Computers and Electronics in Agriculture 196 (2022) 106854

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Assessing environmental control strategies in cage-free aviary housing systems: Egg production analysis and Random Forest modeling

Andrés F. Gonzalez-Mora [a,b,*], Alain N. Rousseau [a], Araceli D. Larios [b,c,d], Stéphane Godbout [b], Sébastien Fournel [d]

[a] *Institut National de la Recherche Scientifique (INRS), Centre Eau Terre Environnement (ETE), 490 rue de la Couronne, Quebec (QC) G1K 9A9, Canada*
[b] *Research and Development Institute for the Agri-Environment (IRDA), 2700 rue Einstein, Quebec (QC) GIP 3W8, Canada*
[c] *Département des sols et de génie agroalimentaire, Faculté des sciences de l'agriculture et de l'alimentation, Université Laval. 2425, rue de l'Agriculture, Québec (QC) G1V 0A6, Canada*
[d] *Tecnológico Nacional de México (TecNM) – Campus Perote. Km 2.5, Carretera Perote – México, Perote 91270, Mexico*

ARTICLE INFO

Keywords:
Egg yield
Aviary systems
Statistical modeling
Air quality
Machine learning
Egg quality

ABSTRACT

Since 1990, worldwide egg production has increased on average 2.8% per year. This increase has drawn the attention of animal welfare advocates. In Canada, new challenges have emerged, among them: increased awareness in animal welfare and environmental footprint and a shift to cage-free egg production systems (CFSs). Welfare assessment of environmental control strategies (ECSs), and deployment of early warning egg production systems based on on-line monitoring, have become an important field of research to understand interactions and mitigate environmental issues related to these CFSs. This study assessed the effect of selected ECSs on hen-day egg production (HDEP) and daily egg cleanliness (EGC) in experimental CFSs, using Principal Component Analysis and Linear Discriminant Analysis. The assumption that air quality conditions could play an important role in HDEP predictions was addressed by developing a machine learning method based on Random Forest models (RF) to predict HDEP daily fluctuations using measured hygrothermal and air quality conditions as input variables. A variable importance analysis further confirmed the governing variables of the egg production time series. Following that, an inquiry-driven scenario analysis was performed to identify potential changes in egg production. Results showed that the ECSs did not disrupt the egg production in the experimental CFSs. Meanwhile, A RF model with a window size of 14 days showed satisfactory performance predicting HDEP daily fluctuations with a RMSE of 0.176% and 0.368%, and a R^2 of 0.94 and 0.78, for training and testing, respectively. The temperature was the dominant governing variable among the predictors, followed by hen's age and relative humidity. Finally, the scenario analysis revealed that a 5% temperature's increase could negatively affect the egg yield. The outcomes of this study aim to contribute to the on-line monitoring and control activities in laying hen facilities as an important aspect in Precision Livestock Farming.

1. Introduction

Since 1990, worldwide annual egg production has achieved on average a 2.8% increase, reaching 76.8 million metric tons in 2018. In Canada, egg production rate and product demand have also grown. In 2019, the demand for eggs increased by 1.25% when compared to that in 2018, while since 2015 the annual egg production had increased by 4.7%. Meanwhile, new challenges have emerged in light of these substantial production rates and farm's capacity (Agriculture and Agri-Food Canada, 2019a). To date, Canadian egg farms' capacity ranges from 100,000 up to 400,000 laying hens, each hen producing on average 340 eggs per year. The Province of Quebec has 20.1% of the country's quota allocation (Agriculture and Agri-Food Canada, 2019b).

This impressive growth of the egg industry has drawn the attention of animal welfare advocates. Thus, egg production systems have adapted over the last 20 years, starting with the EU Council Directive 99/74 whereby conventional laying hen systems have shifted from batteries of cages to alternative systems such as furnished cages or cage-free egg production systems (CFS) (Shields et al., 2017). In Canada, these alternative systems were introduced with the 2017 Code of Practice for laying hens. The Code sets forth beneficial in-farm practices based on the

* Corresponding author.
E-mail address: andres.gonzalez@inrs.ca (A.F. Gonzalez-Mora).

https://doi.org/10.1016/j.compag.2022.106854
Received 22 May 2021; Received in revised form 8 February 2022; Accepted 4 March 2022
Available online 11 March 2022

Nomenclature

ACP	Animal Care Program
AirFlow	Outlet airflow rate (L min^{-1})
CART	Classification and regression trees
CEc	Count of clean eggs
C.I.	Confidence intervals
CFS	Cage-free egg production system
CFR	Cage-free livestock rooms
CO_2	Carbon dioxide
CH_4	Methane
DoA	Laying hen's age in days
ECS	Environmental control strategies
EGC	Daily egg cleanliness index (%)
GHG	Greenhouse gases
HDEP	Hen-day egg production index (%)
H_2O_g	Water vapor
KW_{pw}	Kruskal-Walls test with a pairwise Wilcoxon test
LDA	Linear Discriminant Analysis
mtry	Number of predictors at each split
ntree	Number of trees
MSE	Mean square error
N_2O	Nitrous oxide
NH_3	Ammonia in gas phase
OOB	Out-of-bag subset
OOBp	Out-of-bag permuted
OOB-MSE	Estimate error rate
ppm	parts-per-million unit
PCA	Principal Component Analysis
PM	Particle matter
PM_{10}	Particle matter diameter <10 μm
$PM_{2.5}$	Particle matter diameter <2.5 μm
Random	Variable of random numbers
RMSE	Root Mean Square Error (%)
R^2	Coefficient of determination
RF	Random Forest
RH	relative humidity (%)
SW_t	Shapiro-Wilk test
TLEc	Total laid eggs
THc	Total number of laying hens
TEc	Total number of eggs
Temp	Indoor temperature (°C)
%IncMSE	Increase in MSE in percentage

Animal Care Program (ACP), along with a transition strategy to move from conventional systems to CFSs over a projected period of 19 years. In Quebec, conventional cages have been prohibited since 2015 for all new egg producers (Philippe et al., 2020).

Egg production and international market demand combined with technological developments, environmental changes, and consumer awareness, have produced complex interactions. Moreover, due to food production increase, sustainable laying hen housing systems must meet the demand while minimizing the environmental footprint (Pelletier et al., 2018). For these reasons, the deployment of several strategies to reduce this footprint and understand the complex interactions ensuring high egg production performances, have become highly relevant research subjects.

In laying hen housing systems, manure management and animal density play key roles in gas and dust emissions (David et al., 2015a; David et al., 2015b). Ammonia (NH_3) represents one of these harmful gases produced by the degradation of uric acid from stored feces. High levels of NH_3 (greater than 25 ppm) inside laying hen houses can reduce feed intake and animal growth. Furthermore, high NH_3 concentrations over large periods of time can lead to respiratory illnesses and increase susceptibility to viral diseases, affecting egg quality and egg production (Tong et al., 2019). Laying hen houses can also contribute to greenhouse gases (GHG) such as carbon dioxide (CO_2), water vapor (H_2O), methane (CH_4), and nitrous oxide (N_2O). These GHGs are produced by the bird's respiratory cycles and the biochemical processes associated with manure management, in which high levels could affect bird's health (Ribeiro et al., 2018). Production of particle matter (PM), such as PM_{10} (diameter < 10 μm) and $PM_{2.5}$ (diameter < 2.5 μm), results from the interaction between animal activity levels, access to litter, and environmental conditions. PMs can be harmful because they can be inhaled and cause serious health problems for animals and caretakers (David et al., 2015b).

Studies about conventional and alternative laying hen housing systems have shown that conventional cages tend to lead to better air quality conditions and ensure high egg production performance than alternative systems. However, the lack of space and high animal density reduce the animal welfare conditions (Pelletier et al., 2018). On the other hand, alternative systems have shown to provide positive impacts on hen behaviors and food safety. Nevertheless, there is a risk to hen and workers' health mainly induced by poor indoor air quality conditions arising from these systems, more specifically on CFSs (Pelletier et al., 2018). Thus, sustainable strategies to improve air quality in CFSs have become focal points. Therefore, to curb aggravating indoor air quality conditions, several environmental control strategies (ECSs) have been proposed, such as: (i) amendments of chemical and natural compounds to the litter (Qasim et al., 2017; Schneider et al., 2016; Zhang et al., 2016); (ii) different bedding materials (van Harn et al., 2012); (iii) incorporation of inert material in hen feed (Prasai et al., 2018); (iv) sprinkler application of neutral electrolyzed water to the litter (Chai et al., 2018); (v) air-bulk ionization; and (vi) use of vegetable oil spraying systems on the litter (Aarnink et al., 2011; Winkel et al., 2016).

Besides, egg production has been one primary measure to understand interactions, variability, and laying performance in hen housing systems (Godbout et al., 2020). The hen-day egg production index (HDEP) is usually defined as the ratio of the total number of laid eggs and the total number of producing laying hens housed over different time intervals (days, weeks, years) (Misslin, 2017). Indeed, HDEP time series could serve as a decision-making tool to egg producers, since it could supply information about flock performance, illness patterns, or egg quality and on-farm economic profits. Moreover, tracking egg production provides a good indicator of nutrient intake balances, improvement of on-farm practices plans, and the effect of environmental conditions (Gorgulu and Akilli, 2018; Misslin, 2017). Animal genetic, feed intake, water consumption, lighting program, and ambient temperature could be factors affecting variability on these egg production curves (Ahmad, 2011; Omomule et al., 2020). In addition, the shifting towards new alternative systems has also led to new egg cleanliness challenges, affecting production efficiency and economic gains. In Canada, an egg not intended for human consumption is defined as an egg with a broken or unbroken shell with adhering dirt; that is, blood, yolk, or manure spots (Egg Farmers of Canada, 2018). A high percentage of dirty eggs have been reported in CFSs compared to conventional and furnished cages (Holt et al., 2011). These aspects have encouraged the use of modeling techniques to predict HDEP given the increasing bird populations within CFSs and the need to increase net farm income, along with economic projections and interaction analyses with environmental conditions, animal health, and farming activities (Ramírez-Morales et al., 2017).

Several non-linear regression models have been proposed to identify time-dependent HDEP fluctuations, namely: Gamma, McNally, McMillan, Adams-Bell, and Compartmental model, to name a few (Gorgulu and Akilli, 2018; Narinc et al., 2014). These non-linear models have specific constants related to initial production, rate of increase and decrease of egg production, initial day of egg-laying, or biological aspects. Data-driven models have also emerged in egg production modeling, such as Artificial Neural Networks (Ahmad, 2011; Ramírez-Morales et al., 2017), Least-square Support Vector Machines (Gorgulu and Akilli, 2018) or Fuzzy logic algorithms (Omomule et al., 2020). These models have used several variables, such as feed intake, mortality, live animal counts, and historical registered egg production, as input variables. However, to the best of our knowledge, there is no evidence that air quality data, *i.e.*, gas concentrations, have been used as inputs to data-driven models to predict egg production. The available literature on egg production modeling using air quality variables is still scarce, and additional research is recommended to establish their interaction with egg yield (Gorgulu and Akilli, 2018).

