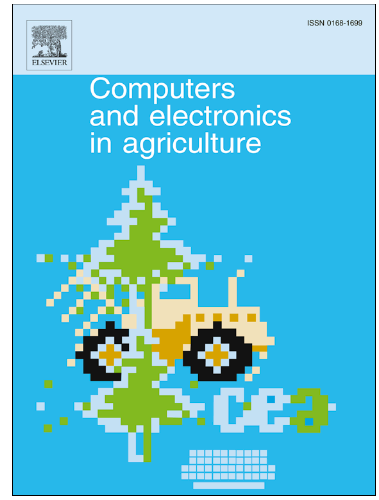
Accepted Manuscript



Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming

Alberto Diez-Olivan, Xavier Averós, Ricardo Sanz, Basilio Sierra, Inma Estevez

PII: S0168-1699(17)31439-4
DOI: https://doi.org/10.1016/j.compag.2018.03.025
Reference: COMPAG 4228

To appear in: *Computers and Electronics in Agriculture*

Received Date: 20 November 2017
Revised Date: 12 March 2018
Accepted Date: 19 March 2018

Please cite this article as: Diez-Olivan, A., Averós, X., Sanz, R., Sierra, B., Estevez, I., Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming, *Computers and Electronics in Agriculture* (2018), doi: https://doi.org/10.1016/j.compag.2018.03.025

Quantile regression forests-based modeling and environmental indicators for decision support in broiler farming

Alberto Diez-Olivan[a,*], Xavier Averós[b], Ricardo Sanz[c], Basilio Sierra[d], Inma Estevez[b,e]

[a] *Tecnalia Research & Innovation, Industrial Systems Unit, Donostia-San Sebastián, Spain*
[b] *Neiker-Tecnalia, Department of Animal Production, Vitoria-Gasteiz, Spain*
[c] *Autonomous Systems Laboratory, Universidad Politécnica de Madrid, Spain*
[d] *Department of Computer Sciences and Artificial Intelligence, UPV/EHU, Spain*
[e] *Ikerbasque, Basque Foundation for Science, Bilbao, Spain*

Abstract

An efficient and sustainable animal production requires fine-tuning and control of all the parameters involved. But this is not a simple task. Animal farming is a complex biological system in which environmental parameters and management practices interact in a dynamic way. In addition, the typical non-linear response of biological processes implies that relationships across parameters that are critical to assure animal welfare and performance are difficult to determine. In this paper a novel decision support system based on environmental indicators and on weights, leg problems and mortality rates is proposed to address this issue. The data-driven modeling process is performed by a quantile regression forests approach that allows estimating growth, welfare and mortality parameters on the basis of environmental

*Corresponding author:
Email address: alberto.diez@tecnalia.com (Alberto Diez-Olivan)

deviations from optimal farm conditions. Resulting models also provide confidence intervals able to deal with uncertainty. They are deployed in farm, offering an accessible tool for farmers, veterinarians and technical personnel. Experimental results involving 20 flocks of broiler meat chickens from different farms show the validity of the system, obtaining robust prediction intervals and high accuracy, namely over 81% for every model. The in-field use of the proposed approach will facilitate an efficient and animal welfare-friendly production management.

Keywords: Data processing, Random forests, Broiler meat chicken, Efficient production, Animal welfare, Machine learning.

---

1. Motivation

Broiler meat chickens are the most abundant farmed animal in the European Union (EU), and a key component of EU food supply. Much data is already collected by the main actors involved in the process, predominately at slaughter plant level [1]. Data acquisition in itself is, however, of little value unless collected data are standardized at national and international levels and further processed to produce useful information to improve broiler health and welfare, and hence offering the potential to reduce antimicrobial use, a real and current concern for producers and legislators.

The broiler meat chicken industry is the second largest meat industry in the world. Yearly, 70 billion birds are produced around the world under similar, well established management practices with similar genetic stocks produced majorly by two international companies; Cobb-Vantress (Cobb) and

[1] Under the Broiler Directive 2007/43/EC

Aviagen (Ross). On-farm environmental models put in practice nowadays (e.g. indoors temperature and relative humidity) are based on the theoretical curves proposed by this two companies [1]. However, the large production volumes and the great complexity of the production chain, implies that the possibility to control system parameters to optimal values is probably fictitious. Additionally, deviation of values from optimality may go undetected due to different reasons including the lack/misuse of information at various stages, or late problem detection. Such circumstances may lead to a loss of efficiency of the system. Lack of an adequate control during each production process step may also lead to impaired animal health and welfare, and to the occurrence of meat quality and safety problems. Therefore, system adjustments to maximize, for instance, energy efficiency, may easily derive in unforeseen animal health or welfare issues. In addition, the economic perspective must also be considered as potential changes in management practices may be impossible in practice when all inputs and outputs of the system, many of them of different nature and coming from diverse sources, are simultaneously considered.

In order to provide the broiler meat chicken industry with an intelligent system able to support an efficient and sustainable broiler production, most recent technological approaches must be put together with traditional broiler production. The model of sustainable production also needs to be based on considerations for animal welfare and environmental responsibility. Advanced knowledge-based, empirical models to optimally manage broiler processes are needed. They are able to provide the players of the production chain such as farmers, veterinarians and technical personnel with key recommendations to optimize production and to design better production strategies. Machine learning algorithms are usually applied to automatically generate such data-driven knowledge models. To do so, features of interest must be identified according to best management practices and most recent scientific knowledge in broiler production, health and welfare. Resulting models can be then deployed in a Decision Support System (DSS) that should be able to provide the necessary tools to assure efficient and sustainable production according to a social responsibility-based production model.

This work aims at developing a DSS that uses environmental parameters automatically collected for each corresponding flock, together with additional live production health and welfare data. This work will allow to generate intelligent practical tools to improve flock performance by better managing the health and welfare of broiler flocks. Our main contribution is to provide such tools on the basis of environmental indicators and a quantile regression forests-based approach. Resulting growth, welfare and mortality models are robust and comprehensive, yet accurate decision support tools in animal farming.

The rest of the article is organized as follows. Section 2 presents a review of related and previous works in animal farming DSS for improving broiler flocks management, and the position of the present work in the context of previous ones. Section 3 explains the proposed DSS, based on quantile regression forests modeling of environmental indicators and growth, welfare and mortality parameters. In Section 4 the test scenario is presented and the experimental results obtained are discussed. Finally, the conclusions achieved in this study are given in the last section.

2. Research background

Historically, advances in animal production have been sustained by mono-disciplinary approaches to the resolution of single problems, mostly focused on increasing animal performance (e.g. considering business aspects by a participatory approach [2]). Nevertheless, current needs of industry demand not only to produce efficiently, but also to adopt production models based on social and environmental responsibility, which will be essential to maintain long term consumers' acceptability and will allow increased competitiveness in the international market. Such models comprise critical aspects such as environmental stewardship (e.g. by optimizing resource use, reducing wastage of resources and greenhouse gas emissions [3] [4]) and animal health-welfare responsibility (e.g. by assuring management practices that optimize animal health and welfare, and therefore increasing meat quality and safety [5] [6]). These aspects are pivotal for the sustainability of the poultry industry or any livestock production.

Advances in animal production have traditionally targeted genetics, nutrition, health, and housing systems, the pillars of production. However, modern animal production must also consider sustainability and therefore product quality and ethical aspects of social responsibility that will influence consumers' acceptance [7].

