https://doi.org/10.17221/129/2022-CJAS

Comparison of some non-linear functions to describe the growth for Linda geese with CART and XGBoost algorithms

Cem Tirink[1]*, *Hasan Önder*[2], *Sabri Yurtseven*[3], *Zeliha Kaya Akil*[3]

[1]*Department of Animal Science, Faculty of Agriculture, Igdir University, Igdir, Turkey*
[2]*Department of Animal Science, Faculty of Agriculture, Ondokuz Mayis University, Samsun, Turkey*
[3]*Department of Animal Science, Faculty of Agriculture, Harran University, Sanliurfa, Turkey*

**Corresponding author: cem.tirink@gmail.com*

Citation: Tırınk C., Önder H., Yurtseven S., Kaya Akıl Z. (2022): Comparison of some non-linear functions to describe the growth for Linda geese with CART and XGBoost algorithms. Czech J. Anim. Sci., 67: 454–464.

Abstract: The aim of this study was to determine the best non-linear function describing the growth of the Linda goose breed. To achieve this aim, five non-linear functions, such as exponential, logistic, von Bertalanffy, Brody and Gompertz, were employed to define the live weight-age relationship for male and female Linda geese. In the study, 2 397 body weight-age records from 75 females and 66 males collected from three days to 17 weeks of age were evaluated using the "easynls" and "nlstools" packages for growth modelling of the Linda goose in R software. Each model was analysed in the live weight records of all the geese separately for males and females. To measure the predictive quality of the growth functions used individually here, model goodness of fit criteria, such as the coefficient of determination (R^2), adjusted coefficient of determination (R^2_{adj}), root mean square error (RMSE), Akaike's information criterion (AIC) and Bayesian information criterion (BIC) were implemented. Among the evaluated non-linear functions, von Bertalanffy model gave the best fit of describing the growth curve of female and male Linda geese. Based on the "rpart", "rpart.plot", and "caret" R packages, the CART and XGBoost algorithms were specified in the prediction of live weight of Linda geese at 17 weeks of age from the growth parameters of the von Bertalanffy model and the sex factor. XGBoost produced better results in superiority compared with the CART algorithm. In conclusion, it could be suggested that the von Bertalanffy model might help geese breeders to determine the appropriate slaughtering time, feeding regimes, and overcome flock management problems. The results of the XGBoost algorithm might present a good reference for breeders to establish breed standards and selection strategies of Linda geese in the growth parameters for breeding purposes.

Keywords: body weight; goose growth curve; non-linear models; XGBoost; CART

The goose occupies an important place among poultry all over the world due to the increasing demand in parallel with the increase in the human population (Ibtisham et al. 2017; Cilavdaroglu et al. 2020). Nowadays, goose breeding is economically important in both Asia and Central Europe. Protein sources obtained from poultry have an important place in the provision of protein sources,

Supported by the Faculty of Agriculture, Harran University (Project No. HUBAB-20052).

https://doi.org/10.17221/129/2022-CJAS

which have an important place in human nutrition (Onk et al. 2018). Goose meat is a healthy food in terms of its high protein, vitamin A, vitamin B, niacin, as well as low cholesterol and fat contents (Stevenson et al. 1989). Growth is one of the important factors for maintaining the supply of this healthy food.

Growth is an important factor in increasing the economic income from animal breeding. The concept of growth in livestock is a process in which the physiological and morphological change in both the weight and volume takes place during the period from hatching to maturity (Topal et al. 2004; Kaplan and Gurcan 2018). Tracking the growth of animals is important for herd management (Do and Miar 2019). In animal breeding, all the stages of growth must be controlled for the efficient and sustainable herd management. It is easier for the breeders to estimate the growth and to determine the required feed amount, drug dose and marketing time for the animals. In this context, the breeders pay attention to the body weight as a result of the growth rates. Within the scope of animal breeding, various methods called growth curves are used to describe changes in the body weight or length over time (Do and Miar 2019). Growth curve modelling is designed to investigate longitudinal data on the growth in biological systems over time (Bahreini et al. 2014). The ability to reach the maximum genetic potential for the cultivated biological material under the current environmental conditions can be explained with the help of growth curves. In order to better understand the growth patterns, it would be more appropriate to benefit from the growth curves associated with the increasing age of the animals in livestock studies. In animal species, growth curve models can also be used in breeding programmes, as the growth parameters are inherited (Do and Miar 2019). For this reason, many researchers have reported that the correct management of selection practices and herd management can be determined effectively with non-linear models (Sakar and Erisek 2019).

Many non-linear methods such as Exponential, Logistic, von Bertalanffy, Brody and Gompertz have been proposed for the determination of growth. Since these non-linear models have a sigmoidal structure, they are more effective than linear models in explaining growth. Together with the non-linear models, it allows the interpretation and understanding of metabolic events in animals during growth periods (Do and Miar 2019). Many research studies have modelled the growth curve in geese. However, there are few research studies on Linda geese (Kaya and Yurtseven 2021). Wu et al. (2011) used 0–8 weeks old Zhejiang White geese to compare three non-linear functions such as Logistic, Gompertz and von Bertalanffy. Onder et al. (2017) aimed to compare some growth curve models, such as Bertalanffy, Brody, Gompertz, logistic and negative exponential for Turkish native geese. Hrncar et al. (2021) used the Gompertz growth curve model for the Landes, Pomeranian and Steinbacher geese breeds.