This paper presents the results of a project focusing on the effect of selected ECSs on egg production and egg quality from an experimental aviary system. This study contributed to a larger project that also investigated the effect of these ECSs on air quality and animal welfare. A panel of experts selected the ECSs following an extensive review of available literature and pre-testing performed in a laboratory of the Research and Development Institute for the Agri-Environment (IRDA), Québec (Canada). A future companion paper will present detailed results about ECSs pre-testing and air quality observations. Regarding animal welfare, the results showed that selected ECSs did not affect natural hen behaviors and spatial occupancy inside the experimental CFSs, besides reducing litter allowance treatment (Gonzalez-Mora et al., 2021). Following that, the goal of the present study was to demonstrate whether or not the selected ECSs could have any effect on egg productivity and egg quality. Moreover, the assumption that air quality conditions could play an important role in HDEP predictions was addressed by the application of a data-driven model. This study aims to contribute to the on-line monitoring and control activities, as an essential aspect of Precision Livestock Farming (Niloofar et al., 2021).

Three specific sub-objectives motivated this study:

(i) evaluate the effect of the selected ECSs on egg production (HDEP) and egg cleanliness index (EGC) from an experimental aviary system, using exploratory methods such as principal component analysis (PCA) and linear discriminant analysis (LDA),
(ii) predict the HDEP daily fluctuations from an experimental aviary system, using a data-driven model, known as Random Forest, and measured hygrothermal and air quality conditions as input variables, and
(iii) assess the effect of varying the main explanatory variable on the predicted egg production curve, obtained by a variable importance analysis.

The paper is organized as follows: Section 2 introduces the materials and methods used in this study; Section 3 the ensuing results; Section 4 the discussion and finally Section 5 the conclusions and recommendations for future studies.

2. Materials and methods

2.1. Animal housing and egg handling

2.1.1. Experimental setup

Twelve bench-scale experimental cage-free rooms (CFR) were used to shelter twelve Lohmann LSL-Lite laying hens per room (n = 144) from February to June 2019. The animal density was one hen per 1115 cm^2, ensuing the space allowance recommendations for non-cage systems by the National Farm Animal Care Council of Canada (NFACC). Animals arrived at 19 weeks of age with the vaccination requirements. Then, hens were individually weighed and randomly placed in the CFRs. At the beginning of the experiment, an acclimatization period of 2 weeks was granted to allow (i) the hens adapt to the new environment, and (ii) the caretakers to carry out flash observations of hen behavior and animal adaptability.

The experiment was divided into two consecutive experimental batches of 8 weeks each (batch 1 and batch 2), according to the experimental setup for the ECSs (Fig. 1). CFRs were located at the livestock building agri-environmental assessment laboratory of IRDA, in Quebec (Canada). The use and treatment of laying hens in this study was approved by the *Comité de protection des animaux* (Animal Welfare Committee) of the research center (CPA-CRSAD – authorization number 19AVCPA01). The bench-scale CFRs (122 cm length, 119 cm wide) were equipped with a variable-speed exhaust fan. CFRs consisted of a ground and an upper floor conception. Ground-level was a litter space initially conditioned with 5-cm thick wood shaving bedding. The upper-level was a square mesh wire cloth furnished with two nest boxes, one linear feeder, two aviary nipple drinkers, and two perch PVC-round pipes. CFR design and livestock units (*e.g.*, linear feeder) met the requirements of the Code of practice for the care and handling of laying hens of the NFACC. A more detailed description of the CFR design and conception is available in Gonzalez-Mora et al. (2021).

The incoming air was pre-conditioned with a conditioning unit and a heated electrical resistance. This way, the incoming air could meet the target temperature inside CFRs throughout the experimental housing period (from 22 °C to 23 °C).

Laid eggs were manually collected, counted, and classified (clean/dirty) daily for each CFR. Total laid eggs, clean and dirty ones, were manually registered in an experimental book and daily reports digitized and stored in a digital database.

2.1.2. Hen-day egg production index

Equation (1) introduces the mathematical expression to generate egg production data using a hen-day egg production index (HDEP) (*i.e.*, time series):.

$$\boldsymbol{HDEP}(\%) = \frac{\boldsymbol{TLEc}}{\boldsymbol{THc}}\boldsymbol{x}100 \tag{1}$$

Where *TLEc* is the total laid eggs count, and *THc* is the total number of laying hens.

2.1.3. Daily egg cleanliness index

An egg cleanliness index (EGC) was defined as the daily ratio of clean eggs (CEc) and the total number of eggs (TEc) counted (Equation (2)). Clean eggs were collected from the nest egg collector following the quality requirements of Egg Farmers of Canada (2018). Otherwise, eggs were classified as dirty. It should be noted that eggs could be laid either in the litter space or wire floor area, however, these eggs were classified as dirty.

$$\boldsymbol{EGC}(\%) = \frac{\boldsymbol{CEc}}{\boldsymbol{TEc}}\boldsymbol{x}100 \tag{2}$$

2.2. Hygrothermal and air quality monitoring

Hygrothermal conditions, *i.e.*, temperature and relative humidity (RH); air quality conditions, *i.e.*, gas concentrations (CO_2, N_2O, CH_4, and NH_3); and ventilation rates were monitored before application of any ECSs (2 weeks before) and throughout both batches 1 and 2, respectively. Temperature and RH were measured with a T&RH probe (Model CS500, Campbell Scientific, Inc., Canada corp.). Gas concentrations were measured upstream and downstream, using an infrared gas analyzer (FTIR, model DX4040, Gasmet, Pulttitie 8A 08,880 Helsinki, Finland) located in a laboratory unit. The gas analyzer uses a Fourier transform infrared spectrometer (FTIR), and a built-in sample gas pump to measure gas concentrations. Ventilation rates were calculated by the pressure difference (ΔP) downstream from a 204-mm iris orifice damper

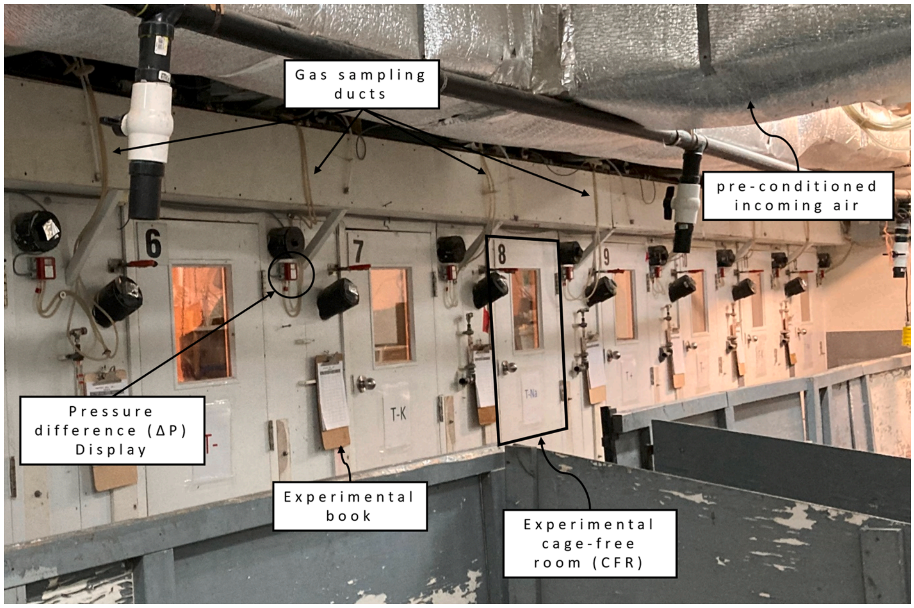

Fig. 1. Experimental set-up at the livestock building agri-environmental assessment laboratory of IRDA.

(Model 200, Continental Fan Manufacturing Inc., Buffalo, NY, USA) installed in the exhaust duct of each CFR. A data logger recorded hygrothermal data, gas concentrations, and ΔP every 15 min. Gas emissions were calculated by multiplying the ventilation rate and gas concentration delta at each CFR.

2.3. Environmental control strategies (ECS)

One control (C0), one treatment proposed by the *Fédération des Producteurs d'Oeufs du Québec* (FPOQ, Quebec's Federation of Egg Producers) (T1), and two combined ECSs (T2 and T3) were applied and equally distributed over the 12 CFRs to evaluate indoor air quality conditions, animal welfare, and egg production (see Table 1). The control was a traditional aviary system with 33% of the entire floor space with litter surface following the NFACC recommendations (n = 3) (Fig. 2a). T1 was proposed by the FPOQ to evaluate the possibility to decrease the litter surface (Fig. 2b). More information is available in (Gonzalez-Mora et al., 2021). A review of different ECSs by a panel of experts led to the selection and implementation of combined ECSs, namely: (i) in-floor heating combined with an oil sprinkling system (Fig. 3a), and (ii) litter absorbent combined with an oil sprinkling system (Fig. 3b). Then, pre-testing was carried out in the agri-environmental assessment laboratory at IRDA to establish the best set up among the selected strategies seeking reduction of ammonia concentration from hen litter box-samples (Godbout et al., 2020).

Table 1
List of treatments applied within the CFRs.

Abb.[a]	ECSs	CFRs[b]	Description[c]
T1	Reduced litter surface area	1-5-11	17% of litter area, reduction of litter surface from 33% to 17% (n = 3).
T2	Heated floor + oil sprinkling	2-6-10	33% of litter area, installation of a heated floor set to 27 °C. Spraying an oily emulsion over litter (1.17 L m^{-2} $week^{-1}$) (n = 3)
T3	Litter absorbent + oil sprinkling	3-7-9	33% of litter area, addition of 10%-litter of acid adsorbent (Active biochar). Spraying an oily emulsion over litter (1.17 L m^{-2} $week^{-1}$) (n = 3)

[a] Abb. = Abbreviation.
[b] Room number of the CFRs at BABE.
[c] CFRs design was based on a traditional aviary system under NFACC.

2.4. Feature and data classification analysis

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were applied to observe interactions of selected ECSs and treatments with the HDEP and EGC observations. PCA is a non-supervised method that uses Euclidian distances to assess feature classification. On the other hand, LDA is a technique primarily used for data classification where data is reduced from a high dimensional to low dimensional space in light of obtained projections, which can be maximized between-class distances (Parent, 2020). The R programming language (R Core Team, 2020) was used to perform the analysis using the *ade4* package (Dray and Dufour, 2007) and the methodology proposed by Parent (2020).

2.5. Egg production modeling

The application of non-linear models was limited due to the availability of initial parameters for the model. In this study, hens were housed for 16 weeks, instead of an entire egg production period. This experimental framework did not allow to have parameters such as: initial production, initial day of egg-laying or rate of increase. Thus, use of a data-driven model known as Random Forest (Appendix A), along with moving averages, was proposed to assess egg production modeling.

2.5.1. Data analysis

Time series of hygrothermal and gas emissions recordings were analyzed using measures of central tendency and dispersion. Confidence intervals were established using the Central Limit Theorem method for data not drawn from a normal distribution and sample sizes greater than 40 samples (n greater than 40). Boxplots in this study present the median, the first and third quartiles (the lower and upper hinges), while the lower and upper whiskers delineate the inter-quartile distance (IQR) between the first and third quartiles (1.5 * IQR), the outliers (values beyond the whiskers), and the notches with a 95% confidence interval for median comparison, calculated following the mathematical expression 1.58 * IQR / sqrt(n) where n is the number of observations. A Kruskal-Wallis Rank Sum test with a pairwise Wilcoxon test (KW_{pw}) were applied to observe data comparisons. Also, Shapiro-Wilk test (SW_t) was used to test normal distribution behavior within the data. A *p*-value of 0.05 was established to accept null hypothesis (H_o) within tests. This analysis allowed to: (i) perform statistical comparisons between the

Fig. 2. Difference in litter surface area inside the cage-free rooms.