Farming optimization would greatly benefit from the latest technological advances, allowing the fine-tuning of systems according to sustainable production models. Precision Livestock Farming (PLF), defined as the management of livestock production using the principles and technology of process engineering [8], will play a basic role [9], creating exciting opportunities to accelerate the development of modern farming systems [10] [11] [12]. For instance the use of integrated, innovative decision tools and automation technologies to reduce the variability in system outputs such as product quality [7], would help improving animal management, productivity and benefits, reduce costs and improve animal welfare and human well-being [13] [12] [14]. PLF has a tremendous potential in achieving these objectives [15], developing systems for real-time animal management [7]. PLF can also provide opportunities for a transparent quality control system of the whole chain from farming to retail [9]. Some authors indicated that PLF approach should be easy to implement to animal production [16], and some applications of PLF technologies, as well as commercial applications within the animal production sector are available [12] [17] [11]. However, PLF applied to farm animals is complex due to aspects regarding animal sentience [8].

Optimal information on animal welfare must be obtained by smartly combining system variables and animal-based indicators [18]. System variables are easily collected and some useful indicators can be inferred from them, but animal-based data collection is challenging, especially in broilers where thousands of animals are housed in one location. In this sense, a novel approach has been proposed, based on transect walk methodology, to assess broiler flock welfare [19]. Due to the relevance of animal welfare in current livestock production systems, and the high relevance given by the EU, other complex systems are being developed, primarily focusing on monitoring animal welfare in itself [20].

Data management after collection is perhaps the main factor determining the power of PLF systems [21], giving the opportunity to create basic early warning systems. However, the development of decision support tools for animal farming that also consider animal welfare is still a very challenging issue [17]. A knowledge-driven DSS must be supported by advanced data processing techniques able to provide powerful problem solving tools and useful recommendations [22] [23]. These techniques may vary depending on the nature and characteristics of the data to be analysed. To this regard Artificial Neural Networks (ANN) models have been used in a wide variety of applications for many years [24]. In [25], for instance, an approach based on a combination of ANNs and Fuzzy_AHP for heart failure risk prediction is proposed. Indeed, models based on Fuzzy Logic are very popular when implementing a DSS. They provide a mapping between real-valued input and output parameters as represented by understandable fuzzy rules [26] [27]. Some recent works in that direction can be found in [28] and [29], for the particular case of rough sets analysis.

Optimization of efficiency for sustainable production requires a wider perspective involving all parameters and factors that contribute to the system. It requires a fine-tuning and an efficient functioning of all the steps of the production chain. Therefore, system adjustments to maximize, for instance, energy efficiency, may easily derive in unforeseen animal health or welfare issues. Economic benefits of increasing production density in broilers might increase mortality and reduce meat quality [30]. Thus, perhaps a more efficient economic model would contemplate, for instance, to increase production by means of a strategic management aiming at reducing on-farm mortalities or deaths on arrival at the slaughterhouse.

The DSS concept in animal farming must be then designed as a data-driven approach to maximize productive efficiency considering animal welfare and environmental responsibility. The main objective of this research is to provide farmers and experts with an intelligent decision-making tool for an efficient and animal welfare-friendly production management, which is based on data-driven methods and machine learning algorithms.

3. Proposed decision support system

This section presents the key parts of the proposed DSS, whose aim is to provide the meat chicken industry with an intelligent system able to support an efficient and sustainable broiler production. They are further discussed in the next subsections and they are the following:

- Automatic data acquisition and cloud-based storage capabilities.
- Environmental indicators, which are defined as deviations from optimal environmental conditions, previously learned, over time. The environmental conditions are measured indoors, inside the farm, and for each parameter there are three sensors located around the farm in order to minimise the variations in environmental conditions given in different locations.
- Quantile regression forests-based approach used to model data and make predictions, dealing with uncertainty.
- Empirical data-driven generation of growth, welfare and mortality models on the basis of quantile regression forests method and on empirical

data corresponding to environmental indicators as predictor variables and on weights, leg problems and mortality rates as target features. These key parameters used as target features have been chosen due to their importance regarding the broiler production, but others can be also used in order to establish the optimal farm conditions.

The system architecture consists of a set of sensors installed to automatically measure and collect different environmental conditions in farm, a cloud-based storage of such information and a set of machine learning-based models learned from historical data to manage the production smartly and in an online fashion.

3.1. Data acquisition and cloud-based storage

Data acquisition devices measure environmental parameters every 15 minutes for each flock of birds from farm arrival to the end of production. For each parameter, there are three sensors located around the farm in order to minimise the variations in environmental conditions given in different locations. Collected data is automatically transferred to a cloud-based service, where it is stored for further analysis. The cloud storage service used in this research work is owncloud [2].

Growth, welfare and mortality parameters (e.g. weights, leg problems and mortality rate) are periodically acquired by experts by transect walk methodology [19] and other similar techniques, due to the complexity of the process and the expertise needed:

[2]https://owncloud.org/

- Growth parameter: weights are collected during week 1, 3, 5 and 6 of age on a representative sample of 50 birds per flock and their weights are averaged to produce an average sample weight, as established in [31]. Weights at arrival are registered but they are not representative for the study since the weights variance and the cumulative deviations in relation to the environmental parameters at that stage of the production process are meaningless.
- Welfare parameter: frequency (%) of lame and immobile birds are considered per transect, which are collected using the transect method as described in [19].
- Mortality parameter: the mortality rate is considered, which is calculated using farmer records from day 1 to slaughter and it includes birds found dead and culled every day.

All this information is also integrated in the cloud-based data system.

3.2. Environmental indicators

Information given by environmental conditions can support predictions made since they have serious effects in growth and animal welfare. Since the acquisition frequency regarding the environmental parameters is much higher than in the case of growth, welfare and mortality parameters, environmental data, $\boldsymbol{env} = \{env_1, ..., env_m\}$, are re-sampled in order to have an average value per hour, $\langle \boldsymbol{env} \rangle = \{\langle env \rangle_1, ..., \langle env \rangle_{n<m}\}$, where $\langle \cdot \rangle$ denotes the average over the sampled data.

Observations that are far from the mean are filtered out, based on 3 standard deviations from the mean, $\mu_{\langle \boldsymbol{env} \rangle}$. They are bad readings caused by hardware issues and can strongly affect the resulting environmental indicators. The resulting relative humidity model from the set of 20 farms used in this study is shown in Figure 1. It consists of the average value and the related upper and lower established limits after filtering out outliers, and it covers the whole production period, from farm arrival to the end of production.

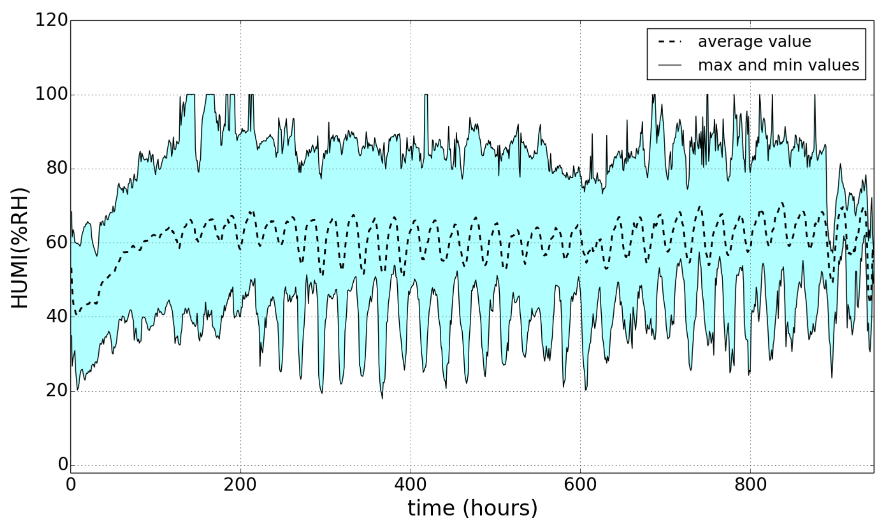


Figure 1: Relative humidity model learned from the set of 20 farms used in this study, from farm arrival to the end of production.

