Prediction of cumulative egg production in japanese quails by using linear regression, linear piecewise regression and MARS algorithm

KOSKAN OZGUR[1], ALTAY YASIN[2*], AKTAN SEDAT[3]

[1] Faculty of Agriculture, Department of Animal Science, Isparta University of Applied Sciences, Isparta, Turkey
[2] Faculty of Agriculture, Department of Animal Science, University of Eskisehir Osmangazi, Eskisehir, Turkey
[3] Faculty of Veterinary Medicine, Department of Animal Science, Kırıkkale University, Kırıkkale, Turkey

SUMMARY

The study aims to predict the cumulative egg production of Japanese quails' by using linear regression, linear piecewise regression, and multivariate adaptive regression splines algorithms including age at sexual maturity, weight at sexual maturity, average weight of the first ten eggs, and partial-egg records (20, 30, 40, 60, 80, 100, and 150 d partial-egg records). All the raw data were acquired from a total of 128 female quails. To compare prediction methods, the fit criterions of 15 different models were examined, moreover the models were compared with the most common criterions.
All prediction methods showed similar results, when the 40, 60, and 80 d partial-egg records included as independent variables in the models. Although the linear regression and the MARS algorithms inferred satisfying performance with 100 and 150 d of partial-egg records, the linear piecewise regression models gave a worse prophesying performance than others did. In conclusion, as an early (indirect) selection criterion, partial-egg records from d 100 can be successfully included as independent variable into the linear regression and MARS models to predict cumulative egg production.

KEY WORDS

Egg production, Partial egg record, Linear regression, Piecewise linear regression, MARS.

INTRODUCTION

Japanese quail is described as a model animal, which has some advantages, such as easy to raise, reach to sexual maturity at an earlier age, highly adaptable to environmental conditions, shorter generation interval, need to less floor space per bird, allow to the effective selection, low-cost egg production, and resist to the diseases[1,2,3].
Besides the meat production, quails are used in egg production, commonly and effectively[4]. In last years, the importance of quail production has been rising to meet increasing demands for eggs in the poultry industry[3].
As mentioned above, earlier puberty and shorter generation interval cause to evaluate them as a proper breeding material[5]. As in most reproductive traits, egg production is mostly affected by environmental factors rather than genotypic ones (*i.e.* relatively low heritability). Therefore, their highly adaptive natures to the environmental challenges cause them to lay more eggs[6].
As in other poultry species, egg production is affected by a series of factors, such as feed intake, feed conversion, diet composition, floor space allowance, flock age, lightening, house temperature and relative humidity, health condition, and genotype[7,8]. In this context, to predict egg production in an earlier age is crucial for breeders, because of the egg production is geared towards to large-scale production of eggs for profit maximization and human consumption[9]. Along the reproductive life cycle, all of these challenging factors can affect both quality and quantity of the egg production.
There are voluminous research articles on egg production, egg weight, internal and external egg quality traits. Modelling the egg production curve has a complexity due to a typical overall performance curve is characterized by non-linear probabilities and unpredictable effectual factors (variables) on the reproductive performance (especially in house conditions that are not fully controlled). Since both pre- and post-sexual maturity processes span a wide range of time, ambiguities in many effectual factors make the egg-production curve non-linear in nature[9].
The random regression model (RRM) is determined as the best-fitted model for partial egg production records[10]. While the non-linear models (such as gamma, McNally, McMillan, Adams-Bell, compartmental, modified compartmental, logistic-curvilinear, Gloor, Lokhorst, and Narushin-Takma) are used, some researchers[11] reported that, as a multiphasic function, the segmented polynomial function could be used to estimate individual egg production and persistency. Some researchers[12] predicted the hen-day egg production rates by using nonlinear regression models (*i.e.* gamma, McNally, Adams-Bell and modified compartment models). In another research[13], individual cumulative egg production was predicted by a multiphasic function, which is developed to express the time in terms of cumulative egg number. Some internal egg quality traits were predicted by using principle component regression analysis[14]. Genetic parameters for egg weight, egg production and age at first

Corresponding Author:
Altay Yasin (yaltay@ogu.edu.tr).

oviposition were estimated by using Bayesian procedures, which uses Gibbs sampling[15]. In another study, researchers[16] estimated the genetic variance components of some egg production related traits (*i.e.* age and live weight at first oviposited egg, egg weight, partial and cumulative egg yield) by using a diverse of estimation methods (*i.e.* REML, Gibbs sampling, maximum likelihood and MIVQUE).

As we summarized above, a plenty of prediction methods were used in previous researches, but there is no report on linear piecewise regression (LPR) and multivariate adaptive regression splines (MARS). Therefore, the intent of this study is to investigate both methods to predict cumulative egg yield by using some parameters, such as age and live weight at first oviposited egg, first ten egg weight mean, and partial-egg records (*i.e.* cumulative egg numbers at 20, 30, 40, 60, 80, 100 and 150 d of laying period).