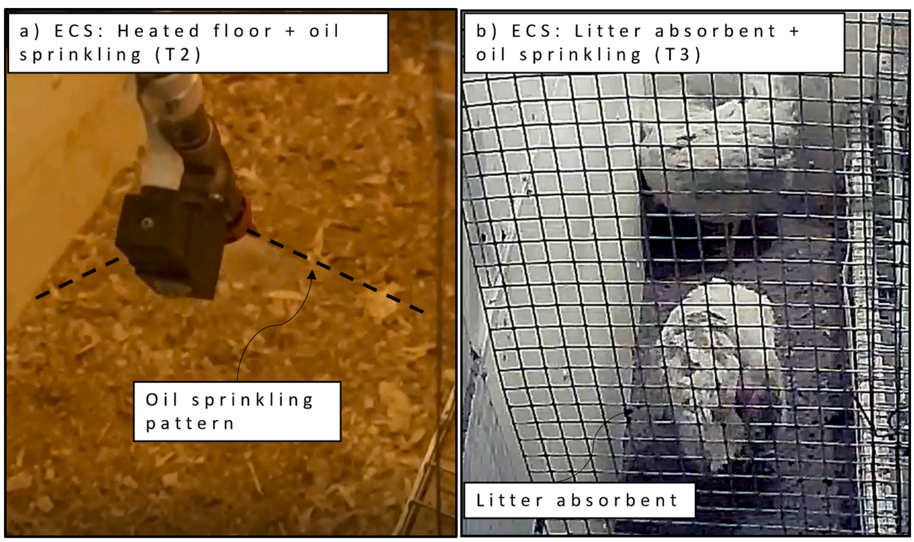

Fig. 3. Oil sprinkling pattern and biochar application inside the cage-free rooms.

treatments (T1, T2, T3) and the control (C0), (ii) to observe an overview of the distribution of the hygrothermal and air quality conditions during the experiment, and (iii) to compare them with literature ensuring conditions within the standards reported in laying hen facilities.

2.5.2. *Moving averages*

Moving averages were used to estimate the egg production trend using several window sizes. Sliding windows allowed to observe egg production trends over time by assuming a locally linear regression from a specific window size (span). These simple mathematical transformations could also reduce noise and follow a similar approach from local weighted regression functions (LOESS). Five intervals were proposed, namely, 3, 5, 7, 10, and 14 days. We assumed that the trend had a linear behavior within the interval considered. Equation (3) describes the mathematical expression used to determine the moving average.

$$u_i = \frac{1}{k+1}\sum_{j=i-k}^{i} y_j \tag{3}$$

Where k is the window size, y_j is the observed value and u_i is the unweighted mean of the measurements for the times.$t_{i-k}, t_{i-k+1}, t_{i-1}, t_i$

2.5.3. *Data-driven model*

The Random Forest model (RF) was selected to predict HDEP daily fluctuations using hygrothermal and air quality data obtained from the traditional experimental aviary system (*i.e.*, C0). Then, RF was used to establish the correlation between either air quality variables or hygrothermal variables and egg production. This model is a user-friendly algorithm with only two parameters to be optimized, namely, the number of predictors to split at each node (*mtry*) and the number of trees to develop in the forest (*ntree*) (Liaw and Wiener, 2002). The reader is referred to the Appendix A for more detailed information about the Random Forest models.

2.5.3.1. *Input data.* RF model was built in the R programming language version 4.0.0 (R Core Team, 2020) using the *caret* package (Kuhn, 2020). The *randomForest* package (Liaw and Wiener, 2002) was also used to perform the model framework. The following variables were selected as input variables, namely: CO_2, CH_4, N_2O and NH_3 concentrations, indoor temperature (Temp), relative humidity (RH), outlet airflow rate (AirFlow) and hen's age expressed in days (DoA). Input variables were transformed in daily basis averaging the observations in a day. A variable with random numbers (Random) was added to the final input dataset to validate the variable importance performed by the RF model (see Section 2.6). The Random variable was generated from a uniform distribution from 0 to 1 using the *runif* function in R. For HDEP modeling, only the data from the control (C0) was use as input.

2.5.3.2. Splitting and resampling method. Training and testing datasets were used to build and validate the RF model. Therefore, the original dataset was randomly split into training (70%) and testing (30%) subsets. In all cases, the training dataset was re-sampled into several training subsets by 10-fold cross-validation with 5-repetitions to increase the diversity within predictions and to get better performance in model estimation as proposed by Polikar (2006) and Kuhn and Johnson (2013).

2.5.3.3. Model parametrization. RF model parametrization was achieved by using the train function from the *caret* package (Kuhn, 2020). The number of predictors at each split (mtry) and number of trees (ntree) were varied from 1 to 9 predictors, and from 1,000 to 3,500 trees with increments of 1 predictor and 500 trees. Optimal values of *mtry* and *ntree* parameters were selected using the root mean squared error (RMSE) as objective function. An iterative algorithm was used to generate different RF models of different window sizes. One RF model, with the lowest RMSE, was selected for each window. Then, the RMSE evolution was observed to select the best RF model.

2.6. Variable importance

A variable importance analysis was used to rank the predictors by evaluating changes in the prediction after training was completed. This analysis provided information about the driving forces involved in the process and highlighted the key variables useful for monitoring activities in livestock farming. Variable importance was assessed for each predictor as follows: (i) for the Out-Of-Bag (OOB) process (Appendix A), the values were randomly permuted (OOBp); (ii) a decision tree was grown using the new data and new predictions calculated; and (iii) a mean squared error (MSE) was estimated from new predictions. Thus, the importance of a predictor was defined by an increase in the difference of the estimated MSE from OOBp and the MSE using the original OOB, expressed in percentage (%IncMSE) (Breiman, 2001; Liaw and Wiener, 2002).

2.7. Model evaluation

Root mean squared error (RMSE) and the coefficient of determination (R^2) were calculated to evaluate model performance (Equation (4) and Equation (5), respectively):.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{N}} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \widehat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \overline{y})^2} \tag{5}$$

Where, y_i and $\widehat{y}_i$ are observed and predicted values, $\overline{y}$ is the mean value of observed data and N is the quantity of data.

2.8. Scenario analysis

An inquiry-driven scenario analysis was performed to identify potential changes in egg production, as well as model response when varying the input of the selected RF model. The following procedure was adapted from Alcamo (2008):

1. Select the RF model with the best performance characteristics to predict time-dependent egg yield fluctuations and use it in a predictive mode.
2. Identify a “driving force” or the variable with more importance observed after applying the variable importance analysis.
3. Apply an increase of 5% to the driving force and add new values to the original dataset.
4. Run the selected predictive model and analyze results.

3. Results

3.1. Hygrothermal and ventilation rate conditions

The hygrothermal conditions and ventilation rates of each treatment for batches 1 and 2 are presented in Fig. 4. A total of about 2050 measures were analyzed for each treatment. Statistical differences were observed in temperatures from T1 (22.06 ± 0.03 °C) and T3 (22.06 ± 0.03 °C) comparing to the control (21.68 ± 0.02 °C) in batch 1 (±, 95% C.I., KW_{PW}: $p < 0.05$). Also, significant differences were observed for T1 (21.99 ± 0.04 °C), T2 (21.51 ± 0.03 °C), and T3 (21.99 ± 0.04 °C) with respect to the control (21.63 ± 0.03 °C) in batch 2 (±, 95% C.I., KW_{PW}: $p < 0.05$). There were no any statistical differences observed in relative

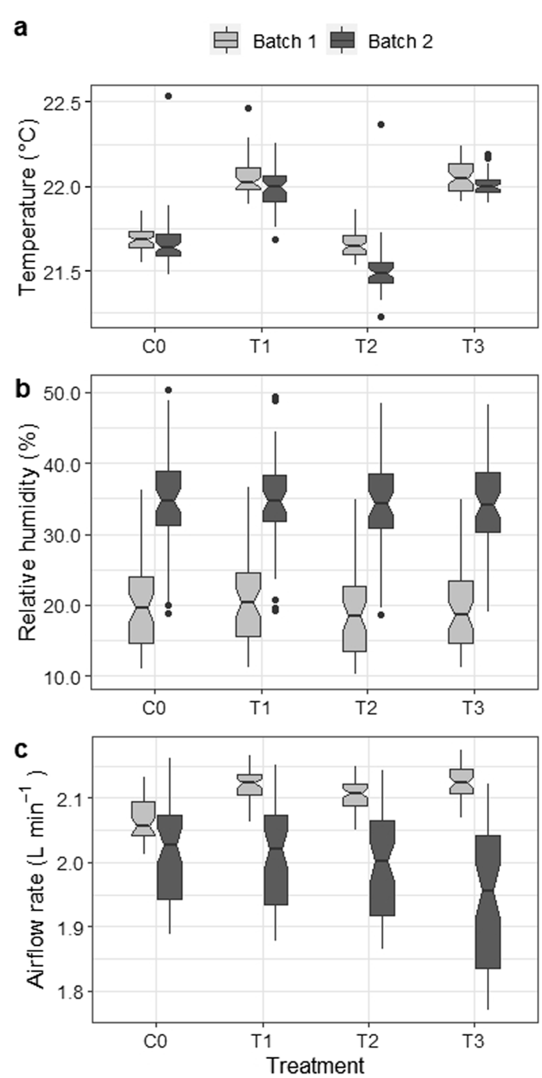

Fig. 4. Boxplot of hygrothermal and ventilation rate conditions of three ECSs and control for both batches 1 and 2: (a) temperature, (b) relative humidity, (c) airflow rate.

humidity (RH) between treatments and control in batches 1 and 2, respectively. However, the overall RH in batch 1 was lower than that reported for batch 2 for all treatments. The average values of RH for T1, T2, T3 and control in batch 1 were 20.6 ± 1.8%, 19.0 ± 1.7%, 19.8 ± 1.6%, and 20.0 ± 1.8% (±, 95% C.I.); while the average of RH in batch 2 were 34.6 ± 1.8%, 34.0 ± 1.8%, 34.1 ± 1.8%, and 35.1 ± 1.9% (±, 95% C.I.), respectively. Higher values of airflow rates were observed from T1 (2.120 ± 0.007 L min^{-1}), T2 (2.106 ± 0.006 L min^{-1}) and T3 (2.126 ± 0.007 L min^{-1}) in contrast to the control (2.065 ± 0.009 L min^{-1}) in batch 1 (±, 95% C.I., KW_{PW}: $p < 0.05$). Besides, lower values of airflow rates were observed for T3 (1.946 ± 0.028 L min^{-1}) in comparison to the control (2.020 ± 0.020 L min^{-1}) in batch 2 (±, 95% C.I., KW_{PW}: $p < 0.05$).