Cumulative deviations from environmental models over time, $\boldsymbol{d} = (d_1, ..., d_n)$ are then computed in order to provide useful indicators. They are calculated for values above, $\boldsymbol{d^a}$, and below, $\boldsymbol{d^b}$, the model, as it can be seen in Equation

1.

$$d_i = \begin{cases} d_i^a = \sum_{j=1}^{i}(\langle env\rangle_j - \langle env\rangle_j') & \forall\ \langle env_j\rangle \geq \langle env\rangle_j' \\ d_i^b = \sum_{j=1}^{i}(\langle env\rangle_j' - \langle env\rangle_j) & \forall\ \langle \mathrm{env_j}\rangle \leq \langle \mathrm{env}\rangle_\mathrm{j}' \end{cases} \tag{1}$$

where $\boldsymbol{\langle env\rangle'} = (\langle env\rangle_1', ..., \langle env\rangle_i')$ are the environmental model values and $\boldsymbol{\langle env\rangle} = (\langle env\rangle_1, ..., \langle env\rangle_i)$ are the current, real environmental values until day $i = 1, ..., n$, respectively.

It is assumed that large cumulative deviations from normal environmental conditions over time have a big impact in the growth process and welfare of the birds and in production [32].

3.3. Quantile Regression Forests

Random forests is an ensemble method that grows an ensemble of trees [33]. It employs averaging to improve the predictive accuracy and control over-fitting. The sub-sample size is always the same as the original input sample size but the samples are drawn with replacement (bootstrap). A large number of trees can be therefore grown. Random forests for regression analysis, or regression forests, are an ensemble of different regression trees in which each leaf draws a distribution for the continuous target feature, $\boldsymbol{y} = (y_1, ..., y_n)$. More precisely, given a set of m features, $X = \{X_1, ..., X_m\}$, where each feature X_i can take a value from its own set of possible values χ_i, and n feature vectors or instances, $\boldsymbol{x_i} = (x_1, ..., x_m) \in \chi = (\chi_1, ..., \chi_m)$, with $i = 1, ..., n$, a random forest is a collection of K tree predictors $T(\theta_k)$, with $k = 1, ..., K$, being θ_k the random parameter vector that determines how the k-th tree is grown, or which features are considered to split on at each node when approximating $\boldsymbol{y}$. For each tree and each node, randomness is employed when selecting a feature to split on and for each tree, a bagged version of the n feature vectors is used. In addition, only a random subset of the predictor features is considered for splitpoint selection at each node.

It is assumed that X and θ_k are independent and identically distributed, and tuples $(\boldsymbol{x_i}, y_i)$ are independently drawn from the joint distribution.

Every leaf of the tree, $l = 1, ..., L$, corresponds to a rectangular subspace of χ denoted as $R_l \subseteq \chi$. Then for every $\boldsymbol{x_i}$ there is one and only one leaf $l(\boldsymbol{x_i}, \theta)$ for tree $T(\theta)$ such that $\boldsymbol{x_i} \in R_l$.

For a new feature vector, $\boldsymbol{x_{new}}$, the prediction of a single tree $T(\theta)$ is the average of the observed values in leaf $l(\boldsymbol{x_{new}}, \theta)$. Let the weighted vector $w_i(\boldsymbol{x_{new}}, \theta)$ be defined as:

$$w_i(\boldsymbol{x_{new}}, \theta) = \begin{cases} 1/\#\{j : \boldsymbol{x_j} \in \chi_{l(\boldsymbol{x_{new}}, \theta)}\} & \text{if } \boldsymbol{x_i} \in \chi_{l(\boldsymbol{x_{new}}, \theta)} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

The prediction of a single tree can be then computed as the weighted average of the target feature values, $\boldsymbol{y} = (y_1, ..., y_n)$, as it is shown in Equation 3.

$$\hat{\mu}(\boldsymbol{x_{new}}) = \sum_{i=1}^{n} w_i(\boldsymbol{x_{new}}, \theta) y_i \tag{3}$$

In the case of random forests, Equation 3 is generalized as the average prediction of K single trees, as it can be seen in Equation 4.

$$\hat{\mu}(\boldsymbol{x_{new}}) = \sum_{i=1}^{n} K^{-1} \sum_{k=1}^{K} w_i(\boldsymbol{x_{new}}, \theta_k) y_i \tag{4}$$

By means of the Quantile Regression Forests approach, confidence intervals can be obtained from predictions made by random forests [34]. It provides valuable information about the dispersion of observations around the predicted value, reinforcing the reliability of predictions made. Instead of recording the mean value of responses given in each tree leaf in the forest, all observed responses in the leaf are recorded. The prediction thus becomes the full conditional distribution $P(\boldsymbol{y} \leq y_i|X = \boldsymbol{x_i})$, with $i = 1, ..., n$, given by the probability that, for $X = \boldsymbol{x_i}$, $\boldsymbol{y} < y_i \in \mathbb{R}$.

The corresponding conditional distribution function $F(\boldsymbol{y}|X = \boldsymbol{x_i})$ can be also expressed as $E(1_{\{\boldsymbol{y} \leq y_i\}}|X = \boldsymbol{x_i})$, which is approximated by the weighted mean over the observations of $1_{\{\boldsymbol{y} \leq y_i\}}$, as it can be seen in Equation 5.

$$\hat{F}(\boldsymbol{y}|X = \boldsymbol{x_i}) = \sum_{i=1}^{n} w_i(\boldsymbol{x_i}) 1_{\{\boldsymbol{y} \leq y_i\}} \tag{5}$$

where $w_i(\boldsymbol{x_i}) = K^{-1} \sum_{k=1}^{K} w_i(\boldsymbol{x_i}, \theta_k)$ is the weighted vector.

The α-quantile, $Q_\alpha(\boldsymbol{x_i})$ is defined such that the probability of $\boldsymbol{y} < Q_\alpha(\boldsymbol{x_i}) = \alpha$. The quantiles give more complete information about the distribution of $\boldsymbol{y}$ as a function of the predictor features, X, than the conditional mean $E(\boldsymbol{y}|X = \boldsymbol{x_i}) = \underset{z}{\operatorname{argmin}}\, E\{(\boldsymbol{y} - z)^2|X = \boldsymbol{x_i}\}$.

Given a new feature vector, $\boldsymbol{x_{new}}$, the estimate of the distribution function in 5 is computed and prediction intervals are created by simply applying the appropriate percentiles of the distribution. A 95% prediction interval for the value of $\boldsymbol{y}$, for instance, will be given by Equation 6.

$$I(\boldsymbol{x_i}) = [Q_{.025}(\boldsymbol{x_i}), Q_{.975}(\boldsymbol{x_i})] \tag{6}$$

Quantile regression forests approach can therefore be used to create prediction intervals that contain very useful information about the dispersion of observations around the predicted value. Besides, quantile regression estimates are more robust in presence of outliers and uncertainty in data.