The main purposes of this present study were to determine the best non-linear function among several non-linear functions, such as Exponential, Logistic, von Bertalanffy, Brody and Gompertz to describe the growth of the male and female Linda geese breed in Turkey for establishing breed characterisation, growth patterns and to make an accurate prediction of the body weight at 17 weeks of age by means of the growth parameters of the best selection model, and sex factor through the CART and XGBoost algorithms.

MATERIAL AND METHODS

The ethical permission was obtained from the Ethics Committee of Harran University for the use of animals in the study (ethics document number: HRU-HADYEK-2020/001-01/07).

The Linda geese used in the trial were obtained from Harran University goose farm in Şanlıurfa Province of Turkey. One-hundred forty-one three-day old goslings were divided into two groups, 75 female and 66 male geese. Both geese groups were given rations containing 23% crude protein and 3 100 kcal/kg ME between 0–6 weeks and rations containing 20% crude protein 3 100 kcal/kg ME between 7–17 weeks. The live weights of the geese were weighed weekly with the help of precision scales.

The live weight of the animals subject to the research were fitted using five non-linear models, such as the Brody, exponential, logistic, von Bertalanffy and Gompertz models. The mathematical description of each of the five selected growth models is given in Table 1.

Unknown parameters of each function can be interpreted as follows: BW is the observed live body

https://doi.org/10.17221/129/2022-CJAS

Table 1. Mathematical description of the selected growth curve models

Model	Equation	Parameters (n)	Reference
Exponential	$BW = a \times \exp^{b \times time}$	2	Arnhold 2017
Logistic	$BW = a \times (1 + b \times \exp^{-c \times time})^{-1}$	3	Bahreini et al. 2014
von Bertalanffy	$BW = a \times (1 - b \times \exp^{-c \times time})^3$	3	von Bertalanffy 1957
Brody	$BW = a \times (1 - b \times \exp^{-c \times time})$	3	Bahreini et al. 2014
Gompertz	$BW = a \times \exp^{(-b \times \exp^{-c \times time})}$	3	Gompertz 1825

weight for each goose of t age. The parameter a is the live body weight at maturity, which is the asymptotic limit of the live body weight for different ages of t. The parameter b is an integration constant adjusted according to the situation where it differs from the initial weight and time, in which, with this coefficient, it is possible to interpret the ratio of the asymptotic mature live body weight to be obtained after birth. The parameter c is a constant that expresses the linear change rate of a logarithmic function of the live body weight over time, which is specific for each of the non-linear equations (Kopuzlu et al. 2014).

To compare and explain the growth curve model performances, the goodness of fit criteria such as R^2, R^2_{adj}, Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) were used. The goodness of fit criteria functions are given below, respectively.

$$R^2 = 1 - \frac{SSE}{SST} \quad (1)$$

$$R^2_{adj} = R^2 - \left(\frac{k-1}{n-k}\right)(1 - R^2) \quad (2)$$

$$AIC = n \ln\left(\frac{SSE}{n}\right) + 2k \quad (3)$$

$$BIC = n \ln\left(\frac{SSE}{n}\right) + k \ln(n) \quad (4)$$

where:
SSE – the sum of square error;
SST – total sum of squares;
log-likelihood – maximum likelihood term, the number of parameters is represented by k;
n – sample size of the study.

To determine the significant effects of the environmental factors on the growth curve of animals, the following fixed effect model was adopted:

$$Y_{ijk} = \mu + S_i + T_j + e_{ijk} \quad (5)$$

where:
Y_{ijk} – the vector of the estimated curve parameters, i.e., a, b and c for the k^{th} Linda geese;
μ – overall mean of the flock;
S_i – one of the fixed effects of the i^{th} sex (i = male, female).

The CART model was used to correlate the individually determined growth parameters, such as a, b and c with the 17 week body weight. For this purpose, the tree-based algorithm called the CART (Classification and Regression Tree) algorithm was used, which is one of the non-parametric statistical methods used to estimate the body weight in sheep that was developed by Breiman et al. (2017). The algorithm is used as a tree-based method among data mining algorithms. The main purpose of the algorithm is to create a binary tree structure algorithm that recursively divides a node into sub-nodes and provides homogeneous subgroups with this division. In the CART algorithm, the process starts with the root node and continues with the process of splits until a homogeneous node expressing the data set is obtained.

The main purpose of creating the tree structure is to select among all the possible divisions in each node to obtain the most homogeneous sub-nodes. Each split changes depending on one explanatory variable only. All the feasible splits include the possible divisions for each explanatory variable.