MATERIALS AND METHODS

Materials

In this study, a kind of data mining was carried out by using the performance data previously recorded in the Faculty of Agriculture poultry houses (Isparta province of Turkey). The experimental birds from which the data were acquired were reproduced from a wild type feathered, unselected, randomly mated (in group cages) breeder Japanese quail core flock. Day old chicks were emerged from daily collected and a week stored eggs. They were kept in electric-heated tier brooders until 4 weeks of age. Standard raising procedures were followed at this period. After sexing based on distinctive feather coloration (*i.e.* sexual dichromatism), a total of 128 female quails were wing-numbered and placed into individual cages. Two of them died during data collecting, and therefore excluded. Daily eggs were recorded until 210 d of age. All the quails were fed *ad libitum* with a commercial cage layer diet (17% CP and 2700 kcal/kg ME). They were daily exposed to 16:8 h hemeral lighting regime. Age and live weights were measured and recorded when they oviposited their first eggs. First ten eggs of each quail were weighed at 16.00 pm (same day) by a 0.01 g sensitive electronic scale. Clutch and pause lengths were calculated and itemized by using daily egg records in an excel sheet. As guess, cumulative egg yield is sum of total clutches.

Methods

As independent variables, first ten egg weight (FTEW), age and live weight at first oviposition (AFE and WFE, respectively), and as models, linear regression (LRM), linear piecewise regression (LPR) and MARS algorithms were used to predict cumulative egg yield.

Linear Regression Model (LRM)

Linear regression analysis is a statistical model which offers an explanatory linear prediction equation to be able to elucidate (less or more) some of the dependent variable, and it mathematically expresses the causal link between dependent and independent variables (aka response and predictors respectively). In other words, independent variable(s) can be a function of the dependent variable[17]. In linear regression analysis, the data should have some assumptions such as normal distribution, constant variance (homoscedasticity), collinearity, and independence of the residuals[18]. The mathematical model of the linear regression, which denotes the mean across change in dependent variable by predictor value's change, is given below in Equation 1[19].

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j + \varepsilon \tag{1}$$

Where,
Y is prediction of the dependent variable,
β_o is the regression coefficient,
β_j is the slope of the jth part,
ε is the error term

Linear Piecewise Regression Model (LPR)

LPR is also known as segmented regression or broken-stick regression, and this regression analysis can be applied in which the predictor is subdivided into intervals and a separate line segment is fit to each interval. It is also useful when the predictors clustered into different groups, exhibit different relationships between the variables in these regions. The boundaries between the segments are called as breakpoints, and they can be important to make decision.

Instead of using a single complex polynomial function of the equation divides each piece into a finite number of equal pieces with breakpoints at predetermined locations where a function will fit. The statistical model of LPR consisting of k pieces is given below in Equation 2.

$$Y = \beta_0 + \beta_1 X + \sum_{j=1}^{k-1} \beta_{j+1} \left(X - \Delta_j\right) I\left(X - \Delta_j\right) + \varepsilon \tag{2}$$

Where,
Y is the prediction of the dependent variable,
β_o is the regression constant
β_j is the slope of the jth part,
Δ_j is the slope change between jth and j+1th part,
$I(X - \Delta_j)$ is the breakpoint, if $X \geq \Delta_j$ then I=1; if $X < \Delta_j$ then I=0
ε is the error term

Multivariate Adaptive Regression Splines (MARS)

MARS is a flexible data mining algorithm, which is announced by Friedman[20], enabling to high precision predictions. MARS is a nonlinear regression method that creates different regression coefficients for different interval values of the independent variables that are important in the regression model. The MARS algorithm can be presumed as a generalized iterative separation method and stepwise linear regression in terms of creating the regression model. The MARS uses appropriate techniques to linearize non-linear relationships between dependent and independent variables[21]. The most prominent disadvantage of the MARS algorithm is that it is adversely affected by the multi-collinearity between the independent variables, and in this context, reliability of the model depends on its generalization ability[22]. The statistical model of the MARS algorithm is given in Equation 3.

$$\mathrm{Y} = \beta_0 + \sum_{m=1}^{M} \beta_m \prod_{k=1}^{K_m} h_m \left(X_{v(k,m)}\right) \tag{3}$$

Where,
Y is the prediction of the dependent variable,
β_o is the regression constant,
β_m is the coefficient of the basis function in which estimating equation,
$h_{km}(X_{v(k,m)})$ is the basis function
$v(k,m)$ is the indice of the independent variable used in the mth component of the kth factor,

K_m is the parameter that limits the order of interaction in the MARS

Model Goodness of Fit Criteria

Evaluation of the quality of the estimation models, which are created amongst the dependent and independent variable(s) is executed in line with the model goodness of fit criteria[23]. Albeit fifteen different model performance evaluations are examined in current study, herein only the outstanding ones are given in Equations 5 to 15.