3.4. Empirical data-driven models

Environmental indicators are used as predictor variables and weights, leg problems and mortality rates are used as target features to generate empirical data-driven growth, welfare and mortality models on the basis of quantile regression forests approach. They are able to make useful predictions and to finally provide optimal decision support in animal farming.

The growth parameter is given by the average of the weights at weeks 3, 5 and 6, $\langle weight_s \rangle$, being $s = 3, 5, 6$. The idea is to anticipate heavy deviations during the growth process on the basis of deviations from optimal environmental values, modifying the farm conditions accordingly in advance. The weight parameter values observed during the growth process in the set of 20 farms used in this study can be seen in Figure 2.

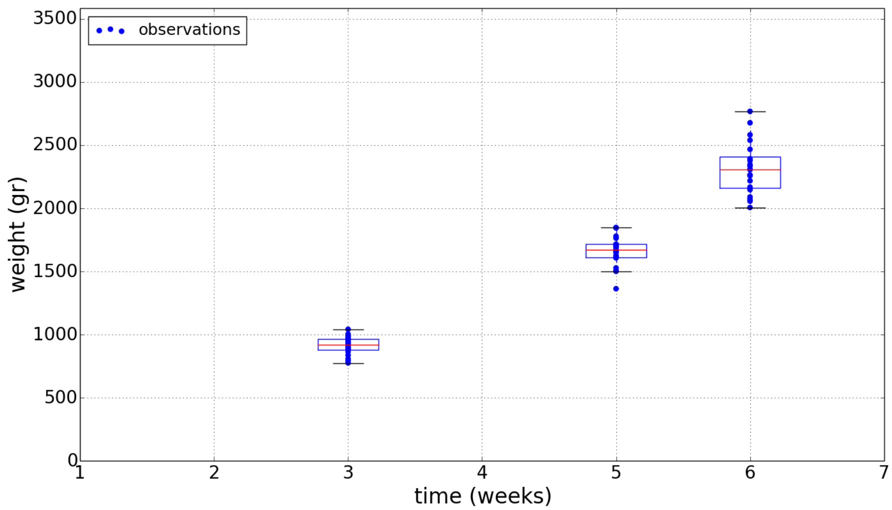


Figure 2: Weight parameter values observed during the growth process in the set of 20 flocks used in this study. Each dot represents the average weight value of a representative sample of 50 birds per flock, from farm arrival to the end of production.

Then, the growth model, $\langle \boldsymbol{weight_s} \rangle \approx G(d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms})$, is defined as it can be seen in Equation 7, being $(d_1, ..., d_m)$ the deviations from the m environmental parameters under study.

$$\hat{G}(\langle \boldsymbol{weight_s} \rangle | X = \{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}) =$$

$$\sum_{i=1}^{n} w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}) 1_{\{\langle \boldsymbol{weight_s} \rangle \leq \langle weight_{si} \rangle\}} \tag{7}$$

with $w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}) = K^{-1} \sum_{k=1}^{K} w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}, \theta_k)$ being the weighted vector.

Furthermore, the proportion of occurrence of leg problems (lame and immobile birds) over the total population (%), lp, at weeks 3, 5 and 6 of the entire growth process, is established as the welfare parameter. Figure 3 shows the welfare parameter values that were observed during the growth process in the set of 20 flocks used in this study.

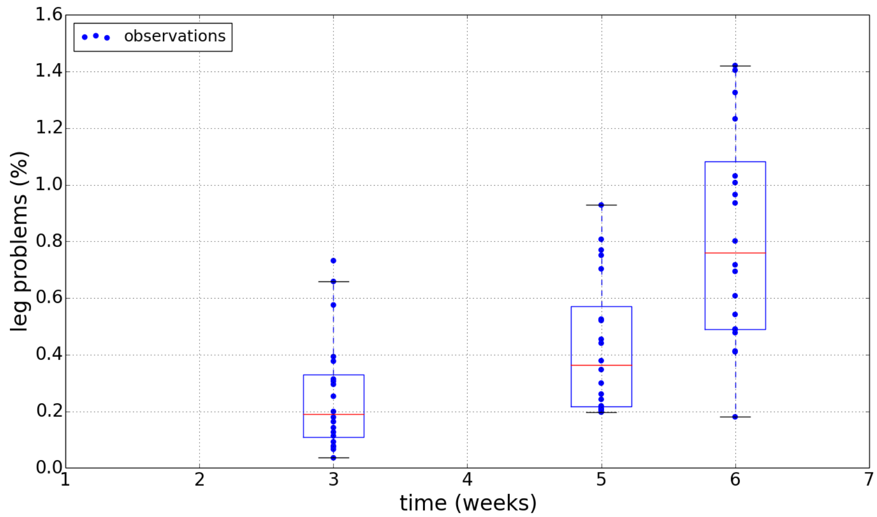


Figure 3: Welfare parameter values observed during the growth process in the set of 20 flocks used in this study, from farm arrival to the end of production.

Similarly as in the case of G, the welfare model, $\boldsymbol{lp_s} \approx W(d_{1s}^a, d_{1s}^b, ..., d_{ms}^a, d_{ms}^b)$, becomes (see Equation 8).

$$\hat{W}(\boldsymbol{lp_s}|X = \{d_{1s}^a, d_{1s}^b, ..., d_{ms}^a, d_{ms}^b\}) =$$

$$\sum_{i=1}^{n} w_i(\{d_{1s}^a, d_{1s}^b, ..., d_{ms}^a, d_{ms}^b\}) 1_{\{\boldsymbol{lp_s} \leq lp_{si}\}} \tag{8}$$

with $w_i(\{d_{1s}^a, d_{1s}^b, ..., d_{ms}^a, d_{ms}^b\}) = K^{-1} \sum_{k=1}^{K} w_i(\{d_{1s}^a, d_{1s}^b, ..., d_{ms}^a, d_{ms}^b\}, \theta_k)$ being the weighted vector.

Finally, the mortality parameter is given by the cumulative, proportion of mortality over the total population (%), $mort$, at weeks 3, 5 and 6. The corresponding mortality parameter values over time in the set of 20 flocks used in this study can be seen in Figure 4.

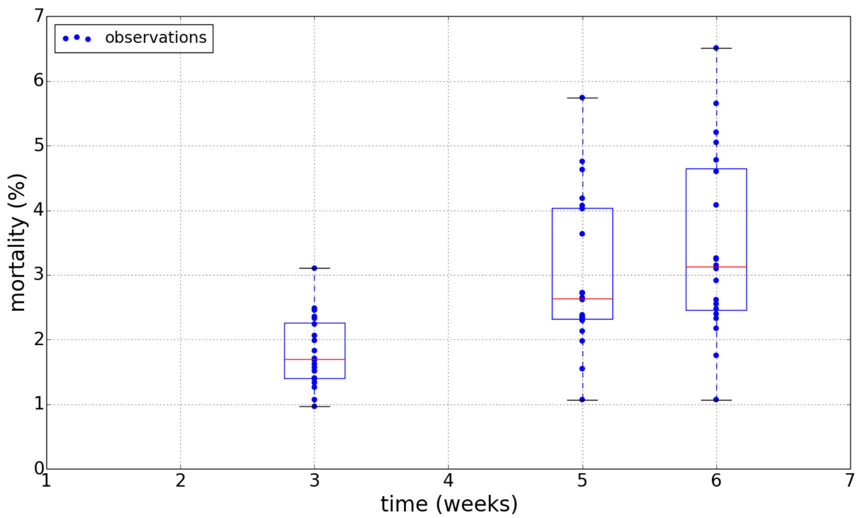


Figure 4: Mortality parameter values observed during the growth process in the set of 20 flocks used in this study, from farm arrival to the end of production.