In addition, XGBoost was proposed by Friedman (2001) as an efficient and scalable implementation algorithm based on gradient boosting. XGBoost is one of the most successful methods developed in machine learning algorithms in recent years (Ma et al. 2018). The XGBoost algorithm is based on the gradient tree-boosting method. In addition, the XGBoost algorithm is a regression tree algorithm with the same decision rules as the deci-

https://doi.org/10.17221/129/2022-CJAS

sion tree. The XGBoost algorithm uses a collection of classification and regression trees as a mapping to fit the training dataset (Yu et al. 2020). In addition, XGBoost is able to utilise the sparsity and overcome the overfitting problem of the data set by applying shrinkage and regularisation methods (Gertz et al. 2020). In the training process, firstly, XGBoost selects the variables that can increase the effectiveness of the model based on the decision tree to distinguish between the two groups. Furthermore, uninformative variables are routinely arranged at the expense of the computation time (Gertz et al. 2020). The main purpose in this process is to build a group of high variance and low bias decision trees.

The prediction performance of the optimum CART tree structure was evaluated using the following goodness of fit criteria (Zaborski et al. 2019):

1. Akaike information criterion (AIC):

$$\begin{cases} \text{AIC} = n \times \ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2\right] + 2k, & \text{if } n/k > 40 \\ \text{AIC}_c = \text{AIC} + \frac{2k(k+1)}{n-k-1} & \text{otherwise} \end{cases} \tag{6}$$

2. Relative root mean square error (rRMSE):

$$\text{rRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2}}{\bar{y}} \times 100 \tag{7}$$

3. Standard deviation ratio (SD_{ratio}):

$$\text{SD}_{\text{ratio}} = \frac{s_m}{s_d} \tag{8}$$

4. Performance index (PI):

$$\text{PI} = \frac{\text{rRMSE}}{1 + r} \tag{9}$$

5. Global relative approximation error (RAE):

$$\text{RAE} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{ip})^2}{\sum_{i=1}^{n} y_i^2}} \tag{10}$$

6. Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_{ip}}{y_i}\right| \times 100 \tag{11}$$

where:

n – training sample size in the data set;
k – number of model parameters;
y_i – real value of the dependent variable (BW);
y_{ip} – predicted value of the dependent variable (BW);
s_m – standard deviation of the model errors;
s_d – standard deviation of the dependent variable (BW).

All the statistical evaluations were performed using the R package (R Core Team 2020). The descriptive statistics of the quantitative characteristics were estimated by using the "psych" package in the R package (Revelle 2020). For performing the growth curve models, the "easynls" and "nlstools" packages were used (Baty et al. 2015; Arnhold 2017). To build the tree with the CART algorithm, the "rpart" and "rpart.plot" packages were used (Therneau and Atkinson 2019; Milborrow 2020). To perform the XGBoost algorithm, the "caret" package was used in the R software (Kuhn 2020). To display the predictive performances of the CART and XGBoost algorithms, the "ehaGoF" package (v0.1.1) developed by Gulbe and Eyduran (2020) was employed.

RESULTS

To describe the data set, the descriptive statistics of the data set for the body weight of the Linda geese from first week to 17 weeks of body weight are given in Table 2. For the male and female Linda geese, the maximum body weight was detected at 15 weeks (3 735.20 ± 687.80 g and 3 144.27 ± 588.34 g, respectively).

The estimated growth parameters for each growth curve model for the male Linda geese are shown in Table 3. The predicted live body weight (*a*) varied between the models and genders. The highest predicted live body weight at 17 weeks was observed as 4 294.487 g for the male geese in the Brody curve model. The exponential growth curve model was the model with the lowest 17-week live weight with a value of 1 219.419 g for the male geese. The logistic growth curve model produced the highest value for parameter *b*, which ranged from 8.049 to 11.568. Parameter *c* indicates the 17-week growth rate which varied between the models from 0.132 ± 0.008 (Brody growth curve model) to 0.416 ± 0.018 (logistic growth curve model) for the males.

https://doi.org/10.17221/129/2022-CJAS

Table 2. Descriptive statistics

Age (week)	Male BW (mean ± SD)	Female BW (mean ± SD)
1	374.7 ± 63.73	363.11 ± 52.66
2	413.45 ± 61.46	415.65 ± 62.00
3	877.56 ± 57.94	858.11 ± 65.77
4	1 513.23 ± 231.65	1 409.57 ± 266.05
5	1 818.62 ± 405.75	1 700.89 ± 415.95
6	2 161.21 ± 532.81	1 945.48 ± 482.29
7	2 369.41 ± 604.88	2 086.55 ± 540.81
8	2 602.74 ± 790.35	2 276.21 ± 691.81
9	2 746.70 ± 771.45	2 402.93 ± 675.61
10	2 929.26 ± 720.14	2 612.17 ± 696.94
11	3 488.36 ± 1040.2	3 020.92 ± 988.09
12	3 392.35 ± 753.00	2 938.00 ± 628.26
13	3 605.00 ± 675.08	3 099.41 ± 583.95
14	3 602.62 ± 656.41	3 160.73 ± 641.55
15	3 735.20 ± 687.80	3 144.27 ± 588.34
16	3 584.85 ± 593.62	3 079.27 ± 533.36
17	3 518.03 ± 615.28	3 049.80 ± 586.51

The estimated growth parameters for each growth curve model for the female Linda geese are shown in Table 4. The predicted live body weight (a) varied between the models and genders. The highest predicted live body weight at 17 weeks was observed as 3 586.195 g for the female geese in the Brody curve model. The exponential growth curve model was the model with the lowest 17-week live weight with a value of 1 115.138 g for female geese. The Logistic growth curve model produced the highest value for parameter b, which ranged from 7.085 to 10.050. Parameter c indicates the 17-week growth rate which varied between the models from 0.144 ± 0.009 (Brody growth curve model) to 0.415 ± 0.020 (logistic growth curve model) for the female geese.