1. Akaike information criterion, (4)

$$AIC = n.\ln\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2\right] + 2k, \quad \text{if } \frac{n}{k} > 40$$
$$else\ AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

2. The root mean square error, $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})^2}$ (5)

3. The relative root mean square error, (6)

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

4. Mean error, $ME = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_{ip})$ (7)

5. Mean absolute deviation, $MAD = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_{ip}|$ (8)

6. Standard deviation error, $SD_{ratio} = \frac{S_m}{S_d}$ (9)

7. Performance indices, $PI = \frac{rRMSE}{1+r}$ (10)

8. Relative approximate error, $RAE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - y_{ip})^2}{\sum_{i=1}^{n} y_i^2}}$ (11)

9. Mean absolute percentage error, (12)

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

10. Pearson correlation coefficient between predictions and independent variable, $r_{y_i y_{ip}} = \frac{Cov(y_i y_{ip})}{S_{y_i} S_{y_{ip}}}$ (13)

11. Determination coefficient, $r^2_{y_i y_{ip}} = \frac{Cov(y_i y_{ip})^2}{S_{y_i} S_{y_{ip}}}$ (14)

Where,
n is the number of observations,
k is the number of model parameters (selected terms),
y_j is the ith independent variable,
y_{jp} is the prediction of the ith independent variable,
S_m is the standard deviation of the model error terms,
S_d is the standard deviation of the independent variable,
$Cov(y_{i,} y_{ip})$ is the covariance between predictions and independent variables,
S_{yi} is the standard deviation of the dependent variable,
S_{yip} is the standard deviation of the predicted values.

A model should have following conditions in order to meet goodness of the prediction. While the model goodness-of-fit measures RMSE, RRMSE, RAE, CV, MRAE, MAPE and MAD are expected to be close to zero, SDR is expected to be less than 0.10.

On the other hand, while the adjusted coefficients of determination should be close to 1, the Pearson correlation coefficient amongst the actual and predicted values should be close to 1[24].

In this study, the linear regression and the MARS algorithms were executed by "earth" and "ehaGoF" packages of the R Studio software[25,26,27].

The linear piecewise regression was performed by using Statistica (v12) software.

RESULTS AND DISCUSSION

Some descriptive statistics of AFE, WFE, FTEW, partial-egg record (PER, at 20, 30, 40, 60, 80, 100, 150 d of laying period) and cumulative egg production (CEP, until 210 d of age) traits were given in Table 1.

The parameters of the seven diverse LRM to predict egg yield, which uses AFE, WFE, FTEW and PER_i (partial egg record at i^{th} age), were given in Table 2.

A continuous decrement in regression constant (β_0) was observed, when the partial-egg record increased (*i.e.* as approaching to more realistic egg yield data). As general, in the context of earlier partial records, although the WFE parameter positively contributed to the egg yield, but same parameter seemed to cause less cumulative egg yield in the model of at 150 d of productive life. It can be interpreted that the last exceptional status is more realistic. Because it is well known that the net effects of genetic increases in growth rate or juvenile (and/or pubertal maturity) body weight on avian reproduction are negative[28,29,30]. Some of these effects appear to be positive, but it appears that physiological imbalances nullify any favorable consequences. While the overall mean egg weight has a positive genetic correlation with the body mass, especially in meat type poultry species, responses to selection and genetic correlations denote that a negative correlation seem to be existed between juvenile body mass and normal egg production[29,30]. Although

Table 1 - Some descriptive statistics of dependent and independent variables.

Variable	N	Min	Max	Mean±SE	Std Dev
WFE	128	188.60	316.30	255.74±2.50	28.24
AFE	128	41.00	65.00	47.88±0.47	5.30
FTEW	128	8.69	14.93	11.29±0.09	1.06
PER_{20}	128	0.00	19.00	11.29±0.41	4.65
PER_{30}	128	6.00	28.00	20.12±0.46	5.23
PER_{40}	128	13.00	37.00	29.01±0.49	5.54
PER_{60}	128	20.00	56.00	46.54±0.56	6.32
PER_{80}	128	39.00	75.00	64.34±0.61	6.93
PER_{100}	128	46.00	94.00	81.63±0.72	8.13
PER_{150}	128	62.00	137.00	122.54±0.99	11.14
CEP	128	73.00	154.00	137.08±1.04	11.79

WFE, weight at first oviposition; AFE, age at first oviposition; FTEW, first ten egg weight; PER_i, partial-egg record on ith d; CEP, cumulative egg production.