3.2. Indoor air quality conditions

Indoor air quality conditions of treatments T1, T2, T3 and the control (C0) for batches 1 and 2 are summarized in Fig. 5. Significant differences were found in CO_2 emissions in T2, with an average value of 26.39 ± 0.72 kg yr^{-1} hen^{-1}, compared to the control in batch 1, where the average of CO_2 emissions was 28.59 ± 0.93 kg yr^{-1} hen^{-1} (±, 95% C.I., KW_{PW}: $p < 0.05$). The average of CO_2 emissions in batches 1 and 2 ranged from 26.29 ± 0.74 to 28.59 ± 0.93 kg yr^{-1} hen^{-1} and from 24.77 ± 0.14 to 26.28 ± 0.16 kg yr^{-1} hen^{-1} (±, 95% C.I.), respectively. On the other hand, there were not any statistical differences observed in CH_4 emissions between the ECSs and the control. The average values of CH_4 emissions in batches 1 and 2 ranged from 31.2 ± 5.88 to 35.1 ± 6.21 g yr^{-1} hen^{-1} and from 30.3 ± 3.57 to 35.4 ± 4.49 g yr^{-1} hen^{-1} (±, 95% C.I.), respectively, showing similarities in methane emissions between batches. N_2O emissions from T1 (1.75 ± 0.10 g yr^{-1} hen^{-1}), T2 (1.78 ± 0.09 g yr^{-1} hen^{-1}) and T3 (1.73 ± 0.20 g yr^{-1} hen^{-1}) were significantly higher than emissions for the control (1.69 ± 0.10 g yr^{-1} hen^{-1}) in batch 1 (±, 95% C.I., KW_{PW}: $p < 0.05$). However, no statistical differences of N_2O emissions were observed between treatments and control in batch 2 where average values ranged from 1.95 ± 0.19 to 2.13 ± 0.23 g yr^{-1} hen^{-1} (±, 95% C.I.). Also, no significant differences in NH_3 emissions were observed between treatments T1, T2 and T3 and the control for batches 1 and 2. Average values of NH_3 emissions ranged from 7.99 ± 0.95 to 10.46 ± 1.66 g yr^{-1} hen^{-1} and from 21.03 ± 2.89 to 34.56 ± 5.91 g yr^{-1} hen^{-1} (±, 95% C.I.) for batch 1 and 2, respectively.

3.3. HDEP and EGC time series

Egg production (HDEP) and egg cleanliness (EGC), in time series, of each treatment (T1, T2, T3) and the control (C0) are presented in Figs. 6 and 7. The overall averages ± SD (standard deviation) of HDEP and EGC were 97.6 ± 4.2% and 87.0 ± 6.9%, respectively. Results show that T1 had a slightly superior HDEP mean index (98%) compared to the other treatments. However, the lowest minimum HDEP index was found for treatments T1 and T3 (83.3%), respectively. Noteworthy, T3 had the lowest HDEP mean index (97.1%) and highest SD (5.2%). This was also noticed for the EGC observations (82.9 ± 8.4%) (Fig. 7). Also, T3

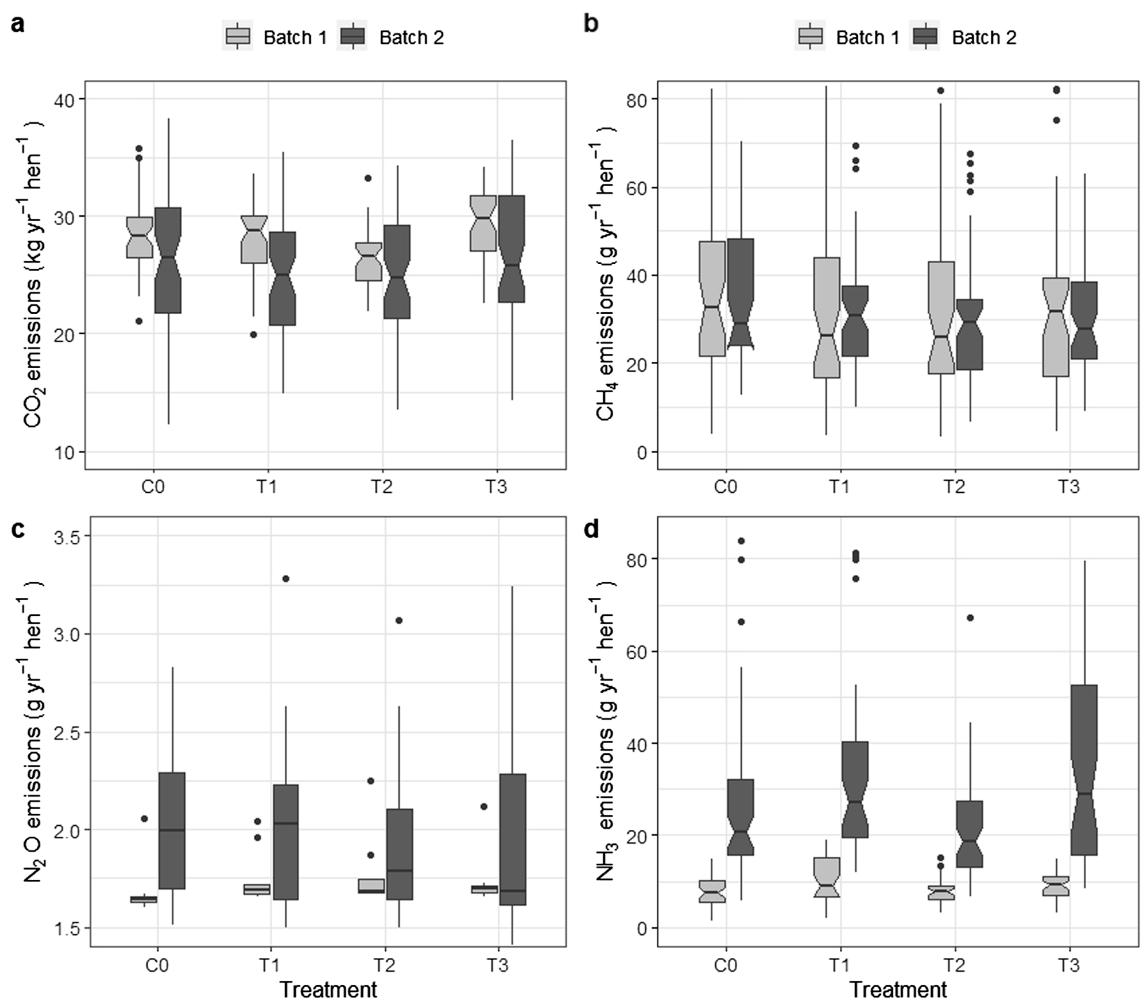

Fig. 5. Boxplot of gas emissions inside the three ECSs and the control for both batches 1 and 2: (a) CO_2, (b) CH_4, (c) N_2O, and (d) NH_3.

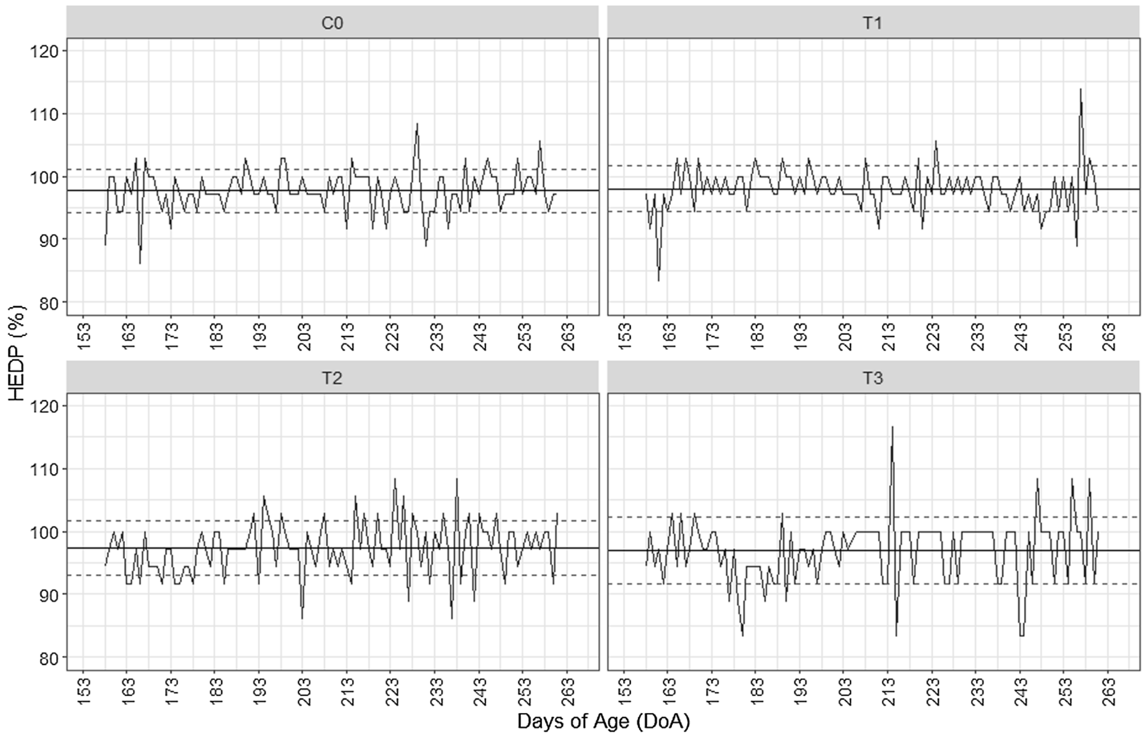

Fig. 6. Hen-day egg production index (HDEP) time series. Bold dashed lines (–): mean values. Dashed lines (–): standard deviations.

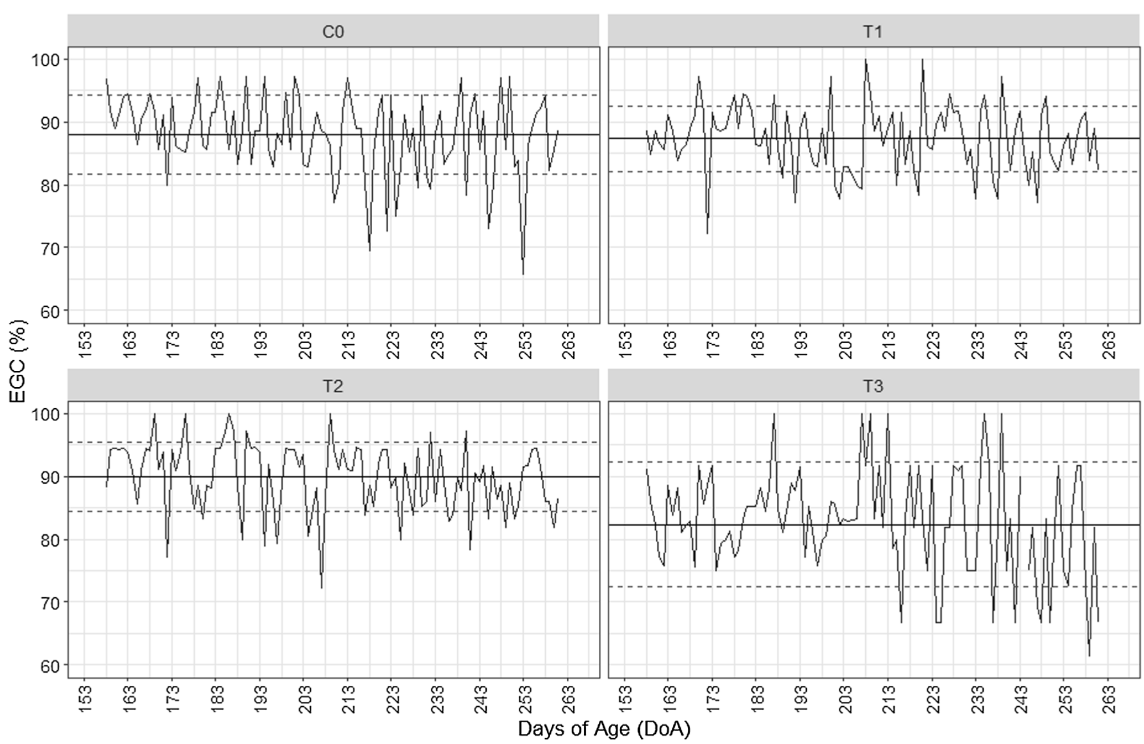

Fig. 7. Egg cleanliness index (EGC) time series. Bold dashed lines (–): mean values. Dashed lines (–): standard deviations.

treatment reached the highest maximum HDEP index (117%) when compared to all treatments, and the same maximum EGC index of T1 and T2 (100%). Furthermore, T2 was the treatment with the highest EGC mean percentage (89.9%), followed by the control and T1. Daily fluctuations of HDEP and EGC were not drawn from a normal distribution, except for treatment T1 (SW_t: $p = 0.3$). Also, significant differences were found in the mean EGC index from T2 (89.9 ± 5.5%) and T3 (82.9 ± 8.4%) compared to control with a mean EGC index of 87.9 ± 6.3% (KW_{PW}: $p < 0.05$).