In this case, $\boldsymbol{mort_s} \approx M(d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms})$ defines the mortality model, as it can be seen in Equation 9.

$$\hat{M}(\boldsymbol{mort_s}|X = \{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}) =$$

$$\sum_{i=1}^{n} w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\})1_{\{\boldsymbol{mort_s} \leq mort_{si}\}} \tag{9}$$

with $w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}) = K^{-1}\sum_{k=1}^{K} w_i(\{d^a_{1s}, d^b_{1s}, ..., d^a_{ms}, d^b_{ms}\}, \theta_k)$ being the weighted vector.

From resulting models, prediction intervals are generated. On the basis of such predictions, useful recommendations to adjust environmental conditions when they are outside the optimal limits can be provided.

4. Test scenario

This section presents the benefits from the use of the proposed DSS in broiler farming data, where empirical models based on real historical data monitored and collected from farms are not yet being applied, taking advantage of advances in monitoring devices, big data management and machine learning paradigms.

A total of 20 samples, which correspond to 20 flocks of broiler meat chickens from different farms around Spain, are checked to show the validity of our approach. They correspond to diverse periods of the year (mainly at the middle and end of the year) and to Ross, Cobb and Ross/Cobb breeds. The system is deployed in the cloud, so that the functionality provided can be easily accessed from different devices and locations by the end users. In addition, big data management strategies can be also considered regarding the monitoring data.

4.1. Description of the experimental setup

Environmental parameters used in this study are the temperature (TEMP°C) and relative humidity (%RH), collected from each flock inside the farm. Weights during the growth period, relative leg problems and cumulative relative mortality are also considered to generate the growth, welfare and mortality models, respectively, as it was defined in Section 3. Some statistics of such parameters are shown in Table 1, namely the maximum (*max*), minimum (*min*), mean (*mean*) and standard deviation (*std*) values and the 75th percentile (75%), or third quartile.

Table 1: Environmental parameters, weights, leg problems and mortality statistics.

		max	min	mean	std	75%
TEMP(°C)		35.2	19.4	27.6	3.2	30
%RH		70.8	40.3	60.7	5.7	65.1
Weights (gr)	**week 3**	1,042.4	775.8	913.6	70.3	965.4
	week 5	1,848.7	1,363.6	1,658.2	115.4	1,715.1
	week 6	2,768.2	2,006.2	2,319.3	203.6	2,411.1
Leg problems (%)	**week 3**	0.72	0.03	0.26	0.19	0.33
	week 5	0.93	0.19	0.42	0.24	0.57
	week 6	1.41	0.18	0.79	0.37	1.02
Cumulative mortality (%)	**week 3**	3.1	0.97	1.82	0.55	2.25
	week 5	5.73	1.06	3.01	1.2	4.03
	week 6	6.5	1.06	3.45	1.42	4.64

Quantile regression forests-based modeling is performed using the configuration parameters presented in Table 2. Although other complementary tests were performed with different number of estimators, from $K = 50$ to $K = 250$, results obtained were less accurate in terms of the number of predictions in the corresponding prediction or confidence intervals. Intuitively, the larger the number of estimators, the greater the precision of prediction, but in practice at a certain point the improvement decreases as the number of trees increases. In our case the prediction accuracy improved until $K = 200$ and started to decrease from that value. Regarding the confidence interval

(CI), 95% is used, based on common conventions and in order to efficiently control the margin of error [35]. This value implies that produced intervals will have a 0.95 probability of containing the population mean.

Table 2: Quantile regression forests configuration parameters.

Parameter	**Value**
Number of estimators (K)	200
Confidence interval (CI)	95%

In the following subsection the analysis process followed by the proposed DSS is shown over a real dataset, comprising a set of 20 flocks of broiler meat chickens from different farms in various locations around Spain. Results obtained are also discussed.

4.2. Experimental results and discussion

Given the temperature and the relative humidity parameters of the 20 flocks of broiler meat chickens from different farms under study, the environmental deviations from the corresponding empirical models learned are computed on a LOOCV (leave-one-out cross-validation) basis for each sample, by segmenting the total set of samples into 20 parts. Environmental model is thus calculated on $n - 1$ samples and cumulative deviations are obtained for the sample left out, as it was presented in Equation 1. The resulting indicators will be used as input features of the quantile regression forests-based growth, welfare and mortality models.

The average of cumulative deviations from temperature and relative humidity models used in this study can be seen in Figure 5 and Figure 6, respectively.

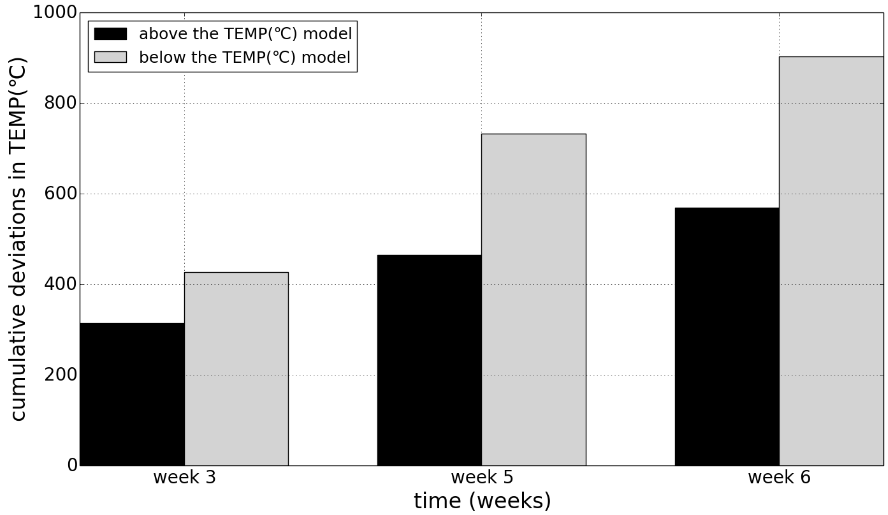


Figure 5: Cumulative deviations in temperature.

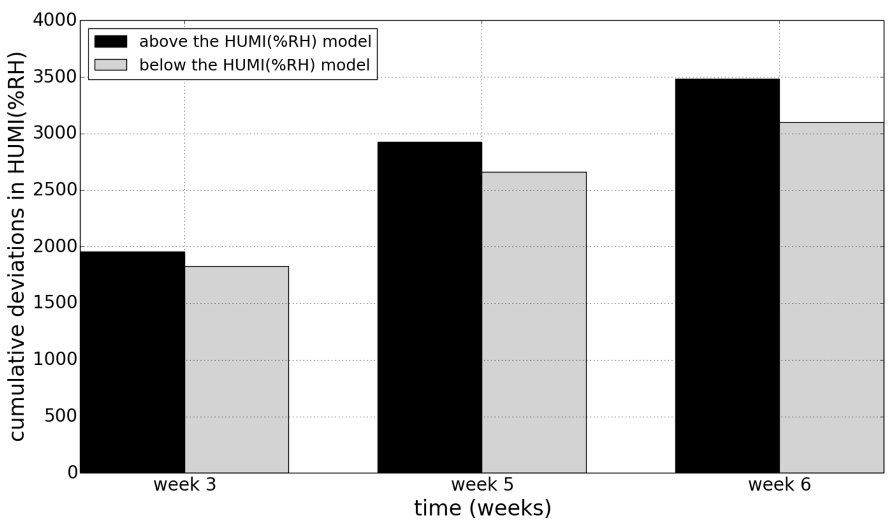


Figure 6: Cumulative deviations in relative humidity.