Table 2 - Parameters of LRM which used to estimate egg production.

Models	β_0	WFE	AFE	FTEW	PER(i)
PER_{20}	143.30	+ 0.03	- 0.21	- 1.35	+ 1.01
PER_{30}	132.80	+ 0.03	- 0.18	- 1.18	+ 0.96
PER_{40}	120.50	+ 0.03	- 0.13	- 1.23	+ 1.01
PER_{60}	92.30	+ 0.04	- 0.07	- 1.10	+ 1.11
PER_{80}	55.90	+ 0.03	+ 0.06	- 0.92	+ 1.27
PER_{100}	33.90	+ 0.02	+ 0.08	- 0.54	+ 1.23
PER_{150}	19.80	- 0.01	- 0.03	- 0.12	+ 0.99

WFE, weight at first oviposition; AFE, age at first oviposition; FTEW, first ten egg weight; PER_i, partial-egg record on ith d.

it seems to be contradicted, but there are some reasonable explanations (*e.g.* increased incidence of abnormal eggs, such as double-yolked, extra-calcified shelled, compress- or slab-sided, internal laid eggs etc. and progressive regression of developing follicles and/or double follicle hierarchy) to clarify "increased ova production" versus 'reduced egg production' imbalance[31]. After all, above-mentioned exceptional and the longest partial-egg record interval has more data for the cumulative egg number (also note that the increments in partial-egg records are asymmetric), therefore why individuals with high WFE produce fewer eggs can be attributed to the number of abnormal eggs that increase by aging[31]. Herein it can be discussed that, whether abnormally oviposited eggs, leastways detectable and tenable ones can be included to the clutches to more accurate predictions, and/or it should also be considered that which partial-egg record is more optimal to reveal more realistic overall performance prediction. At this point, we also highlight that, in the multiphasic approach, total egg production was determined by not only sum of total clutches, but also by including even non-captured yolks by *infundibulum* just following ovulation (these internally laid eggs were included to the total egg production phase by using an assumption) [32].

When the AFE parameters are elaborated, it was revealed that only two partial egg records (PER_{80} and PER_{100}) had a positive effect on CEP, on the other hand, in both earlier and later partial records had a negative effect on the same trait. This reveal can be interpreted as follows: While earlier partial records have no sufficient data to predict cumulative reproductive performance[32], some causes of the egg abnormalities, such as shell defects and misshaped (due to age-related uterus/shell gland fatigue), increased incidence of the double-yolked eggs etc., occur in later periods more frequently[31]. Advanced age related issues decrease either saleable or hatching egg numbers.

All of the parameters of FTEW had a reducing effect on CEP. It is well known that the mean egg weight is influenced by a series of factors such as, hen body weight, lighting regime, health status, ambient temperature, nutritional factors, flock age, genotype *etc*[9]. Rather than table egg production, the FTEW foretells how far to reach optimum hatching (suitable for setting) egg weight in breeder flocks[33]. As it was also summarized by several researchers[24,29,30] that about existence of the genetic correlations involving sexual maturity, egg production, and egg weight, it can be briefed that heavier individuals produce lesser but heavier eggs because of both genetic and non-genetic associations, in general. All of the partial egg records had an augmenting effect on CEP. This is an expected consequence, because wherever the partial-record interval is considered, all the partial-records are an integral part of the overall reproductive performance.

Model goodness of fit criterions of the LRM, which are used to predict egg production, are given in Table 3. While the AFE, WFE and FTEW are included into the prediction models as constant independent variables, model differences are due only to the partial egg production records.

In general, results can be interpreted that the RMSE, RRMSE, RAE, CV, MRAE, MAPE, MAD, AIC, and CAIC criterions gradually approached to zero as the partial-records' range prolonged (*i.e.* approaching to full record). Looking at the proportion of the variation in the dependent variable that is predictable from the independent variables, while the adjusted coefficient of determination is 0.203 in the PER_{20} model, it raises up to 0.886 in the PER_{150} model (from least to most comprehensive partial-record, respectively). It can be concluded that the overall egg production could be predicted by linear regression from day 100.

The reproductive cycle in domesticated species with high-lay rating (*e.g. gallus domesticus, coturnix japonica*) is multiphasic[13,32]. When the layer flocks commence egg production, there is a slow increase in lay-rate at first, but then reaches to the peak with an exponential increase. This first phase is characterized by a sharp rise. The second and the most steady phase continues along the peak where the lay-rate is over ninety percent (peak lay-rate can

Table 3 - Model goodness of fit results of LRM.