3.3.1. HDEP trend

HDEP values displayed a constant trend from week 22 to week 37, ranging mostly from 95% to 100% (Fig. 8). A clear reduction in noise was observed when using a window size larger than 7 days, whereas

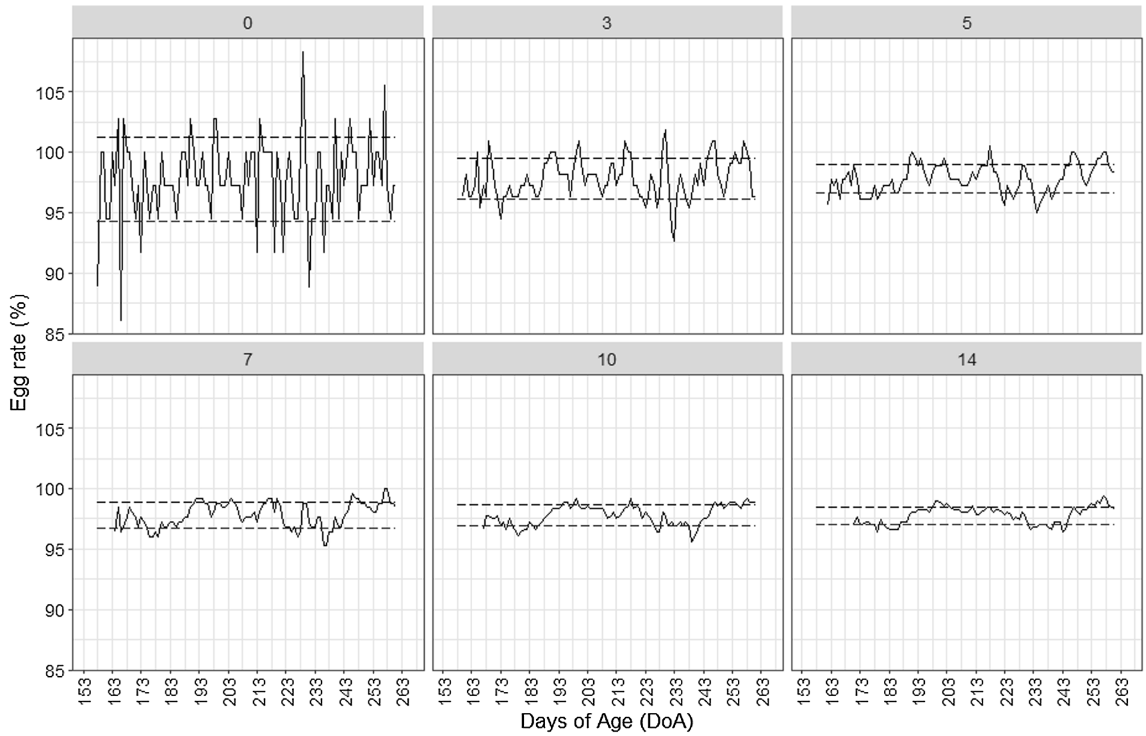

Fig. 8. Egg production curves of the control (C0) using several averaging window sizes, expressed in days. Dashed lines (–): Standard deviation intervals.

marginal reductions could be appreciated between 7, 10, and 14 days. This could be noticed by the narrowing pattern of the standard deviation intervals (black long dashed lines) throughout the windows.

3.4. Effect of treatments over HDEP and EGC

Linear Discriminant Analysis (LDA) was performed with three continuous variables: days of age (*doa*), egg yield (*hdep*) and egg cleanliness (*egc*), and one categorical variable (*i.e.*, *treatments*) (Fig. 9). The first axis (LD1) was mostly associated with the *egc* and reproduced 42.7% of the data variability. The second axis (LD2) was mainly associated with the *hdep* and explained 30.3% of the variability. Hence, 73.1% of the variability was explained by the first two axes. Hen's age did not influence data variability since the data seemed to follow uniform trends along the days of age (see Section 3.3). The score plot did not show any significant differences in HDEP nor EGC index between treatments and control with a confidence interval of 95%. However, the confidence regions of the mean discriminant scores (white ellipses) for T2 and T3 seemed to have a marginal distancing compared to the control regarding the LD1 axis.

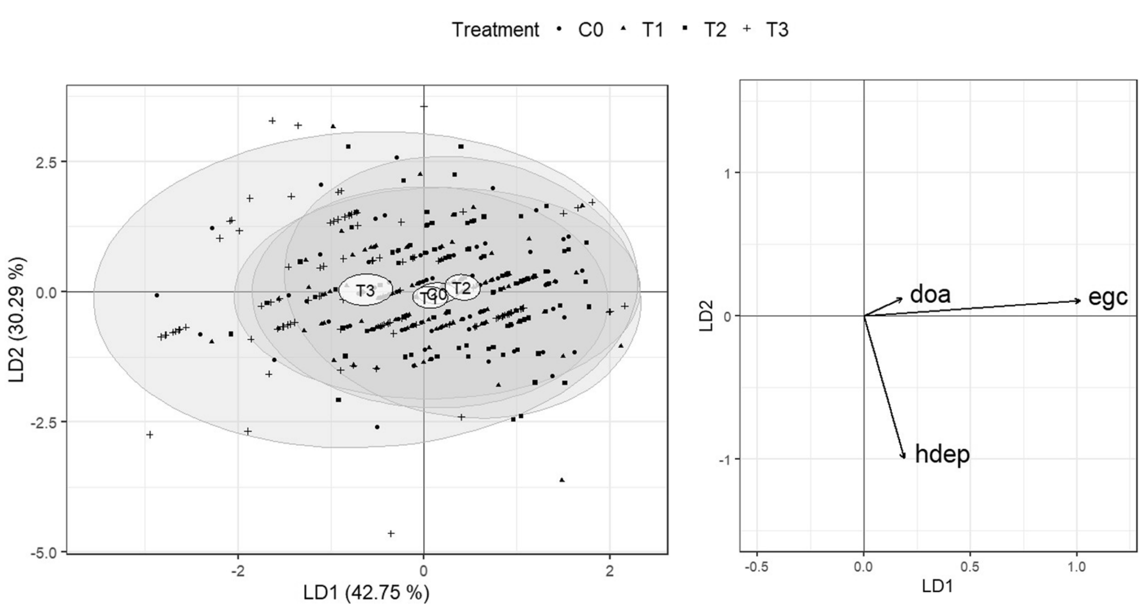

Fig. 9. Linear discriminant analysis (LDA). Left: Score plot. Right: Biplot of loadings.
Adapted from: Parent (2020).

3.5. Random Forest for HDEP predictions

3.5.1. Optimal model parameters

Optimal parameters of different RF models, using different averaging window sizes, revealed that the best model performance was achieved with a window size of 14 days (RMSE = 0.30%, R^2 = 0.84) (Fig. 10). It is noteworthy there was an increase in model accuracy as the window size increased. However, optimal parameters, *i.e.*, *mtry* and *ntree*, were different depending on the training procedure of each RF model and the window size. For this work, a 14-day RF model, with *mtry* and *ntree* parameter values of 2 predictors and 3 000 trees, was selected to predict HDEP daily fluctuations.

3.5.2. Simulation of HDEP

Simulated fluctuations contrasted well with HDEP observations (Fig. 11). A RF model with a window size of 14 days was able to reproduce 94.7% and 78.8% of data variability over the training and testing datasets, respectively. The RMSE was lower for the training (0.176%) than the testing (0.368%), where mean HDEP fluctuations were 97.77% and 97.78%. Observed and simulated HDEP time series are presented in Fig. 12. The testing dataset, *i.e.*, data not used over the training process, was applied to generate Figs. 11 and 12.

3.6. Variable importance

The variable importance analysis showed that temperature was the most important variable (*i.e.*, main governing factor) along with hen's age and relative humidity in predicting egg production daily fluctuations among hygrothermal and gas concentrations variables (Fig. 13). Moreover, the random variable did not show any importance at the end of the analysis. Although the temperature range in our experiment was under the recommended threshold for aviary housing (<35 °C) (Oloyo, 2018), the model was more sensitive to changes in this variable when gas emissions were under the threshold values recommended in the literature (see Section 3.2).

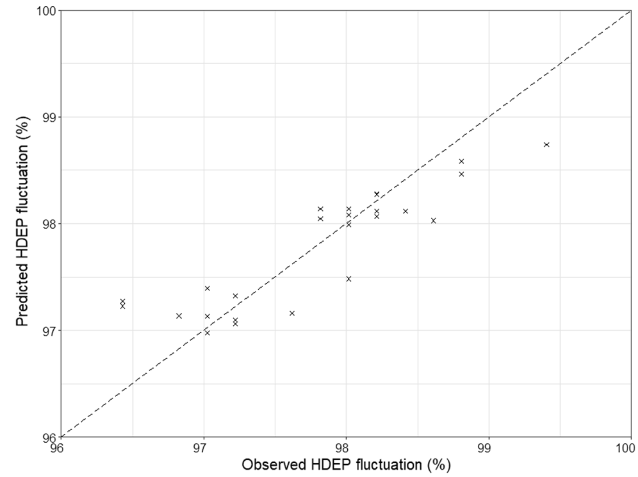

Fig. 11. Observed and simulated HDEP fluctuations using a Random Forest model with a window size of 14 days, mtry = 2 and ntree = 3 000.

3.7. Scenario analysis

Fig. 14 displays egg production curves of original (with a window size of 14 days) and simulated data after applying an increment of 5% to temperature (Scenario: Inc. Temp. 5%) from 171 to 261 days of age. HDEP daily fluctuations are within an interval of ± 2%. It should be noted that sudden HDEP drops and rises were related to the original HDEP daily fluctuations after applying the moving average, not to any animal illnesses or housing deficiency. The increment in temperature (+5%) was proposed to observe the effect in laying performance, as well as the model response to these changes.