367 It can be appreciated how the cumulative deviations values increase as

the growth period gets close to its end in week 6, especially in relation to deviations in temperature below the model. The conditions inside the farm varies a lot from day to night and, in addition to that, heavy changes and variations were observed depending on the control system applied in each farm that strongly affect the production process.

The growth, welfare and mortality models are then obtained using the previously computed environmental indicators and targeting the available weights, leg problems and mortality rates. In order to estimate the generalization performance of the quantile regression forests-based modeling approach, the same LOOCV strategy was applied. Therefore, all models were trained on $n-1$ flocks or samples, including data related to weeks 3, 5 and 6, and tested for each sample by comparing estimated target parameters in weeks 3, 5 and 6 to unseen, real values.

Results are presented in graph and table formats. Graphs show the real and predicted values for each target parameter, weights, leg problems and mortality rates, and the corresponding quantile interval learned in relation to weeks 3, 5 and 6. These values and the three quantile intervals are provided for each of the 20 flocks under study, labeling each flock with a different sample id in the graphs. Of special interest are those cases that are detected outside the 95% CI at early stages of the production process, namely at week 3 and above or below the CI depending on the indicator to be addressed. Prediction intervals are able to deal with uncertainty, adapting estimations made to learned models and given environmental conditions. The aim is to reverse a negative situation by tuning the environmental parameters accordingly and thus reaching optimal values at the end of the process. Similarly, tables summarize the number of flocks that are above, below and within the confidence intervals regarding each week of the process.

In the case of the growth model, the prediction intervals in weeks 3, 5 and 6 for all farms can be seen in Figure 7. Prediction intervals in week 3 for samples with id 13, 16, 18 and 20 are quite precise, showing very small sizes. Some interesting trends regarding prediction intervals can be appreciated in several samples, e.g. samples with id 3, 4 or 8. Additionally, in many cases a low final weight could have been anticipated at an early stage of the process, at week 3, e.g. in samples with id 1, 5, 6, 11, 12, 16 or 20, since the upper bound of predicted intervals are very low.

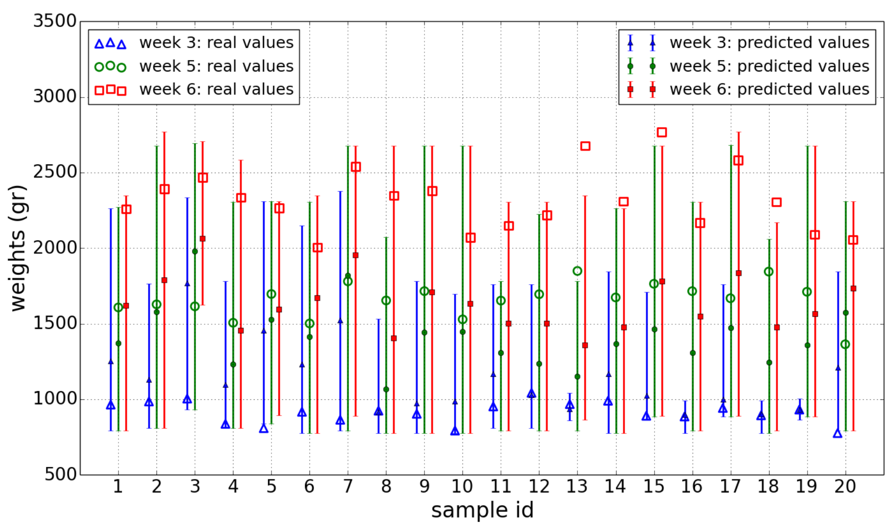


Figure 7: Random forests-based growth model. Illustration of real and predicted averaged weight values and the corresponding quantile intervals in weeks 3, 5 and 6 for each sample (flock) under study.

The results presented in Table 3 are obtained, containing the average outputs of the 20 folds. From samples detected outside the prediction intervals, there are more samples above than below them, which means that other factors may also influence the growth curve (e.g. CO^2 emissions or ammonia levels).

Table 3: Results obtained by the growth model.

	Week 3	**Week 5**	**Week 6**
Predicted in the 95% CI	18	19	16
Above the upper 95% CI limit	0	1	4
Below the lower 95% CI limit	2	0	0

The prediction intervals obtained by the welfare model in weeks 3, 5 and 6 for all samples is presented in Figure 8. Note that in the case of the flock 17th there were no leg problems registered in week 6, since in that case the growth period ended before the 6th week. In some cases, e.g. in samples with id 8, 9, 10 or 11, big differences could be found regarding the evolution of prediction intervals from weeks 3 to 6, highly increasing the predicted values and the size of the intervals. This negative welfare effect must be dealt with in advance by, for instance, readjusting the environmental parameters in accordance with the optimal conditions, in order to minimise its impact in the production at the end of the growth period.

ACCEPTED MANUSCRIPT

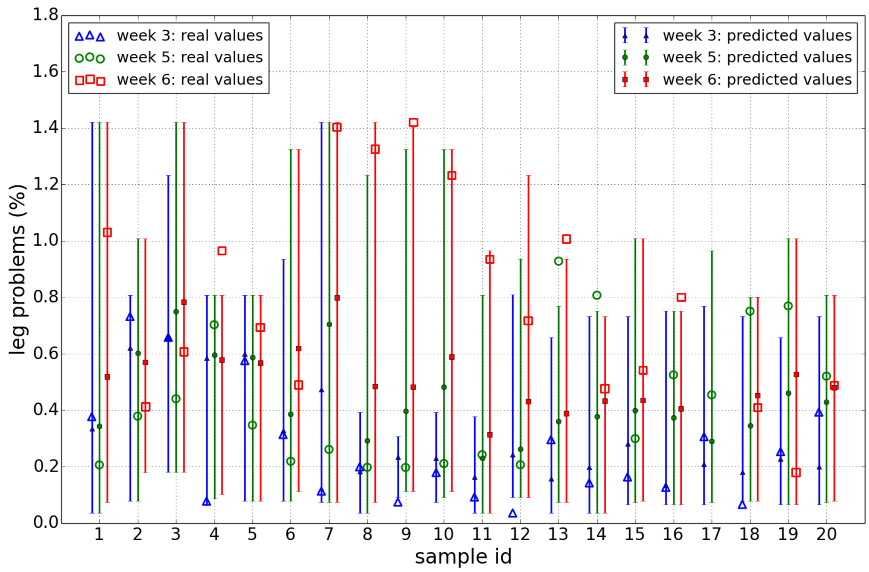


Figure 8: Random forests-based welfare model. Illustration of real and predicted frequency (%) of lame and immobile birds values and the corresponding quantile intervals in weeks 3, 5 and 6 for each sample (flock) under study.

The LOOCV process with relation to the welfare model produced the following results (see Table 4). Samples outside the prediction intervals are equally distributed below and above them in weeks 3 and 6, respectively. This means a sharp change in the global trend that could have been anticipated in week 5, when it turns out that a couple of samples are above the prediction intervals.