Pearson's correlation coefficients between the parameters of the growth curve models fitted to the live weight from the data of the different weeks are presented in Table 5 and Table 6 for males and females, respectively. In Table 5, Pearson's correlation coefficient between parameters a and b was found to be –0.945, –0.364, –0.576, –0.709, –0.515 for the exponential, logistic, von Bertalanffy, Brody and Gompertz models, respectively. The whole correlation coefficient between parameters a and c is negative with a value of –0.654, –0.850, –0.950 and –0.799 for the logistic, von Bertalanffy, Brody and Gompertz models, respectively. The correlation between parameters b and c is 0.884, 0.863, 0.845 and 0.868 for the logistic, von Bertalanffy, Brody and Gompertz models, respectively.

In Table 6, for the female Linda breed, the correlation coefficient between parameters a and b was found –0.942, –0.350, –0.552, –0.682, –0.494 for the exponential, logistic, von Bertalanffy, Brody and Gompertz models, respectively. The whole cor-

Table 3. Estimated parameters for the selected growth curve models and their 95% confidence interval for male Linda geese

Model	Parameters	Estimate	Standard error	Confidence interval (95%)	
				lower bound	upper bound
Exponential	a	1 219.419	33.00	1 154.662	1 284.177
	b	0.074 4	0.002 1	0.070 2	0.078 5
Logistic	a	3 651.928	41.41	3 570.686	3 733.171
	b	9.808	0.896 1	8.049	11.568
	c	0.416	0.018	0.380	0.451
von Bertalanffy	a	3 852.514	61.61	3 731.631	3 973.396
	b	0.758	0.033	0.692	0.823
	c	0.238	0.012	0.214	0.261
Brody	a	4 294.487	114	4 070.728	4 518.245
	b	1.114	0.023	1.069	1.158
	c	0.132	0.008	0.115	0.149
Gompertz	a	3 770.762	53.04	3 666.703	3 874.822
	b	3.247	0.179	2.895	3.600
	c	0.284	0.013	0.258	0.310

https://doi.org/10.17221/129/2022-CJAS

Table 4. Estimated parameters for the selected growth curve models and their 95% confidence interval for female Linda geese

Model	Parameters	Estimate	Standard error	Confidence interval (95%)	
				lower bound	upper bound
Exponential	*a*	1 115.138	28.160	1 059.892	1 703.830
	b	0.070	0.001	0.066	0.074
Logistic	*a*	3 134.781	34.366	3 067.360	3 202.202
	b	8.568	0.756	7.085	10.050
	c	0.415	0.020	0.379	0.451
von Bertalanffy	*a*	3 285.372	49.28	3 188.687	3 382.056
	b	0.721	0.030	0.660	0.782
	c	0.244	0.012	0.220	0.268
Brody	*a*	3 586.195	83.82	3 421.757	3 750.633
	b	1.106	0.023	1.060	1.152
	c	0.144	0.009	0.126	0.162
Gompertz	*a*	3 224.617	43.05	3 140.154	3 309.080
	b	3.029	0.163	2.709	3.350
	c	0.289	0.014	0.262	0.316

relation coefficient between parameters *a* and *c* is negative with a value of –0.649, –0.839, –0.940 and –0.789 for the logistic, von Bertalanffy, Brody and Gompertz models, respectively. The correlation between parameters *b* and *c* is 0.887, 0.854, 0.835 and 0.860 for the logistic, von Bertalanffy, Brody and Gompertz model, respectively.

In Table 7, to compare the different non-linear growth curve model functions, the goodness of fit criteria were computed for the exponential, logistic, von Bertalanffy, Brody and Gompertz models in each gender. The highest R^2 and R^2_{adj} values were determined for von Bertalanffy growth model for each sex. Also, the lowest AIC and BIC values were computed for the von Bertalanffy growth model. To the lowest AIC and BIC and the highest R^2 and R^2_{adj} was determined for von Bertalanffy growth model. Therefore, the von Bertalanffy growth model was to best fitting model for Linda geese in each sex.

When the results are evaluated according to the goodness of fit criteria, the models are ranked in the order von Bertalanffy > Gompertz > logistic > Brody > exponential. Therefore, the von Bertalanffy model was chosen as the best non-linear model in explaining the relationship between the live weight and the weeks in the Linda geese.