After applying the increment of temperature, the HDEP undergoes a decrease mostly at the peaks of egg production (from days 180 to 227 and from 245 until to the end). While observed and simulated HDEP values below the average remain quite similar. Statistical differences were observed between the original egg production and after applying an increase of temperature of 5% ($p < 0.05$, Wilcoxon's signed rank test,

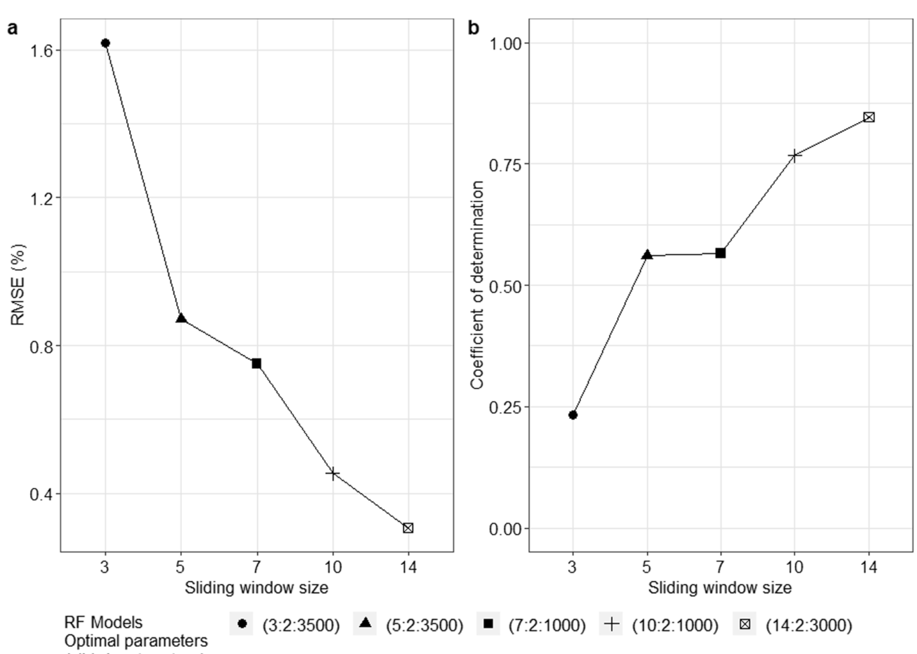

Fig. 10. Optimal parameters of Random Forest models by sliding window size: (a) RMSE and (b) R^2. Optimal parameters (slidwin: sliding window, mtry: number of predictors, ntree: number of trees).

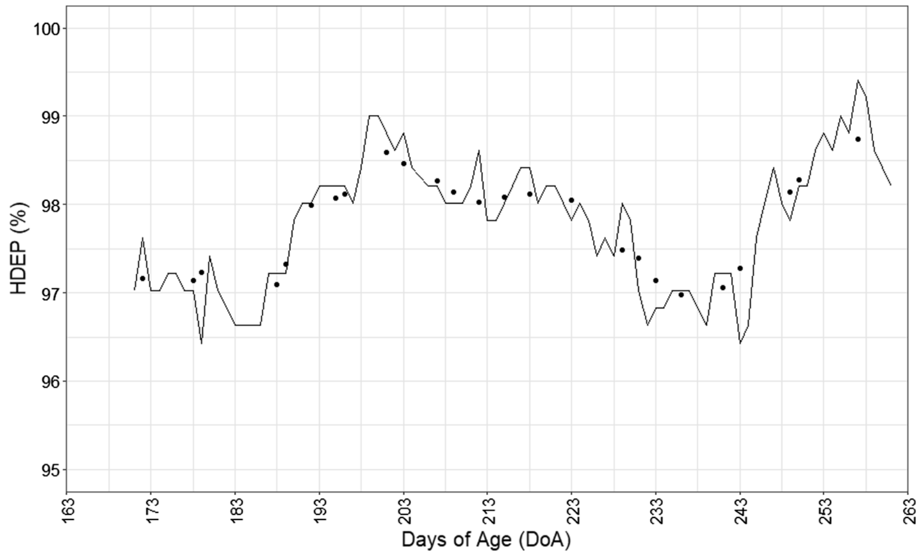

Fig. 12. Simulated (•, black points) and observed (–, black line) HDEP time series using a window size of 14 days.

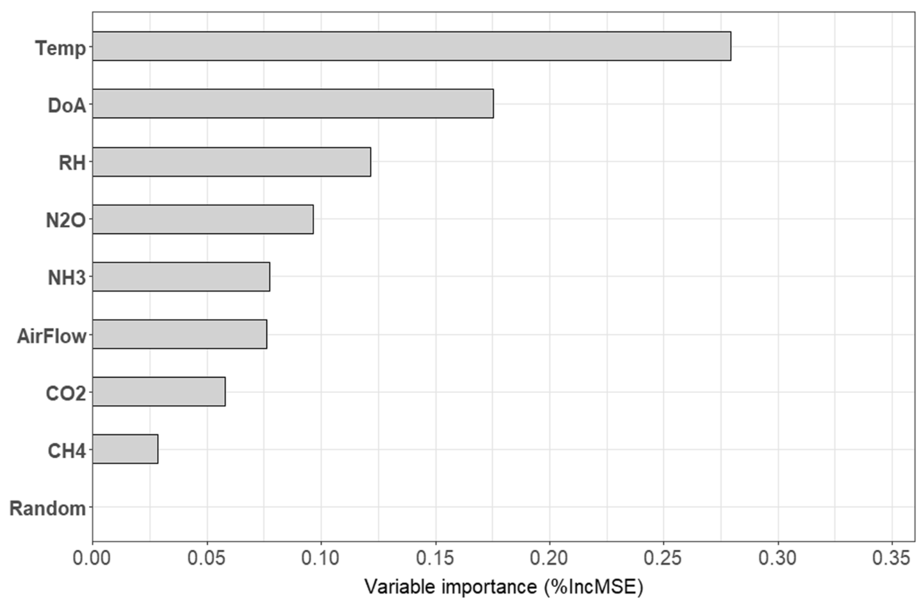

Fig. 13. Predictive variable importance by the increment of the MSE (%).

n = 90). The average values of each egg production curve were about of 97.77% and 97.43% for the original data and the scenario, respectively, showing a slightly difference (0.34%). Nevertheless, the difference observed could be product of the model-predictive error (see Section 3.5.2).

4. Discussion

4.1. Hygrothermal and air quality conditions

Temperatures observed in treatments T1 and T3 were slightly higher than those measured in the control, while most of the time treatment T2 reported similar temperature values when compared to the control. Marginally higher temperatures in T1 and T3 could be associated with several aspects, namely: (i) high activity observed in hens because of the litter reduction treatment (Gonzalez-Mora et al., 2021), probably, increasing natural animal heat production in T1, (ii) litter accumulation under the metallic wire floor because of less presence of hens in litter space increasing some gas emissions and thereby temperature (see Section 3.2), and (iii) increase in litter thickness from agglomerations, generated by the interaction between the biochar and the excreta in the case of treatment T3, producing also an increase in gas emission and temperature. Relative humidity (RH) was lower than the RH reported and recommended in cage-free layer houses (50–80%) (Lin et al., 2017, 2018; Oloyo, 2018; Yilmaz Dikmen et al., 2016). This could be associated with the inlet air conditioning process and the ventilation rates applied throughout the experiment. It is noteworthy that ventilation rates can influence several variables, such as temperature or RH (Lin et al., 2018). In this case, a reduction of ventilation rates induced an increase in RH, as observed in batch 2 (Fig. 4b and 4c). Nevertheless, temperature and RH conditions were considered appropriate for hen housing over all treatments and the control according to Oloyo (2018).

Carbon dioxide (CO_2) and methane (CH_4) emissions were quite similar between batches 1 and 2, respectively. CO_2 emission are in line with those cited by Shepherd et al. (2015) and Hayes et al. (2013) for aviary housing systems (cage-free systems), being 27.01 kg yr^{-1} hen^{-1} and 28.36 kg yr^{-1} hen^{-1}. Also, Fournel et al. (2012) observed similar CO_2 emissions in an experimental conventional system set up with a manure

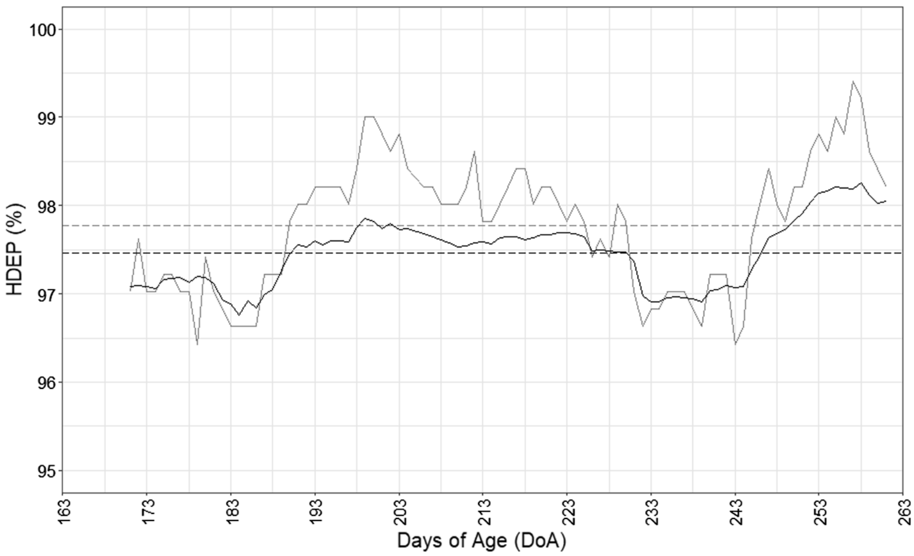

Fig. 14. Scenario analysis by increasing temperature by 5%. Original HDEP (grey line). Simulated data (black line). Original HDEP mean (grey dashed line). Simulated HDEP mean (black dashed line).

belt system and forced air drying. The authors reported emissions of about 28.7 kg yr^{-1} hen^{-1}. Average of CH_4 emissions were marginally higher than those reported by Shepherd et al. (2015), Hayes et al. (2013) and Fournel et al. (2012), where emissions were about 25.5 g yr^{-1} hen^{-1}, 26.0 g yr^{-1} hen^{-1} and 27.7 g yr^{-1} hen^{-1}, respectively. However, these emissions were similar to those cited by Alberdi et al. (2016) for enriched cages, and slightly lower than emissions reported by Zhu et al. (2011) for naturally ventilated cage layer housing, being 32.8 g yr^{-1} hen^{-1} and 40.8 g yr^{-1} hen^{-1}, respectively.

On the other hand, nitrous oxide (N_2O) and ammonia (NH_3) emissions were marginally lower in batch 1 than in batch 2. Emissions from both batches may differ due to a decrease in ventilation rate in batch 2 (Fig. 4c). N_2O emissions were lower than those cited by Zhu et al. (2011) in cage layer houses, being 3.43 g yr^{-1} hen^{-1}. However, emissions rates in N_2O for this experiment should be treated with caution, since measured N_2O concentrations were close to the atmospheric concentration threshold (0.331 ppm, WMO (2020)), limiting the detection of this gas inside the gas analyzer. This phenomena was also observed by Hayes et al. (2013) in aviary systems, where N_2O concentrations were excluded because of the low concentrations observed.