Model goodness of fit criterion	PER_{20}	PER_{30}	PER_{40}	PER_{60}	PER_{80}	PER_{100}	PER_{150}
RMSE	10.40	10.30	10.04	9.25	7.89	6.32	3.93
RRMSE	7.59	7.51	7.32	6.75	5.76	4.61	2.87
SDR	0.89	0.88	0.86	0.79	0.67	0.54	0.34
CV	7.62	7.54	7.35	6.77	5.78	4.63	2.88
PC	0.47	0.48	0.52	0.62	0.74	0.84	0.94
PI	5.18	5.07	4.82	4.17	3.31	2.50	1.48
ME	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RAE	0.01	0.01	0.01	0.01	0.00	0.00	0.00
MRAE	0.01	0.01	0.01	0.01	0.01	0.00	0.00
MAPE	5.56	5.52	5.28	4.98	4.40	3.67	2.11
MAD	6.85	6.79	6.47	6.13	5.48	4.72	2.84
R^2	0.22	0.23	0.27	0.38	0.55	0.71	0.89
Adj-R^2	0.20	0.22	0.26	0. 370	0.54	0.71	0.89
AIC	603.53	600.93	594.44	573.38	532.80	476.04	354.32
CAIC	603.63	601.02	594.54	573.48	532.90	476.13	354.42

PER_i, partial-egg record on i^{th} d; RMSE, root mean square error; RRMSE, relative root mean square error; SDR, standard deviation ratio; CV, coefficient of variation; PC, Pearson's correlation coefficients; PI, performance index; ME, mean error; RAE, relative approximation error; MRAE, mean relative approximation error; MAPE, mean absolute percentage error; MAD, mean absolute deviation; R^2, coefficient of determination; Adj-R^2, adjusted coefficient of determination; AIC, Akaike's information criterion; CAIC, corrected Akaike's information criterion.

be revised a bit for different species or according to the raising goals in-species). Lastly, lay-rate starts to decrease slowly in the third phase until the onset of molting. This three-phased reproductive curve (egg production versus time curve) is also known as the generic egg-laying pattern[9,13]. Above-mentioned influencing factors can change the course of the curve, thus the CEP. Earlier and limited partial-egg records (insufficient data) cannot reflect the overall performance, and therefore, power of the prediction can be unsatisfactory.

The LPR model consists of combining more than one LRM. The data is cut from a proper breakpoint and two different LRM are created to cover the previous and the next breakpoint. Accordingly, there are two different LRM for each egg production prediction. The parameters of the LPR models, which are used to predict egg production, are given in Table 4.

While the partial-records increase, the breakpoints decrease, except for a limited increase in 150 d partial record. When all models were evaluated together, it is determined that, although the regression constants, WFE, AFE and FTEW independent variables positively or negatively contributed to egg production, on the other hand partial-records always contributed positively. As we discussed above, all of the partial-records are an integral part of the overall reproductive performance.

The model goodness of fit criterions of the LPR models, which are used to predict CEP of the Japanese quails, are given in Table 5.

It was ascertained that the RMSE, RRMSE, RAE, CV, MRAE, MAPE, MAD, AIC, and CAIC criterions diminished when the partial-records prolonged. While a 43.9 percent of the variance (adjusted coefficient of variation) in total egg production can be explained by 30 d partial-egg record, an unexpected decrease is observed in both 40 and 60 d partial-egg records. Besides, only 64.10 % of the overall variance for same dependent variable could be explained by 150 d partial-record (*i.e.* in the most ranged and the closest one to cumulative record). The non-linear nature of the typical egg production curves can compel us to use non-linear and/or fuzzy models to predict CEP instead of linear models or regression analysis[9]. Herein, unpredictable shifts in any effectual factor probably caused failures in predictions' success.

Table 4 - Parameters of LPR models used in egg production estimation.

Models	β_0^1	WFE[1]	AFE[1]	FTEW[1]	PER_i	β_0^2	WFE[2]	AFE[2]	FTEW[2]	PER_i	Breakpoint
PER_{20}	-1.41	0.38	0.27	0.54	0.95	-0.26	0.28	1.06	-0.95	2.47	132.07
PER_{30}	-0.13	0.36	0.41	-1.01	1.36	0.02	0.32	0.57	0.09	1.39	135.37
PER_{40}	0.44	0.36	-0.48	1.32	1.53	0.27	0.30	0.50	-0.73	1.61	133.28
PER_{60}	0.78	0.29	0.40	-2.53	1.29	0.60	0.14	0.82	0.53	1.28	129.01
PER_{80}	0.19	0.32	-0.86	3.25	0.42	-0.35	0.09	0.24	0.28	1.55	116.27
PER_{100}	0.91	0.17	0.50	-1.41	0.71	1.97	0.01	0.46	-0.49	1.42	116.45
PER_{150}	-1.41	0.27	0.11	0.39	0.32	1.29	-0.02	0.27	0.30	1.03	126.94

WFE, weight at first oviposition; AFE, age at first oviposition; FTEW, first ten egg weight; PER_i, partial-egg record on i^{th} d.