There are many studies regarding the growth curves modelling by means of different non-linear functions for different species of farm animals; however, there is no information about the body

Table 5. Correlations between the parameters for the male geese

		a	*b*	*c*
Exponential	*b*	–0.945	–	–
Logistic	*b*	–0.364	–	–
	c	–0.654	0.884	–
von Bertalanffy	*b*	–0.576	–	–
	c	–0.850	0.863	–
Brody	*b*	–0.709	–	–
	c	–0.950	0.845	–
Gompertz	*b*	–0.515	–	–
	c	–0.799	0.868	–

Table 6. Correlations between the parameters for the female geese

		a	*b*	*c*
Exponential	*b*	–0.942	–	–
Logistic	*b*	–0.350	–	–
	c	–0.649	0.877	–
von Bertalanffy	*b*	–0.552	–	–
	c	–0.839	0.854	–
Brody	*b*	–0.682	–	–
	c	–0.940	0.835	–
Gompertz	*b*	–0.494	–	–
	c	–0.789	0.860	–

Table 7. Goodness of fit criteria for the selected growth curve models

Model	Male	Female	Male	Female	Male	Female	Male	Female
Exponential	18 188	20 364	18 203	20 380	60.54	57.38	60.50	57.34
Logistic	17 641	19 807	17 661	19 828	75.82	72.51	75.78	72.46
von Bertalanffy	17 620	19 785	17 640	19 805	76.26	73.00	76.22	72.96
Brody	17 644	19 801	17 664	19 822	75.76	72.63	75.72	72.59
Gompertz	17 622	19 787	17 642	19 807	76.23	72.95	76.19	72.91

weight prediction from the growth parameters that has been highlighted for the geese species in the literature. In this context, the relationship between the growth curve parameters and the body weight was explained with the CART and XGBoost algorithms in this study.

The CART and XGBoost algorithms were examined within the scope of the goodness of fit criteria. In this context, the highest Pearson's correlation coefficient (PC), coefficient of determination (R^2), the lowest relative root mean square error (rRMSE), standard deviation ratio (SDR), relative approximation error (RAE), mean absolute percentage error (MAPE), Akaike's information criterion (AIC) values provide the best fit. For the data set, the estimation performance results of the model obtained from the CART and XGBoost algorithms are given in Table 8 within the scope of goodness of fit criteria. Table 8 shows that the XGBoost algorithm provides more reliable performances than the CART algorithm. The correlation between the actual and the predicted body weight is quite strong for the CART and XGBoost algorithms ($P < 0.01$). The XGBoost algorithm produced a more reliable model that has the greatest performance with four predictors, i.e., sex, *a*, *b* and *c* growth parameters compared with the visible results created for the CART algorithm that uses four predictors, i.e., sex, *a*, *b* and *c* growth parameters. It is seen that the most optimum explanatory performance in both algorithms is made by using all the variables. However, when examined within the scope of the goodness of fit criteria, it was determined that the XGBoost method gave more reliable results.

A sensitivity analysis was performed to calculate the relative importance values of the explanatory variables, such as the growth parameters and sex on the body weight (Figure 1).

Table 9 shows the results of the CART algorithm with the scope of the cross-validation technique. In Table 9, the algorithm produced an optimal tree structure of nine terminal nodes (size of tree) with the smallest xerror (a cross-validation error: 0.314) and relative error (0.113), which means that the cross-validation R^2 and coefficient of determination (R^2) were close to each other with 0.686 (1 – 0.314) and 0.887 (1 – 0.113), respectively.

In Figure 2, the important predictors on the BW as explanatory variables were SEX, *a*, *b* and *c* growth parameters. At the top of the regression diagram, the 17-week live weight value of the Linda geese was determined as 3 257 g. In the first depth of the tree, 2 799 g is the average 17-week live weight for the Linda geese, if growth parameter *a* is lower than 2 449 g. It was lighter by nearly 1 000 g than the average Linda geese live weight when growth parameter *a* is ≥ 2 449 g. In the second depth of the tree, the average 17-week live weight was 2 526 g

Table 8. Results of the performances for the CART and XGBoost algorithms in the training and test set

Goodness of fit criteria	CART		XGBoost	
	train	test	train	test
Relative root mean square error (rRMSE)	6.455	9.458	6.421	8.585
Standard deviation ratio (SDR)	0.336	0.463	0.334	0.421
Pearson's correlation coefficients (PC)	0.942	0.886	0.943	0.908
Performance index (PI)	3.324	5.014	3.304	4.499
Relative approximation error (RAE)	0.004	0.009	0.004	0.007
Mean absolute percentage error (MAPE)	4.985	6.75	4.253	5.377
Coefficient of determination (Rsq)	0.887	0.785	0.888	0.823
Akaike's information criterion (AIC)	1 088.361	467.44	1 079.301	451.693

https://doi.org/10.17221/129/2022-CJAS

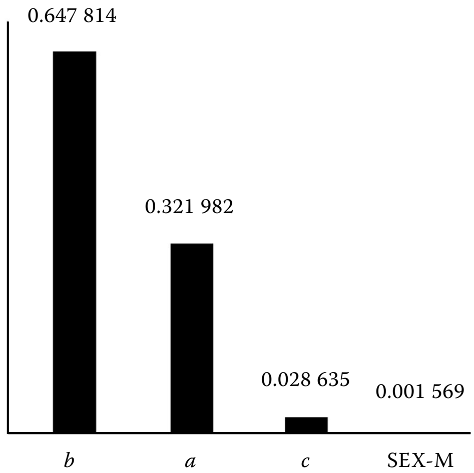

Figure 1. Relative importance for the XGBoost algorithm

Table 9. Results of the CART algorithm based on the cross-validation technique

	CP	Nsplit	Rel error	Xerror	Xstd
1	0.592	0.000	1.000	1.016	0.131
2	0.114	1.000	0.408	0.527	0.065
3	0.096	2.000	0.294	0.389	0.063
4	0.026	3.000	0.198	0.288	0.052
5	0.019	4.000	0.172	0.309	0.054
6	0.018	5.000	0.153	0.306	0.054
7	0.011	6.000	0.134	0.286	0.053
8	0.010	7.000	0.123	0.300	0.064
9	0.010	8.000	0.113	0.314	0.074