Besides, the average of NH_3 emissions were lower than those reported by Shepherd et al. (2015) and Hayes et al. (2013), being 40.8 g yr^{-1} hen^{-1} and 55.0 g yr^{-1} hen^{-1}, respectively. Lower values of NH_3 emissions should be treated also with prudence, because they were influenced by the number of housed animals, the ventilation rate applied, as well as the RH and temperature conditions used at laboratory scale. Hygrothermal conditions, gas emissions, as well as airborne dust emissions will be discussed in more details in a future companion paper. Furthermore, temperature, RH, ventilation rates, and gas concentrations did not draw from a normal distribution (SW_t: $p < 0.05$). For this reason, the non-parametric statistical method, *i.e.*, the Kruskal-Wallis Rank Sum Test, was applied for statistical analysis.

4.2. HDEP and EGC index

Daily fluctuations can be seen over the entire selected housing period. HDEP and EGC trends seem to have a marginally constant trend from 158 to 261 days of age (See Section 3.3.1). The constant trend over the production period was also observed by Ahmad (2011) for brown and white shelled commercial laying hens. Also, Ramírez-Morales et al. (2017) observed this constant behavior for daily egg production per bird in a selected flock (No. 8). It should be noted that higher HDEP index were observed comparing to the standard egg production curves for the Lohmann laying hen strain within the interval of 22–37 weeks. These higher values could be linked with the experimental setup and the animal density used in this study. In this case, the total laid eggs count was higher than the total number of animals housed inside the CFRs (12 hens per room).

On the other hand, higher fluctuations for treatment T3 could be associated with the increase in litter thickness produced by the application of the active adsorbent material (biochar). Thus, laying hens were prompted to lay eggs in the litter space. These laid eggs are more susceptible to break or get dirty and they have a direct impact on the EGC index values. Also, T3 was not replicated beyond the 200 days of age because of an interest in testing the last treatment with only oil sprinkling routines (data not shown), which probably influenced the data analysis.

Outcomes from the statistical test showed that the ECSs did not disrupt the laying hen performance in the experimental aviary system (Fig. 9). A distancing of the confidence region, in treatment T3, with respect to the other treatments, highlights a probable impact in egg cleanliness induced mainly by the litter space available for hens. Nevertheless, further studies are needed to assess this impact in cage-free layer systems.

4.3. HDEP modeling using Random Forest

In our study, moving averages were used with RF algorithms to predict HDEP daily fluctuations using hygrothermal and air quality characteristics as input data. The use of averaging (*i.e.*, sliding) windows showed a satisfactory HDEP modeling tool using Random Forest (RF). Window sizes of either 10 or 14 days seemed suitable and reliable regarding the Goodness of fit in HDEP predictions with RMSE below 0.45% and R^2 above 75% (Fig. 10). Moving averages were also adopted by Ramírez-Morales et al. (2017) to predict egg yield fluctuations using artificial neural networks and features related to egg production as inputs. However, the selection of a window size depends on two main factors, namely: model accuracy and on-farm practices and management. Indeed, egg producers should prefer short periods to determine whether or not there are anomalies in a commercial egg production facility. Thus, large windows could be risky in light of making decisions when facing sudden drops in the egg production curve.

Several statistical models have been used to predict egg production curves from laying houses. Ahmad (2011) applied a general regression neural network-predicted model to simulate egg production from a US commercial strain throughout an intermediate production period (from weeks 22 to 36) using feed consumption as an input variable. Artificial neural networks (ANN) were also applied by Ramírez-Morales et al. (2017) using a multilayer perceptron (MLP) along with sliding windows as an optimization methodology to determine anomalies in egg production. In their study, numbers of eggs, mortality, live animal, and cracked eggs were the relevant features used for modeling. As a result of their work, a warning classifier using a window size of 18 days was successfully developed with high accuracy. Akilli and Gorgulu (2020) aimed to compare multivariate, nonlinear, fuzzy regression techniques with a classical nonlinear regression to simulate egg yield curves, namely: ANN and least squares support vector machines (LSSVM). Akilli and Gorgulu (2020) successfully applied these two statistical models, getting high performance to estimate daily and weekly egg production values, showing suitable methods for early warning and egg production curve analysis. Other works have also been proposed to fit egg production curves using support vector machines and fuzzy logic (Gorgulu and Akilli, 2018; Morales et al., 2016; Omomule et al., 2020). This study presents the development of a 14-day RF model to satisfactory assess daily HDEP fluctuations in an experimental cage-free layer system, using non-conventional variables such as hygrothermal and air quality conditions (Fig. 11).

4.4. Variable importance and scenario analysis

It is noteworthy the temperature is one of the major factors in livestock production (Rojas-Downing et al., 2017). Livestock design and hygrothermal control systems have been subject of interest to ensure proper indoor environmental conditions for animals. Indoor temperature also has an important influence on animal heat transfer and energy consumption. Then, confined livestock units are mainly designed to manage temperature, and relative humidity since these variables govern heat stress and livestock production (Fournel et al., 2017; Rojas-Downing et al., 2017). Further, it has been observed that abrupt temperature changes could enhance adverse effects on bird health, mortality, feed intake, water consumption, body weight gain, or egg production (Oloyo, 2018). These parameters, along with naturally produced sensible and latent heat fluxes by laying hens in large flocks of commercial scale in Canada (*i.e.*, up to 400,000 hens), singled out indoor temperature as a key variable to monitor. It should be noted that the variable importance analysis was validated by the random variable, which did not reveal any importance among all predictors.

Indeed, the indoor temperature continues to be a relevant variable in climate control (combined with hygrothermal and air quality monitoring), considering that other variables, such as CO_2, NH_3, CH_4, or NO_2, are under the recommended thresholds for laying hen housing. Oloyo (2018) has also mentioned drops in egg production under high temperatures, highlighting the importance of controlling the temperature in laying hen houses. Furthermore, temperatures higher than 35 °C can influence heat stress events for animals, leading to a decrease in feed intake, egg quality and quantity, as well as health and well-being in hens (Xin et al., 2011). Nevertheless, continuing the RF model's training process with large data sets is deemed recommended to increase confidence in model accuracy when predicting outputs from inquiry-driven scenario analysis.

5. Conclusion

Environmental control strategies (ECS) are alternatives to reduce environmental impact from enriched and cage-free aviary housing systems. Evaluating outcomes, from applying these strategies, have become a subject of interest in identifying suitable laying hen production systems with low negative environmental effects and high profitability. In this study, an integral egg production analysis along with a statistical modeling application was performed to evaluate the effect of different ECSs on egg yield curves and predict HDEP daily fluctuations, highlighting possible correlations within indoor environmental and hygrothermal conditions. The proposed combined ECSs have potential applications in CFSs from a point of view of productivity since they have not shown negative effects on egg production and egg cleanliness measured for experimental CFS. Nevertheless, additional research is needed to evaluate the implementation of ECSs at the commercial scale. Meanwhile, moving averages can be useful to track HDEP trend, though sliding window sizes must be selected with caution because of the risk of dismissing sudden drops in egg production curves. These should also be selected depending on on-farm practices and egg producer needs.

Random Forest (RF) modeling was used in this study as another suitable approach among statistical models to generate early warning systems in egg production systems. The Goodness of fit of the RF model showed a satisfactory performance to predict daily fluctuations in egg yield using indoor environmental and hygrothermal variables. These results highlight the potential of machine learning approaches to deal with biosystems and their complexities. This machine learning application's contribution has the potential to be useful for egg production modeling and monitoring. Likewise, it is highly recommended to apply critical thinking to validate statistical models' outcomes with physical, chemical, or biological phenomena related to egg production systems. Besides, scenario analysis with RF represents a good framework to evaluate new potential applications. Nevertheless, large data sets are paramount to enhance the training process illustrating all possible events recorded in the past. This study contributes to the recent research filed on monitoring systems, and Precision Livestock Farming applied for laying hen production systems.

CRediT authorship contribution statement

Andrés F. Gonzalez-Mora: Conceptualization, Methodology, Software, Validation, Formal analysis, Visualization, Investigation, Data curation, Writing – original draft, Writing – review & editing. **Alain N. Rousseau:** Methodology, Validation, Visualization, Supervision, Project administration, Writing – review & editing. **Araceli D. Larios:** Conceptualization, Visualization, Supervision, Resources, Project administration, Investigation, Writing – review & editing. **Stéphane Godbout:** Conceptualization, Funding acquisition, Resources, Writing – review & editing. **Sébastien Fournel:** Supervision, Writing – review & editing.

Declaration of Competing Interest

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.

Acknowledgment

Our sincere thanks go to the *Ministère de l'Agriculture, Pêcheries et Alimentation* (MAPAQ) of Quebec - Canada, for the financial support through the *Innov'Action agroalimentaire* program. We would like to acknowledge the valuable contribution and assistance of technicians and fellow researchers from the livestock building agri-environmental assessment laboratory at IRDA. Also, we would like to thank Professor Silvio José Gumiere and a fellow graduate student Andres F. Silva at Université Laval, for suggestions and comments about use of Random Forest algorithms and statistical test for data analysis.

Funding

This work was supported by the *Ministère de l'Agriculture, des Pêcheries et de l'Alimentation* [Innov'Action program/IA117740] in the Quebec region – Canada and supported by the Quebec's Federation of Egg Producers (FPOQ).

Appendix A. . Random Forest model

Random Forest (RF) is a machine learning algorithm proposed by Breiman (2001). The decision trees and the bagging re-sampling techniques are the theoretical basis of this data-driven model. Decisions trees can be constructed by the classification and regression trees (CART) (Breiman et al., 1984). CART algorithm draws trees expanding first a very large quantity of nodes followed by a pruning step; that is removal of branches which reflect non-improvement in the prediction performance, using a pruning criterion. Nevertheless, RF based on CART uses unpruned trees and they grow until reaching a terminal node (Grömping, 2009).

Regression trees are well-defined if all the data have only numerical values, thus a numerical output will be obtained once the model is trained. On the other hand, bagging (or "bootstrap aggregating") is a re-sampling technique that creates several training subsets (or "bags") by randomly selecting the data using a with-replacement sampling; that is, a value could be taken more than once inside a bag. This method aims to increase the diversity in the input data used for the training process. Moreover, decisions trees are relatively unstable models in light of reproduced diversity in the response for small perturbations of data. In this case, no learning is performed at each training subset generated. Once the bag is created, the training process starts. Several decisions trees are generated for each training subset leading to a virtual forest. For each decision tree, about 1/3 of the training subset is left out which has no participation in the model development (known as the Out-of-Bag: OOB). Thus, model-predictive responses are averaged to provide an ensemble prediction of the target variable. An estimate error rate (OOB-MSE; Eq. (A.1)) associated with this final prediction is calculated using the OOB on the training set to estimate the accuracy of the RF model (Breiman, 2001).

$$\boldsymbol{OOB}-\boldsymbol{MSE}=\frac{1}{\boldsymbol{nN}}\sum_{j=1}^{N}\sum_{i=1}^{n}\left(y_{i,j}-\mathring{A}\cdot_{i,j}\right)^{2} \tag{A.1}$$

Where $y_{i,j}$ is the observation i at the j regression tree, $\mathring{A}\cdot_{i,j}$ is the prediction from observation i at the same regression tree, n is the number of observations in the OOB and N is the total number of trees, while MSE stands for mean squared error.