Table 5 - Model goodness of fit results of LPR Models.

Model goodness of fit criterion	PER_{20}	PER_{30}	PER_{40}	PER_{60}	PER_{80}	PER_{100}	PER_{150}
RMSE	14.25	14.09	13.03	10.16	7.99	6.66	6.98
RRMSE	10.40	10.28	9.51	7.41	5.83	4.86	5.09
SDR	1.21	1.18	1.10	0.86	0.68	0.57	0.59
CV	10.43	10.19	9.49	7.43	5.84	4.87	5.11
PC	0.43	0.42	0.51	0.64	0.76	0.84	0.82
PI	7.27	7.23	6.29	4.53	3.30	2.65	2.79
ME	0.28	2.24	1.42	0.60	0.53	0.41	0.07
RAE	0.01	0.01	0.01	0.01	0.00	0.00	0.00
MRAE	0.01	0.01	0.01	0.01	0.01	0.00	0.00
MAPE	8.03	7.71	7.31	5.46	4.70	3.93	3.38
MAD	10.26	9.86	9.35	6.93	6.02	5.10	4.13
R^2	0.50	0.46	0.25	0.25	0.54	0.68	0.65
Adj-R^2	0.47	0.44	0.23	0.24	0.53	0.67	0.64
AIC	684.14	681.26	661.21	597.55	535.93	489.31	501.40
CAIC	684.23	681.36	661.31	597.65	536.03	489.41	501.50

PER_i, partial-egg record on i^{th} d; RMSE, root mean square error; RRMSE, relative root mean square error; SDR, standard deviation ratio; CV, coefficient of variation; PC, Pearson's correlation coefficients; PI, performance index; ME, mean error; RAE, relative approximation error; MRAE, mean relative approximation error; MAPE, mean absolute percentage error; MAD, mean absolute deviation; R^2, coefficient of determination; Adj-R^2, adjusted coefficient of determination; AIC, Akaike's information criterion; CAIC, corrected Akaike's information criterion.

As we mentioned above, the MARS is an algorithm that can create a prophesying equation by determining breakpoints of the independent variables that are important in the model. A total of seven models are created by considering partial-egg record intervals including independent variables. In order to eliminate the multi-collinearity problem that is frequently encountered in the MARS algorithms, the models were established by taking the penalty as 2 (because of the negative and/or positive correlations among the included independent variables that we mentioned above), and the adjusted R^2 values were ensured to be close to each other with the generalized cross-validation value. The model parameters created by the MARS algorithms are given in Table 6.

In the PER_{20} model, (*i.e.* 20 d partial-egg record is independent variable in the model), for example, when we assume that a quail has twelve eggs during her 20 d partial-record, then it is expected that a quail of 139.57 eggs during overall reproductive period, approximately (CEP = 141–1.43 max (0, 13–12) = 139.57). For another example, in the PER_{150} model (*i.e.* 150 d partial-egg record is independent variable in this model), if a quail produced lesser than 110 eggs till this time, it will presumably produce averagely 1.12 lesser egg across the whole reproductive phase. Contrarily, if a quail produced more than 114 eggs during the same period, it will be expected to produce averagely 1.06 more eggs through the whole laying period.

The model goodness of fit merits of the seven MARS algorithms are tabularized below (Table 7). When the MARS algorithms model goodness fit criterions are compared with ones of the linear regressions, it could be said that they are generally similar to each other. It is thought that this is due to the absence of the interaction terms in the models that use MARS algorithms. It could be considered as an advantage that MARS algorithms use fewer independent variables to predict overall egg production. As in both previous prediction methods, it was also determined that the RMSE, RRMSE, RAE, CV, MRAE, MAPE, MAD, AIC, and CAIC model goodness fit criterions in the MARS algorithm decreased with increasing partial-egg record interval. It was determined that the explanation of the variation in CEP by the independent variables in the models increases as the partial-egg record range widened.

Table 6 - Parameters of the MARS algorithms used to predict CEP.