Table 4: Results obatined by the welfare model.

	Week 3	**Week 5**	**Week 6**
Predicted in the 95% CI	16	18	16
Above the upper 95% CI limit	0	2	4
Below the lower 95% CI limit	4	0	0

Regarding the mortality model, Figure 9 presents computed prediction intervals for all samples in weeks 3, 5 and 6. A high rate of mortality can be appreciated in the 10th and 15th samples, which could have been anticipated in week 5, when a value extremely high, above the prediction interval, was detected.

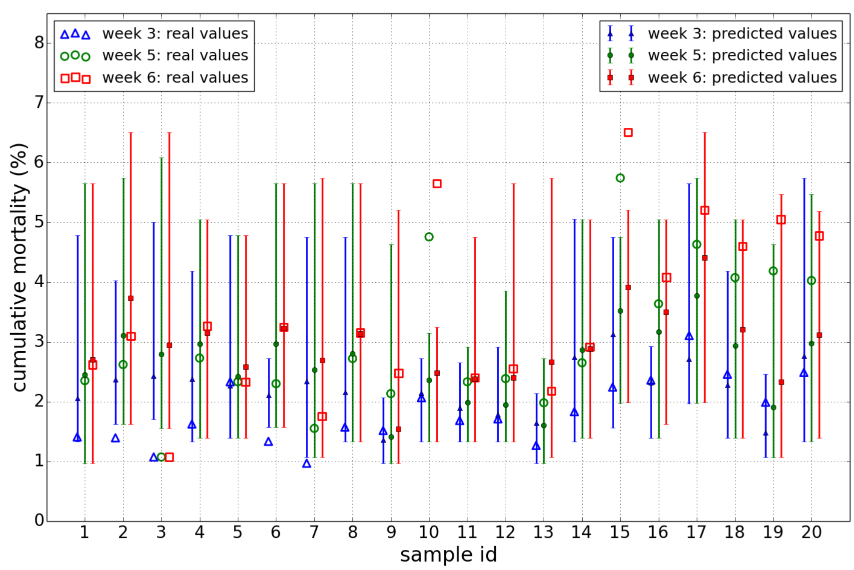


Figure 9: Random forests-based mortality model. Illustration of real and predicted cumulative mortality rate (%) values and the corresponding quantile intervals in weeks 3, 5 and 6 for each sample (flock) under study.

Following are the mortality model results (see Table 5), in which the average outputs of the 20 folds for samples predicted within the 95% CI, and above and below it, are presented. Interestingly, in general terms, it turns out that there are more samples below the prediction intervals than above them (6 and 4, respectively), especially in week 3, which means that there are low mortality rates despite the environmental conditions.

Table 5: Results obtained by the mortality model.

	Week 3	**Week 5**	**Week 6**
Predicted in the 95% CI	16	17	17
Above the upper 95% CI limit	0	2	2
Below the lower 95% CI limit	4	1	1

As a whole, it can be observed that predictions that are more accurate have smaller prediction intervals. This is due to the fact that they are related to more common, expected input environmental parameters, closer to the optimal conditions, and with slight variations in the considered flocks. Therefore, the corresponding output parameters are easier to predict.

The comparison of Tables 3, 4, and 5, suggests that the method consistently over-predicts at week 3 (the number of farms below the CI is higher that above the CI) and under-predicts at weeks 5 and 6 (the opposite relationship is evident). Therefore, it can be concluded that the transition from week 3 to 5 is critical in the growth process regarding the environmental indicators defined. In order to mitigate this effect the setting of the farm conditions should be more carefully accomplished during that period of time.

In Table 6, the global prediction accuracy and interval size obtained by growth, welfare and mortality models at weeks 3, 5 and 6 are shown. They are computed as the average of the 20 folds. In general, prediction accuracy is very high during weeks 3 and 5, and it decreases in week 6. More precisely, the best estimates are achieved in week 5, being 0.95, 0.9 and 0.85 for growth, welfare and mortality models, respectively. Only the mortality model obtained slightly better results in week 6 than in week 3, rising from 0.8 to 0.85. This is probably due to the increase in environmental conditions variance during the last days of the growth period, and to some extent to external factors that could affect the farm conditions, e.g. the way a particular farmer works. By defining a 95% CI, larger intervals are obtained and few samples are classified outside the limits. They can be considered as outliers.

Table 6: LOOCV average score results of growth, welfare and mortality models.

	Week	**Interval size (% over the total range)**	**Prediction accuracy**	
Growth model	**3**	920.1 (46.17%)	**0.9**	**88.32%**
	5	1,558.9 (78.25%)	**0.95**	
	6	1,635.5 (82.1%)	**0.8**	
Welfare model	**3**	0.73 (52.96%)	**0.8**	**81.67%**
	5	0.95 (69.33%)	**0.9**	
	6	0.95 (68.24%)	**0.75**	
Mortality model	**3**	2.55 (45.95%)	**0.8**	**83.32%**
	5	3.41 (61.73%)	**0.85**	
	6	3.98 (72.02%)	**0.85**	

Global prediction accuracy is over 81% in every proposed model, which means that given predictions can highly support the in-farm decision making by providing useful recommendations to adjust environmental conditions when they are outside the optimal limits. Although the obtained global prediction interval size denotes a high degree of uncertainty, it is still encouraging that the correctly predicted values rise above the interval size. These results demonstrate the validity of the proposed DSS for a representative set of farms, involving 20 flocks of broiler meat chickens from different locations in Spain. Moreover, it can be easily applied in other countries and environments, since it is a data-driven method that learns from data. The data-driven, automatic estimation of key production parameters is one of the main advantages over the traditional methods.

The proposed quantile regression forests-based growth, welfare and mortality models are robust and comprehensive, yet accurate decision support tools in animal farming. They are based on deviations from optimal environmental conditions, automatically collected by a set of sensors and on machine learning algorithms. The environmental models put in practice nowadays are based on theoretical curves proposed by two international companies: Cobb and Ross [1]. Therefore, the improvement with relation to the current strategies is clear.

5. Conclusions and further work

In this work we presented a DSS in animal farming that integrates last technological approaches with traditional broiler production. It provides advanced and useful recommendations to support an efficient and animal welfare-friendly production according to empirical data-driven models. One main advantage is that our approach to estimate weights, leg problems and mortality rates facilitates to the players of the production chain such as farmers, veterinarians and technical personnel crucial information to optimize production.

Discussed test scenario involving 20 flocks of broiler meat chickens from different farms showed an accurate prediction tool at different, critical, stages of the growth period, namely during the third, fifth and sixth week. Obtaining a global accuracy over 81% in every proposed model, given predictions can highly support the decision making, mainly by providing useful recommendations to adjust the environmental conditions on farm, which highly affect the growth, welfare and mortality parameters. Therefore, weights can be increased and leg problems and mortality rates minimised.

Further work should include the use of additional environmental parameters (e.g. CO^2 emissions or ammonia levels) and the application of more sophisticated paradigms, as non-parametric regression [36], in order to improve the prediction capabilities of the proposed models. Although the presented work was focused on broiler flocks, the same DSS can be applied to other farm animals. The in-field deployment of the approach will make it easier for users to manage the production smartly, anticipating deviations from optimal conditions.

Acknowledgements

This project was funded by the Spanish Ministry of Economy and Competitivity, General Directorate for Science and Technology, National Research Program 'Retos de la Sociedad' Project #AGL2013-49173-C2-1-R P.I. Inma

Estevez and #AGL2013-49173-C2-2-R. The authors wish to thank to AN and the farmers for facilitating access to their farms for data collection.