for $b < 209$. It was lighter by nearly 500 g than the average (3 061 g) of the 17-week live weight when $a <$ 2 449 and $b > 209$. For $a < 2\,449$ and $b < 165$, the 17-week live weight is average 2 063 g. In addition, when a is lower than 2 449 and b is between 165 and 209, the average 17-week live weight is 2 586 g. If the $a \geq 2\,449$, the tree showed two branches for $b < 337$. When b is lower than 337, the average 17-week live weight is 3 556 g, which is lighter than $b \geq 4\,219$ g. The $b < 337$ node showed two branches for $b < 292$ with an average of 3 351 g. In addition, $b \geq 292$ and $b < 302$ showed two branches according to the sex. If the sex is female, the average 17-week live weight is 3 417 g which is lower than close to 300 g for the male Linda geese. If b is between 302 and 337, the average 17-week live weight is 4 167 g. The last leaf of the tree is for $a \geq$ 2 449 and $b \geq 337$. This leaf showed two branches according to growth parameter $c < 8.6$ which has an average 4 132 g live weight for 17 weeks, which is lower than close to 400 g for $b \geq 8.6$.

DISCUSSION

Understanding and predicting growth processes in animal breeding is important for herd management. The growth curve, which is one of the methods used for this purpose, is a useful method for the biological evaluation, research and explanation

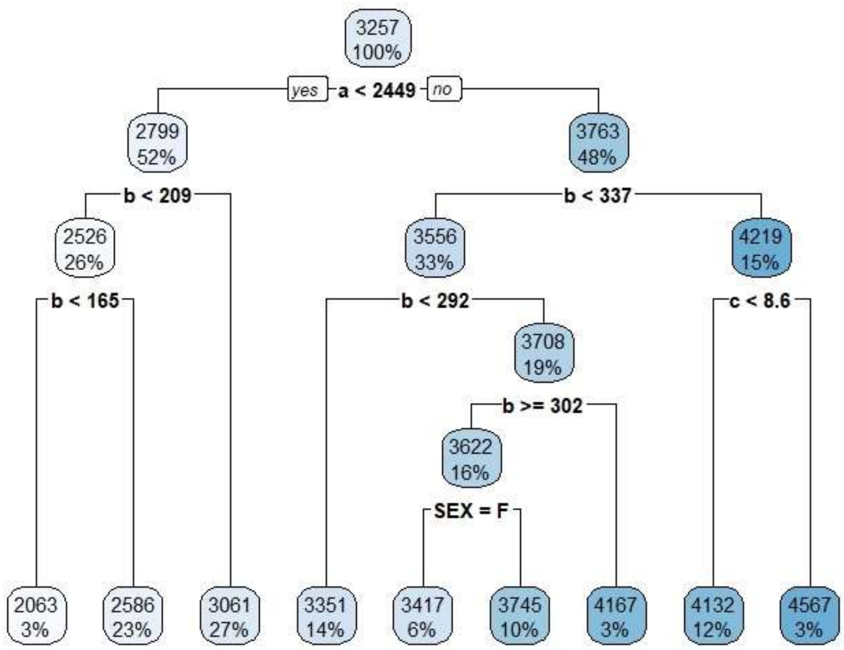

Figure 2. Regression tree diagram constructed by the CART algorithm

https://doi.org/10.17221/129/2022-CJAS

of growth relationships (Ramos et al. 2013; Ibtisham et al. 2017). In this study, the von Bertalanffy model, which gave the best results among the applied growth curves, was chosen. The study was aimed at finding the most suitable model with the selected von Bertalanffy model based on various goodness of fit criteria, with the sex and individually calculated growth parameters for the Linda geese. To this aim, the CART and XGBoost algorithms were used, and the XGBoost algorithm produced the best fit according to the goodness of fit criteria. There are few studies in Linda geese within the scope of the growth curve. However, there is no study to explain the relationship between the individually calculated growth curve parameters and the live weight. In this context, the MARS and CART algorithms have been used for different species, i.e., sheep, but algorithms, such as CART and XGBoost, were not used for geese within the scope of explaining the relationship between the live weight, sex and growth parameters.