References

Aarnink, A.J.A., van Harn, J., Van Hattum, T.G., Zhao, Y., Ogink, N.W.M., 2011. Dust reduction in broiler houses by spraying rapeseed oil. Am. Soc. Agric. Biol. Eng. (ASABE) 54, 1479–1489.
Agriculture and Agri-Food Canada, 2019a. Canada's poultry and egg industry profile. Canada, G.o. (Ed.).
Agriculture and Agri-Food Canada, 2019b. Canada's table and processed egg industry. Government of Canada.
Ahmad, H.A., 2011. Egg production forecasting: Determining efficient modeling approaches. J. Appl. Poult. Res. 20 (4), 463–473.
Akilli, A., Gorgulu, O., 2020. Comparative assessments of multivariate nonlinear fuzzy regression techniques for egg production curve. Trop. Anim. Health Prod. 52 (4), 2119–2127.
Alberdi, O., Arriaga, H., Calvet, S., Estellés, F., Merino, P., 2016. Ammonia and greenhouse gas emissions from an enriched cage laying hen facility. Biosyst. Eng. 144, 1–12.
Alcamo, J., 2008. Environmental futures: the practice of environmental scenario analysis. Elsevier.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. Chapman & Jall/CRC, Florida, USA.
Chai, L., Xin, H., Zhao, Y., Wang, T., Soupir, M., Liu, K., 2018. Mitigating Ammonia and PM Generation of Cage-Free Henhouse Litter with Solid Additive and Liquid Spray. Trans. ASABE 61, 287–294.
David, B., Mejdell, C., Michel, V., Lund, V., Moe, R., 2015a. Air quality in alternative housing systems may have an impact on laying hen welfare. Part II—Ammonia. Animals 5 (3), 886–896.
David, B., Moe, R., Michel, V., Lund, V., Mejdell, C., 2015b. Air quality in alternative housing systems may have an impact on laying hen welfare. Part I—Dust. Animals 5 (3), 495–511.
Dray, S., Dufour, A.-B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22, 1–20.
Egg Farmers of Canada, 2018. PD-PT^{MC} Lignes directives à l'intention des producteurs (pondeuses), In: Canada, A.a.A. (Ed.), https://oeuf.ca/wp-content/uploads/2020/02/sc-sc-farmer-guidelines-layer-v2.0-rev-2.4-2020-01-01-fre.pdf.
Fournel, S., Pelletier, F., Godbout, S., Lagacé, R., Feddes, J., 2012. Greenhouse gas emissions from three cage layer housing systems. Animals 2 (1), 1–15.
Fournel, S., Rousseau, A.N., Laberge, B., 2017. Rethinking environment control strategy of confined animal housing systems through precision livestock farming. Biosyst. Eng. 155, 96–123.
Godbout, S., Larios, A.D., Duchaine, C., Létourneau, V., Belzile, L., Li, J., Morin, C., Lavallée, H., Gonzalez-Mora, A.F., St-Germain, M.-W., Fournel, S., Palacios, J.H., Gordon, C., 2020. Environnement, santé et bien-être en élevage ovocole alternatif : Phase 1 : Mitigation. Rapport final. Institut de recherche et de développement en agroenvironnement Inc. (IRDA), Québec, Canada, p. 30.
Gonzalez-Mora, A.F., Larios, A.D., Rousseau, A.N., Godbout, S., Morin, C., Palacios, J.H., Grenier, M., Fournel, S., 2021. Assessing environmental control strategies in cage-free egg production systems: Effect on spatial occupancy and natural behaviors. Animals 11, 1–13.
Gorgulu, O., Akilli, A., 2018. Egg production curve fitting using least square support vector machines and nonlinear regression analysis. Europ. Poult. Sci. 82.
Grömping, U., 2009. Variable importance assessment in regression: linear regression versus random forest. Am. Stat. 63, 308–319.
Hayes, M., Xin, H., Li, H., Shepherd, T., Zhao, Y., Stinn, J., 2013. Ammonia, greenhouse gas, and particulate matter emissions of aviary layer houses in the Midwestern US. Trans. ASABE 56, 1921–1932.
Holt, P., Davies, R., Dewulf, J., Gast, R., Huwe, J., Jones, D., Waltman, D., Willian, K., 2011. The impact of different housing systems on egg safety and quality. Poult. Sci. 90, 251–262.
Kuhn, M., 2020. caret: Classification and Regression Training, In: 6.0-86, R.p.v. (Ed.).
Kuhn, M., Johnson, K., 2013. Applied predictive modelling. Springer Verlag New York.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. The Newsletter of the R Project. R Project, R news, pp. 18–22.
Lin, X., Zhang, R., Jiang, S., El-Mashad, H., Xin, H., 2017. Emissions of ammonia, carbon dioxide and particulate matter from cage-free layer houses in California. Atmos. Environ. 152, 246–255.
Lin, X., Zhang, R., Jiang, S., El-Mashad, H.M., Xin, H., 2018. Fan and ventilation rate monitoring of cage-free layer houses in California. Trans. ASABE 61, 1939–1950.
Misslin, C., 2017. Le suivi d'élevage en filière poule pondeuse: de l'accouvage à la production d'oeufs. Université Claude-Bernard -, Lyon I, p. p. 138 p..
Morales, I.R., Cebrián, D.R., Blanco, E.F., Sierra, A.P., 2016. Early warning in egg production curves from commercial hens: A SVM approach. Comput. Electron. Agric. 121, 169–179.
Narinc, D., Uckardes, F., Aslan, E., 2014. Egg production curve analyses in poultry science. World's Poult. Sci. J. 70, 817–828.
Niloofar, P., Francis, D.P., Lazarova-Molnar, S., Vulpe, A., Vochin, M.-C., Suciu, G., Balanescu, M., Anestis, V., Bartzanas, T., 2021. Data-driven decision support in livestock farming for improved animal health, welfare and greenhouse gas emissions: Overview and challenges. Comput. Electron. Agric. 190, 106406.
Oloyo, A., 2018. The use of housing system in the management of heat stress in poultry production in hot and humid climate: a review. Poult. Sci. J. 6, 1–9.
Omomule, T.G., Ajayi, O.O., Orogun, A.O., 2020. Fuzzy prediction and pattern analysis of poultry egg production. Comput. Electron. Agric. 171, 105301.
Parent, S.-É., 2020. Analyse et modélisation d'agroécosystèmes.
Pelletier, N., Doyon, M., Muirhead, B., Widowski, T., Nurse-Gupta, J., Hunniford, M., 2018. Sustainability in the Canadian Egg Industry—Learning from the Past, Navigating the Present, Planning for the Future. Sustainability 10, 3524.
Philippe, F., Mahmoudi, Y., Mars, D.C., Lefrançois, M., Moula, N., Palacios, J., Pelletier, F., Godbout, S., 2020. Comparison of egg production, quality and composition in three production systems for laying hens. Livestock Sci. 103917.
Polikar, R., 2006. Ensemble based systems in decision making. IEEE Circuits Syst. Mag. 6, 21–45.
Prasai, T.P., Walsh, K.B., Midmore, D.J., Jones, B.E.H., Bhattarai, S.P., 2018. Manure from biochar, bentonite and zeolite feed supplemented poultry: Moisture retention and granulation properties. J. Environ. Manage. 216, 82–88.
Qasim, W., Moon, B.E., Phonsuwan, M., Jo, J.S., Lee, M.H., Nafees, M., Kim, H.T., 2017. Effects of an aluminum sulfate and ferric chloride blend on poultry litter characteristics in vitro. J. Appl. Poult. Res. 27, 92–102.
R Core Team, 2020. R: A language and environment for statistical computing, In: Computing, R.F.f.S. (Ed.), 4.0.0 ed, Vienna, Austria.

Ramírez-Morales, I., Fernández-Blanco, E., Rivero, D., Pazos, A., 2017. Automated early detection of drops in commercial egg production using neural networks. Br. Poult. Sci. 58, 739–747.

Ribeiro, R., Casanova, D., Teixeira, M., Wirth, A., Gomes, H.M., Borges, A.P., Enembreck, F., 2018. Generating action plans for poultry management using artificial neural networks. Comput. Electron. Agric.

Rojas-Downing, M.M., Nejadhashemi, A.P., Harrigan, T., Woznicki, S.A., 2017. Climate change and livestock: Impacts, adaptation, and mitigation. Clim. Risk Manage. 16, 145–163.

Schneider, A.F., Almeida, D.S.D., Yuri, F.M., Zimmermann, O.F., Gerber, M.W., Gewehr, C.E., 2016. Natural zeolites in diet or litter of broilers. Br. Poult. Sci. 57, 257–263.

Shepherd, T.A., Zhao, Y., Li, H., Stinn, J.P., Hayes, M.D., Xin, H., 2015. Environmental assessment of three egg production systems—Part II. Ammonia, greenhouse gas, and particulate matter emissions. Poult. Sci. 94, 534–543.

Shields, S., Shapiro, P., Rowan, A., 2017. A decade of progress toward ending the intensive confinement of farm animals in the United States. Animals 7, 40.

Tong, X., Hong, S.-W., Zhao, L., 2019. CFD modeling of airflow, thermal environment, and ammonia concentration distribution in a commercial manure-belt layer house with mixed ventilation systems. Comput. Electron. Agric. 162, 281–299.

van Harn, J., Aarnink, J.A.A., Mosquera, J., van Riel, W.J., Ogink, N.W.M., 2012. Effect of Bedding Material on Dust and Ammonia Emission from Broiler Houses. Trans. ASABE 55, 219–226.

Winkel, A., Mosquera, J., Aarnink, A.J.A., Groot Koerkamp, P.W.G., Ogink, N.W.M., 2016. Evaluation of oil spraying systems and air ionisation systems for abatement of particulate matter emission in commercial poultry houses. Biosyst. Eng. 150, 104–122.

WMO, 2020. Greenhouse gas concentrations, in atmosphere reach yet another high. World Meteorological Organization.

Xin, H., Gates, R.S., Green, A.R., Mitloehner, F.M., Moore, P., Wathes, C.M., 2011. Environmental impacts and sustainability of egg production systems. Poult. Sci. 90, 263–277.

Yilmaz Dikmen, B., İpek, A., Şahan, Ü., Petek, M., Sözcü, A., 2016. Egg production and welfare of laying hens kept in different housing systems (conventional, enriched cage, and free range). Poult. Sci. 95, 1564–1572.

Zhang, C., Weiss, A., Lin, C., Li, H., Joerger, R., Chiu, P., 2016. Effects of Multiple Litter Amendment Applications in Commercial Broiler Houses on Ammonia Emissions and Litter Microflora. Trans. ASABE 59, 1393–1401.

Zhu, Z., Dong, H., Zhou, Z., Xin, H., Chen, Y., 2011. Ammonia and greenhouse gases concentrations and emissions of a naturally ventilated laying hen house in Northeast China. Trans. ASABE 54, 1085–1091.