References

[1] T. Jones, C. Donnelly, M. Stamp Dawkins, Environmental and management factors affecting the welfare of chickens on commercial farms in the united kingdom and denmark stocked at five densities, Poultry Science 84 (8) (2005) 1155–1165.

[2] J. Van Meensel, L. Lauwers, I. Kempen, J. Dessein, G. Van Huylenbroeck, Effect of a participatory approach on the successful development of agricultural decision support systems: The case of pigs2win, Decision Support Systems 54 (1) (2012) 164–172.

[3] L. A. Kurkalova, L. Carter, Sustainable production: Using simulation modeling to identify the benefits of green information systems, Decision Support Systems.

[4] B. Recio, J. Ibáñez, F. Rubio, J. A. Criado, A decision support system for analysing the impact of water restriction policies, Decision Support Systems 39 (3) (2005) 385–402.

[5] M. S. Dawkins, Animal welfare and efficient farming: is conflict inevitable?, Animal Production Science 57 (2) (2017) 201–208.

[6] B. Clark, G. B. Stewart, L. A. Panzone, I. Kyriazakis, L. J. Frewer, A systematic review of public attitudes, perceptions and behaviours to-

wards production diseases associated with farm animal welfare, Journal of Agricultural and Environmental Ethics 29 (3) (2016) 455–478.

[7] C. Schulze, J. Spilke, W. Lehner, Data modeling for precision dairy farming within the competitive field of operational and analytical tasks, Computers and electronics in agriculture 59 (1) (2007) 39–55.

[8] C. Wathes, H. H. Kristensen, J.-M. Aerts, D. Berckmans, Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall?, Computers and Electronics in Agriculture 64 (1) (2008) 2–10.

[9] M. T. Scholten, I. De Boer, B. Gremmen, C. Lokhorst, Livestock farming with care: towards sustainable production of animal-source food (2013).

[10] I. Thysen, Agriculture in the information society, Journal of agricultural engineering research 76 (3) (2000) 297–303.

[11] W. Maohua, Possible adoption of precision agriculture for developing countries at the threshold of the new millennium, Computers and electronics in agriculture 30 (1) (2001) 45–50.

[12] T. M. Banhazi, H. Lehr, J. Black, H. Crabtree, P. Schofield, M. Tscharke, D. Berckmans, Precision livestock farming: An international review of scientific and commercial aspects, International Journal of Agricultural and Biological Engineering 5 (3) (2012) 1–9.

[13] T. Banhazi, M. Tscharke, W. Ferdous, C. Saunders, S. Lee, Improved image analysis based system to reliably predict the live weight of pigs

on farm: Preliminary results, Australian Journal of Multi-disciplinary Engineering 8 (2) (2011) 107–119.

[14] J. V. Stafford, Implementing precision agriculture in the 21st century, Journal of Agricultural Engineering Research 76 (3) (2000) 267–275.

[15] D. Berckmans, M. Guarino, Preface, Computers and Electronics in Agriculture 1 (64) (2008) 1.

[16] A. Frost, An overview of integrated management systems for sustainable livestock production, BSAS OCCASIONAL PUBLICATION (2001) 45–50.

[17] T. Banhazi, L. Babinszky, V. Halas, M. Tscharke, Precision livestock farming: Precision feeding technologies and sustainable livestock production, International Journal of Agricultural and Biological Engineering 5 (4) (2012) 54–61.

[18] X. Averós, M. A. Aparicio, P. Ferrari, J. H. Guy, C. Hubbard, O. Schmid, V. Ilieski, H. A. Spoolder, The effect of steps to promote higher levels of farm animal welfare across the eu. societal versus animal scientists perceptions of animal welfare, Animals 3 (3) (2013) 786–807.

[19] J. Marchewka, T. Watanabe, V. Ferrante, I. Estevez, Welfare assessment in broiler farms: Transect walks versus individual scoring, Poultry science 92 (10) (2013) 2588–2599.

[20] K. Rutherford, R. D. Donald, G. Arnott, J. A. Rooke, L. Dixon, J. Mehers, J. Turnbull, A. B. Lawrence, Farm animal welfare: assessing

risks attributable to the prenatal environment, Animal Welfare 21 (3) (2012) 419–429.

[21] H. Linseisen, Development of a precision farming information system, in: Proceedings of the Third European Conference on Precision Agriculture, Montpellier, Citeseer, 2001, pp. 689–694.

[22] D. Delen, H. Zaim, C. Kuzey, S. Zaim, A comparative analysis of machine learning systems for measuring the impact of knowledge management practices, Decision Support Systems 54 (2) (2013) 1150–1160.

[23] A. Hadjimichael, J. Comas, L. Corominas, Do machine learning methods used in data mining enhance the potential of decision support systems? a review for the urban water sector, AI Communications (Preprint) (2016) 1–10.

[24] R. Rojas, Neural networks: a systematic introduction, Springer Science & Business Media, 2013.

[25] O. W. Samuel, G. M. Asogbon, A. K. Sangaiah, P. Fang, G. Li, An integrated decision support system based on ann and fuzzy_ahp for heart failure risk prediction, Expert Systems with Applications 68 (2017) 163–172.

[26] L. A. Zadeh, Fuzzy sets, Information and control 8 (3) (1965) 338–353. doi:10.1016/S0019-9958(65)90241-X.
URL http://www.sciencedirect.com/science/article/pii/S001999586590241X

[27] M. Mehri, A comparison of neural network models, fuzzy logic, and multiple linear regression for prediction of hatchability, Poultry science 92 (4) (2013) 1138–1142.

[28] G. Fan, D. Zhong, F. Yan, P. Yue, A hybrid fuzzy evaluation method for curtain grouting efficiency assessment based on an ahp method extended by d numbers, Expert Systems with Applications 44 (2016) 289–303.

[29] Y. Kim, D. Enke, Developing a rule change trading system for the futures market using rough set analysis, Expert Systems with Applications 59 (2016) 165–173.

[30] I. Estevez, Density allowances for broilers: where to set the limits?, Poultry Science 86 (6) (2007) 1265–1272.

[31] I. Fontana, E. Tullo, A. Butterworth, M. Guarino, An innovative approach to predict the growth in intensive poultry farming, Computers and Electronics in Agriculture 119 (2015) 178–183.

[32] M. S. Dawkins, C. A. Donnelly, T. A. Jones, Chicken welfare is influenced more by housing conditions than by stocking density, Nature 427 (6972) (2004) 342–344.

[33] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[34] N. Meinshausen, Quantile regression forests, Journal of Machine Learning Research 7 (Jun) (2006) 983–999.

[35] D. Altman, D. Machin, T. Bryant, M. Gardner, Statistics with con-

fidence: confidence intervals and statistical guidelines, John Wiley & Sons, 2013.

[36] S.-K. Chao, K. Proksch, H. Dette, W. K. Härdle, Confidence corridors for multivariate generalized quantile regression, Journal of Business & Economic Statistics 35 (1) (2017) 70–85.

Highlights

- A decision support system for sustainable animal farming is proposed
- It relies on a data-driven learning framework based on quantile regression forests
- Fully comprehensive yet accurate models are obtained and deployed in the cloud
- The system is tested on real data concerning 20 different flocks of broilers
- Obtained global accuracy in growth, welfare and mortality models is over 81%