13. Furthermore, the most informative wavelengths for in ovo sexing were found in the 749 to 861 nm range, which is accessible by smaller and more affordable spectrophotometers. More specifically, the most discriminating wavelengths were located in the NIR region around 782 to 792 nm. Presumably, this is the relevant range for brown eumelanin pigments in the feathers that cause an increased absorption in females. Furthermore, the robustness of this approach was demonstrated on d 14 by correctly predicting 98.63% of the whole dataset from a different flock and moment of analysis. Gender prediction before d 13 was considerably less accurate because the feathers are not enough developed for a reliable distinction. In ovo sexing later than d 14 with this technique will be more difficult since the developing embryos absorb more light resulting in lowered transmittance. This was demonstrated by the decreased 94.62% accuracy on d 18. It was also shown that prediction improvements can be obtained by correcting for any possible variability in eggshell properties using d 0 eggshell corrections. More specifically on d 13, an increase in prediction accuracy of 1.52 to 99.05% was obtained with this approach. Moreover, as the embryos might already perceive pain around days 13 to 14, this technique could be combined with electrical stunning to improve animal welfare. Finally, given the fact that this non-invasive technique is only applicable on embryos from layer lines with sex-specific phenotypic color differences, it is suggested for other commercial breeds to apply invasive methods for biomarker analysis using

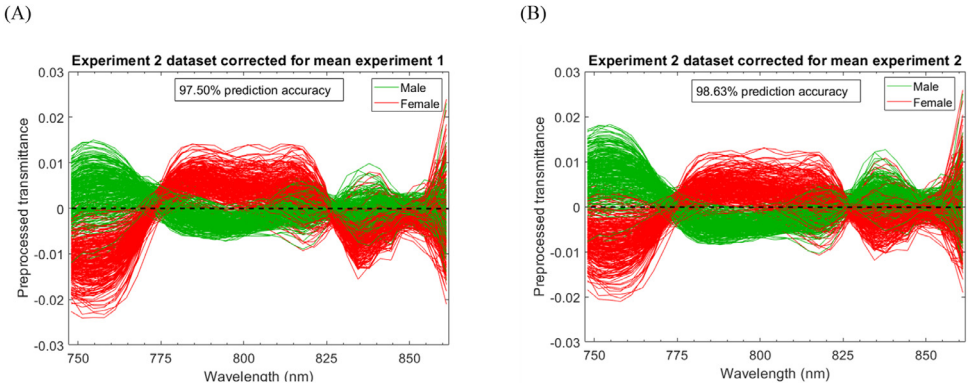


Figure 8. Demonstration of the impact of mean centering on preprocessed data of experiment 2 using a Standard Normal Variate (SNV), a second-order Savitzky-Golay smoothing with second-order derivative and window of 9, followed by a mean centering. (A) Using the mean of the calibration set of experiment 1. (B) Using the mean of experiment 2.

immunosensory, mass spectrometry, PCR, or spectroscopy.

ACKNOWLEDGMENTS

This work has received funding from the Flemish environment department [3E160053] and the Research Foundation − Flanders [SB project 1SC7219N]. Furthermore, gratitude is expressed to the KU Leuven Laboratory of Livestock physiology of prof J. Buyse for their assistance in hatchery practice.

DISCLOSURES

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

SUPPLEMENTARY MATERIALS

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.psj.2022.101782.

REFERENCES

Battelier, F., M. Govoroun, and J. P. Brillard. 2004. Sex-ratio chez les oiseaux sauvages et domestiques. INRAE Prod. Anim 17:365–372.

Deutscher Bundestag. 2017. Sachstand zum schmerzempfinden von hühnerembryonen. Wissenschaftliche Dienste 3000(075/20):5–8. https://wwwbundestag.de/resource/blob/805020/58284d172e611640db4dc17ec59d0865/WD-8-075-20-pdf-data.pdf (Accessed Jan. 2022).

Bruins, W., and W. Stutterheim. 2017. Method and system for the non-destructive in ovo determination of fowl gender. Accessed Mar. 2022. https://worldwide.espacenet.com/patent/search/family/059366469/publication/WO2017204636A2?q=pn%3DWO2017204636A2&queryLang=en%3Ade%3Afr.