Hrncar et al. (2021) use the Gompertz growth curve model to describe the growth of Slovakian goose breeds, such as Landes, Pomeranian and Steinbacher. The research showed that growth parameter a was 5 332.51 g for Landes, 6 186.14 g for Pomeranian and 5 048.27 g for Steinbacher geese. For parameter *b*, the results were determined as 1 960.48 g, 2 274.32 g and 1 855.98 g, respectively. In addition, growth parameter *c*, which determined the maturing rate in each breed, was similar with a value of 0.05 g. According to our results, there were differences between the growth parameters. These differences may be due to the fact that the breed and management were not the same. Ibtisham et al. (2017) used different non-linear growth functions, i.e., logistic, Gompertz, von Bertalanffy and Richards to describe the growth of two Chinese geese breeds, specifically the Shitou and Sichuan White. The research showed that the logistic growth model was the best fitting one describing the growth of two Chinese geese breeds. Onder et al. (2017) used several non-linear functions to describe the growth for Turkish native geese. To this aim, the von Bertalanffy, Brody, Gompertz, logistic and negative exponential models were used. The Gompertz and von Bertalanffy models were the best fitting models with coefficient of determination values of 0.994 and 0.995, respectively, describing the growth of Turkish native geese. The results of Onder et al. (2017) study was better fitting than our results, which may be due to the higher variance of the Linda geese in this study.

The XGBoost algorithm was found to be superior to the CART algorithm according to the study of Bharti et al. (2021) with regards to a slope stability study. Yang et al. (2019) mentioned that the XGBoost algorithm was more effective than the CART algorithm to predict the diagnostic accuracy of prediabetes. Abedi et al. (2022) mentioned that the XGBoost algorithm was better than the CART algorithm, but worse than the Random Forest algorithm for a flash-flood susceptibility mapping study. Li et al. (2022) compared some machine learning methods on multiple cattle unitary behaviours and movements and declared that the XGBoost algorithm gave better fitting results than the other methods. The XGBoost algorithm was found to be more reliable than the CART algorithm to predict lameness in dairy herds (Warner et al. 2020). All these research results support our finding that the XGBoost algorithm is superior to the CART algorithm. The superiority of the XGBoost algorithm comes from the feature that tree boosting can be seen to adaptively determine the local neighbourhoods of the model.

CONCLUSION

For the stability of herd management in animal production, it is a very important issue to follow the developmental stages of the animals in the herd. In this context, a sustainable and efficient management mechanism can be provided with growth curve modelling. In the present study, different non-linear growth functions (Brody, exponential, logistic, von Bertalanffy, and Gompertz) were evaluated to define the body weight-age relationship in the Linda goose breed. In this context, it was determined that the von Bertalanffy model was the one with the best fit among the non-linear models when compared in terms of the performance. The CART and XGBoost algorithms were applied to the growth parameters calculated separately from the determined von Bertalanffy model, which was aimed at expressing the relationship between the sex and the growth parameters. In this way, it also demonstrates the potential for the CART and XGBoost algorithms to identify the body weight-age relationship in Linda geese. In the light of all this information, this method can help increase

https://doi.org/10.17221/129/2022-CJAS

the income from the livestock, as it will be beneficial to many herd management factors, such as determining the best growth curve, optimum feed consumption of the animals, and the best slaughter time of the animals.

Conflict of interest

The authors declare no conflict of interest.

REFERENCES

Abedi R, Costache R, Shafizadeh-Moghadam H, Pham QB. Flash-flood susceptibility mapping based on XGBoost, random forest and boosted regression trees. Geocarto Int. 2022;37(19):5479-96.

Arnhold E. easynls: Easy nonlinear model. R package version 5.0. 2017 Nov 15 [cited 2021 Dec 8]. 9 p. Available from: https://CRAN.R-project.org/package=easynls

Bahreini BMR, Aslaminejad AA, Sharifi AR, Simianer H. Comparison of mathematical models for describing the growth of Baluchi sheep. J Agric Sci Technol. 2014; 16(1):57-68.

Baty F, Ritz C, Charles S, Brutsche M, Flandrois JP, Delignette-Muller ML. A toolbox for nonlinear regression in R: The package nlstools. J Stat Softw. 2015;66(5):1-21.

Bharti JP, Mishra P, Moorthy U, Sathishkumar VE, Cho Y, Samui P. Slope stability analysis using Rf, gbm, cart, bt and xgboost. Geotech Geol Eng. 2021 Jun;39(5):3741-52.

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. 1st ed. New York: Routledge; 2017. 368 p.

Cilavdaroglu E, Yamak US, Boz MA. Geese meat production. BSJ Agri. 2020;3(1):66-70.

Do DN, Miar Y. Evaluation of growth curve models for body weight in American mink. Animals. 2019 Dec 20; 10(1): 10 p.

Friedman J. Greedy function approximation: A gradient boosting machine. Ann Stat. 2001 Oct 1;29(5):1189-232.

Gertz M, Grobe-Butenuth K, Junge W, Maassen-Francke B, Renner C, Sparenberg H, Krieter J. Using the XGBoost algorithm to classify neck and leg activity sensor data using on-farm health recordings for locomotor-associated diseases. Comput Electron Agric. 2020 Jun 1; 173: 105404.

Gompertz B. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. In a letter to Francis Baily. London, UK: Royal Society; 1825. p. 513-83.