PER_{20}		PER_{30}		PER_{40}		PER_{60}		PER_{80}		PER_{100}		PER_{150}	
β_j	Term	β_j	Term	β_j	Term	β_j	Term	β_j	Term	β_j	Term	β_j	Term
141.00	β_0	127.00	β_0	142.00	β_0	123.00	β_0	125.00	β_0	128.00	β_0	128.00	β_0
-1.43	max(0, 13.00-PER_{20})	+7.10	max(0, AFE-44.00)	-1.23	max(0, 32.00-PER_{40})	+1.33	max (0, PER_{60}-36.00)	-2.09	max (0, 53.00-PER_{80})	+3.91	max (0, AFE-44.00)	-1.12	max (0, 110.00-PER_{150})
		-7.62	max(0, AFE- 45.00)					+1.06	max (0, PER_{80}-53.00)	-4.02	max (0, AFE-45.00)	+1.06	max (0, PER_{150}-114.00)
		-69.20	max(0, FTEW-11.30)							-2.13	max (0, 70.00-PER_{100})		
		+84.00	max(0, FTEW-11.40)							+1.25	max (0, PER_{100}-77.00)		
		-121.00	max(0, FTEW-12.40)										
		+114.00	max(0, FTEW-12.50)										
		+1.05	max(0, PER_{30}-12.00)										

WFE, weight at first oviposition; AFE, age at first oviposition; FTEW, first ten egg weight; PER_i, partial-egg record on i^{th} d.

Table 7 - Model goodness of fit criterions of the MARS algorithms.

Model goodness of fit criterion	PER_{20}	PER_{30}	PER_{40}	PER_{60}	PER_{80}	PER_{100}	PER_{150}
RMSE	10.41	9.41	10.10	9.30	7.81	5.78	3.89
RRMSE	7.59	6.87	7.37	6.78	5.70	4.22	2.83
SDR	0.89	0.80	0.86	0.79	0.67	0.49	0.33
CV	7.62	6.89	7.40	6.81	5.72	4.23	2.85
PC	0.46	0.60	0.51	0.61	0.75	0.87	0.94
PI	5.19	4.30	4.88	4.21	3.26	2.26	1.46
ME	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RAE	0.01	0.01	0.01	0.01	0.00	0.00	0.00
MRAE	0.01	0.01	0.01	0.01	0.01	0.00	0.00
MAPE	5.52	5.26	5.24	4.92	4.28	3.37	2.05
MAD	6.78	6.53	6.41	6.06	5.37	4.48	2.78
R^2	0.22	0.36	0.26	0.37	0.56	0.76	0.89
Adj-R^2	0.22	0.36	0.26	0.37	0.56	0.76	0.89
AIC	599.67	573.97	592.03	570.88	526.25	449.22	347.44
CAIC	599.67	573.97	592.03	570.88	526.25	449.22	347.44

PER_i, partial-egg record on i^{th} d; RMSE, root mean square error; RRMSE, relative root mean square error; SDR, standard deviation ratio; CV, coefficient of variation; PC, pearson's correlation coefficients; PI, performance index; ME, mean error; RAE, relative approximation error; MRAE, mean relative approximation error; MAPE, mean absolute percentage error; MAD, mean absolute deviation; R^2, coefficient of determination; Adj-R^2, adjusted coefficient of determination; AIC, Akaike's information criterion; CAIC, corrected Akaike's information criterion.

CONCLUSIONS

When considering in general, the linear models (including some early detectable variables and partial-egg records), which are used to predict CEP, their achievements are aligned following order: While the MARS algorithms and the LRM gave similar results, both were more successful than the LPR models. This method comparison is valid only for these CEP data. Only four of the twenty-one different linear models using different partial-egg records were successful to explain CEP. These were MARS and LRM, when both used only 100 and 150 d partial-egg records. It cannot be said that the prediction performances of the LPR models, which are thought to be used as an alternative to the non-linear models to meet the expectations. Moreover, the existence of the above-mentioned possibility of multi-collinearity between independent variables (AFE, WFE and as a kind of egg weight FTEW) may have limited the success of the MARS algorithms.

In conclusion, although the LRM without breakpoint, the LPR with breakpoint in its dependent variable, and the MARS algorithm with breakpoint in its independent variable are insufficient to explain the variation of the CEP in earlier ages, they can provide information to the breeders. As an inspiring suggestion to further studies, both linear and non-linear models can be comparatively investigated, and merits of the models can be examined by taking account of weights of each egg in the clutches.

Acknowledgements

This study orally presented in the 2nd International Applied Statistics Conference that is held in Tokat (Turkey), on June 29-July 2 2021, and is published as only abstracted in the proceedings book. We also thank to Faculty of Agriculture (Isparta University of Applied Sciences) for providing animal materials, housing and feeding supports.

Statement of conflict of interest

Authors have declared no conflict of interest.