Bruins, W., and W. Stutterheim. 2014. Gender, viability and/or developmental stage determination of avian embryos in ovo. Accessed Mar. 2022. https://worldwide.espacenet.com/patent/search?q=pn%3DWO2014021715A3.

European Commission. 2018. EU agricultural outlook for markets and income 2018-2030. *Directorate-General for Agriculture and Rural Development*, https://ec.europa.eu/info/sites/info/files/food-farming-fisheries/farming/documents/medium-term-outlook-2018-report_en.pdf (Accessed Aug. 2021).

European Commission. 2021. Eggs market situation. https://ec.europa.eu/info/food-farming-fisheries/animals-and-animal-products/animal-products/eggs_en (Accessed Aug. 2021).

Galli, R., G. Preusse, C. Schnabel, T. Bartels, K. Cramer, M. E. Krautwald-Junghanns, E. Koch, and G. Steiner. 2018. Sexing of chicken eggs by fluorescence and Raman spectroscopy through the shell membrane. PLoS One 13(2):e0192554.

Galli, R., G. Preusse, O. Uckermann, T. Bartels, M.-E. Krautwald-Junghanns, E. Koch, and G. Steiner. 2016. In ovo sexing of domestic chicken eggs by Raman spectroscopy. Anal. Chem. 88:8657–8663.

Giersberg, M., and N. Kemper. 2018. Rearing male layer chickens: a German perspective. Agriculture 8:176.

Göhler, D., B. Fischer, and S. Meissner. 2016. In-ovo sexing of 14-day-old chicken embryos by pattern analysis in hyperspectral images (VIS/NIR spectra): a non-destructive method for layer lines with gender-specific down feather color. Poult. Sci. 96:1–4.

Mellor, D. J., and T. J. Diesch. 2007. Birth and hatching: key events in the onset of awareness in the lamb and chick. N. Z. Vet. J. 55:51–60.

Phelps, P., A. Bhutada, S. Bryan, A. Chalker, B. Ferrell, S. Neuman, C. Ricks, H. Tran, and T. Butt. 2003. Automated identification of male layer chicks prior to hatch. Worlds Poult. Sci. J. 59:33–38.

Rajalahti, T., R. Arneberg, F. S. Berven, K. M. Myhr, R. J. Ulvik, and O. M. Kvalheim. 2009. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. Chemom. Intell. Lab. Syst. 95:35–48.

Reithmayer, C., and O. Mußhoff. 2019. Consumer preferences for alternatives to chick culling in Germany. Poult. Sci. 98:4539–4548.

Riley, P. A. 1997. Molecules in focus: melanin. Int. J. Biochem. Cell Biol. 29:1235–1239.

Rosenbruch, M. 1997. The sensitivity of chicken embryos in incubated eggs. ALTEX 14:111–113.

Saeys, W., N. Nguyen Do Trong, R. Van Beers, and B. M. Nicolaï. 2019. Multivariate calibration of spectroscopic sensors for postharvest quality evaluation: a review. Postharvest Biol. Technol. 158:110981.

Seemann, G. 2003. Organisational framework for hatcheries. Worlds Poult. Sci. J. 59:59–61.

Siekmann, L., S. Janisch, R. Wigger, J. Urban, J. Zentek, and C. Krischek. 2018. Lohmann dual: a dual-purpose chicken as an alternative to commercial broiler chicken? aspects of meat quality, lipid oxidation, shear force and muscle structure. Eur. Poult. Sci. 82.

Weigel, M. C., K. Hofmann-Peiker, and M. Kleine. 2017. Method for gender identification in domestic chicken. Accessed Mar. 2022. https://worldwide.espacenet.com/patent/search?q=pn%3DWO2017076957A1.

Weissmann, A., S. Reitemeier, A. Hahn, J. Gottschalk, and A. Einspanier. 2013. Sexing domestic chicken before hatch: a new method for in ovo gender identification. Theriogenology 80:199–205.

Zumbrink, L., B. Brenig, A. Foerster, J. Hurlin, and M. von Wenzlawowicz. 2020. Electrical anaesthesia of male chicken embryos in the second third of the incubation period in compliance with animal welfare. Eur. Poult. Sci. 84:1–11.