Gulbe A, Eyduran E. ehaGoF: Calculates goodness of fit statistics. R package version 0.1.1. 2020 Aug 11 [cited 2021 Dec 8]. 28 p. Available from: https://CRAN.R-project.org/package=ehaGoF

Hrncar C, Josef B, Widya PBP. The growth curve of Gompertz model in body weight of mixed-sex goose breeds in Slovakia. Genet Biodiv J. 2021 Jan 25;5(1):28-32.

Ibtisham F, An L, Li T, Niu Y, Xiao M, Zhang L, Jia R. Growth patterns of two Chinese native goose breeds. Braz J Poult Sci. 2017 Apr;19(2):203-9.

Kaplan S, Gurcan EK. Comparison of growth curves using non-linear regression function in Japanese quail. J Appl Anim Res. 2018 Jan 1;46(1):112-7.

Kaya Z, Yurtseven S. Effects of feeding systems on yield and performance of Linda geese (Anserinae sp.) in hot climatic conditions. S Afr J Anim Sci. 2021;51(5):628-34.

Kopuzlu S, Sezgin E, Esenbuga N, Bilgin OC. Estimation of growth curve characteristics of Hemsin male and female sheep. J Appl Anim Res. 2014 Apr 3;42(2):228-32.

Kuhn M. caret: Classification and regression training. R package version 6.0-86. 2020 [cited 2021 Dec 8]. 224 p. Available from: https://CRAN.R-project.org/package=caret

Li Y, Shu H, Bindelle J, Xu B, Zhang W, Jin Z, Guo L, Wang W. Classification and analysis of multiple cattle unitary behaviors and movements based on machine learning methods. Animals. 2022 Apr 20;12(9): 13 p.

Ma X, Sha J, Wang D, Yu Y, Yang Q, Niu X. Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning. Electron Commer Res Appl. 2018 Sep 1;31:24-39.

Milborrow S. rpart.plot: Plot 'rpart' models: An enhanced version of 'plot.rpart'. R package version 3.0.9. 2020 [cited 2021 Dec 8]. 25 p. Available from: https://CRAN.R-project.org/package=rpart.plot

Onder H, Boz MA, Sarica M, Abaci SH, Yamak US. Comparison of growth curve models in Turkish native geese. Eur Poult Sci. 2017 Jul 6;81:1-8.

Onk K, Sari M, Gurcan IS, Adiguzel Isik S. Live weight and body measurements of male and female native ducks raised in different raising systems. Braz J Anim Sci. 2018 May 3;47: 7 p.

R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. 2020 [cited 2021 Dec 8]. Available from: https://www.R-project.org/

Ramos SB, Caetano SL, Savegnago RP, Nunes BN, Ramos AA, Munari DP. Growth curves for ostriches (Struthio camelus) in a Brazilian population. Poult Sci. 2013 Jan 1;92(1):277-82.

Revelle W. psych: Procedures for personality and psychological research. R package version 2.0.12. Evanston, IL,

https://doi.org/10.17221/129/2022-CJAS

USA: Northwestern University. 2020 [cited 2021 Dec 8]. Available from: https://CRAN.R-project.org/package=psych

Sakar CM, Erisek A. Development of Akkaraman lambs in Cankiri region from birth to 120 days. Black Sea J Agric. 2019;2(1):16-20.

Stevenson H, Bly B, Stevenson H. Nutrition of domestic geese. Proc Nutr Soc. 1989 Jan;48(1):103-11.

Therneau T, Atkinson B. rpart: Recursive partitioning and regression trees. R package version 4.1-15. 2019 [cited 2021 Dec 8]. Available from: https://CRAN.R-project.org/package=rpart

Topal M, Ozdemir M, Aksakal V, Yildiz N, Dogru U. Determination of the best nonlinear function in order to estimate growth in Morkaraman and Awassi lambs. Small Rumin Res. 2004 Oct 1;55(1-3):229-32.

von Bertalanffy L. Quantitative laws in metabolism and growth. Q Rev Biol. 1957 Sep 1;32(3):217-31.

Warner D, Vasseur E, Lefebvre DM, Lacroix RL. A machine learning based decision aid for lameness in dairy herds using farm-based records. Comput Electron Agric. 2020 Feb 1;169: 105193.

Wu HL, Zhou B, Wang HY, Liu Y, Hu YD, Zhu Q, He DQ. Analyze on growth curve fitting of Zhejiang White goose. China Anim Husb & Vet Med. 2011;38(5):134-7.

Yang X, Fang T, Li Y, Guo L, Li F, Huang F, Li L. Pre-diabetes diagnosis based on ATR-FTIR spectroscopy combined with CART and XGBoots. Optik. 2019 Feb 1;180:189-98.

Yu X, Wang Y, Wu L, Chen G, Wang L, Qin H. Comparison of support vector regression and extreme gradient boosting for decomposition-based data-driven 10-day streamflow forecasting. J Hydrol. 2020 Mar 1;582: 124293.

Zaborski D, Ali M, Eyduran E, Grzesiak W, Tariq MM, Abbas F, Waheed A, Tirink C. Prediction of selected reproductive traits of indigenous Harnai sheep under the farm management system via various data mining algorithms. Pak J Zool. 2019 Apr 30;51(2):421-31.

Received: August 14, 2022
Accepted: November 11, 2022
Published: November 28, 2022