Orcid

Ozgur Koskan: 0000-0002-5089-6250; Yasin Altay: 0000-0003-4049-8301; Sedat Aktan: 0000-0001-6182-5450

References

1. Tsudzuki M. (1994). Excalfactoria quail as a new laboratory research animal. Poultry Science. 73: 763-768.
2. Narayan N., Agarwal S. K., Singh B. P., Singh D. P., Majumdar S. and Singh R. V. (1998). Development of specialized strains of meat and egg type quails in hot climate. 10th European Poultry Conference. June. 21-26. Jerusalem. Israel.
3. Alkan S., Galiç A., Karabağ K. and Balcıoğlu M. S. (2008). Effects of selection for live weight and egg weight on hatching and 6 week body weights in Japanese quails (Coturnix coturnix japonica). Animal Production, 49(1): 16-19.
4. Dilmen S. and Özgen H. (1971). A new protein source is quail (Coturnix coturnix Japonica). Ankara University Faculty of Veterinary Medicine Publications No: 280.
5. Marks H. L. (1985). Direct and correlated responses to selection for growth. In: Hill, W.G., J.M. Manson, and D. Hewitt, editors. Poultry Genetics and Breeding. British Poultry Science Symposium. 18.
6. Cighi V., Oroian T., Paşcalău S., Dronca, D., and Georgescu, B. (2011). Relationship between genotype-technological environment and productive performances in Japanese quail (Coturnix coturnix). Animal Biology & Animal Husbandry, 3(2), 110-114.
7. Yannakopoulos A. L. and Tserveni-Gousi A. S. (1986). Quality characteristics of quail eggs. British Poultry Science. 27: 171-176.
8. Nazlıgül A., Bardakcıoğlu H. E., Türkyılmaz M. K., Cenan N. and Oral D. (2001). The effect of settlement frequency on egg production, egg weight and feed consumption in Japanese quail (Coturnix coturnix japonica). Journal of Istanbul University Faculty of Veterinary Medicine, 27(2): 429-438.
9. Omomule T. G., Ajayi O. O. Orogun A. O. (2020). Fuzzy prediction and pattern analysis of poultry egg production. Computers and Electronics in Agriculture, 171, 105301, doi:10.1016/j.compag.2020.105301.
10. Abou Khadiga G. Mahmoud B. Y. Farahat G. S., Emam A. M. and El-Full E.A. (2017). Genetic analysis of partial egg production records in Japanese quail using random regression models. Poultry Science. 96(8), 2569-2575.
11. Narinc D. Uckardes F. and Aslan E. (2014). Egg production curve analyses in poultry science. World's Poultry Science Journal. 70(4), 817-828.
12. Narinc D. Karaman E., Aksoy T. and Firat M. Z. (2013). Investigation of nonlinear models to describe long-term egg production in Japanese quail. Poultry Science. 92(6), 1676-1682.
13. Koops W. J. and Grossman M. (1992). Characterization of poultry egg production using a multiphasic approach. Poultry Science. 71:399-405.
14. Rathert T.Ç. Üçkardeş F. Narinç D. and Aksoy T. (2011). Comparision of principal component regression with the least square method in prediction of internal egg quality characteristics in Japanese quails. Kafkas Universitesi Veteriner Fakultesi Dergisi, 17(5).
15. Hidalgo A. M., Martins E. N., Santos A. L. D., Quadros T. C. O. D. Ton A. P. S. and Teixeira R. (2011). Genetic characterization of egg weight, egg production and age at first egg in quails. Revista Brasileira de Zootecnia, 40(1), 95-99.
16. Narinç D. Karaman E. Firat M. Z. and Aksoy T. (2011). Estimation of Multiple-trait Genetic Parameters and BLUP Using Different Estimation Methods For Some Egg Traits in Japanese Quails. Kafkas Universitesi Veteriner Fakultesi Dergisi, 7 (1): 117-123.
17. Kesici T. and Kocabaş Z. (2007). Biostatistics, Ankara University Faculty of Pharmacy Publications: 94.
18. Şahinler S. (2000). The basic principles of constructing a linear regression model using the least squares method. Journal of Mustafa Kemal University Faculty of Agriculture, 5(1-2): 57-73.
19. Kocabaş Z., Özkan M. M. and Başpınar E. (2013). Basic Biometrics. Ankara University Faculty of Agriculture Publications: 1606.
20. Friedman J. H. (1991). Multivariate adaptive regression splines. Annals of Statistics. 19: 1-67.
21. Temel G. O., Ankaralı H. and Yazıcı A. C. (2010). An Alternative Approach to Regression Models: MARS. Turkiye Klinikleri Journal of Biostatistics 2(2):58-66.
22. Kuhn M. and Johnson K. (2013). Applied Predictive Modeling. New York: Springer.
23. Grzesiak W. and Zaborski D. (2012). Examples of the use of data mining methods in animal breeding. In: Data mining applications in engineering and medicine (Ed. A Karahoca). InTech, Rijeka, Croatia, in IntechOpen 303-324.
24. Akın M., Eyduran S. P. and Eyduran E. (2020). Mars Algorithm in Solving Regression and Classification Type Problems in Agricultural Sciences with R Software, Nobel Academic Publishing, Ankara, Turkey.
25. R Core Team. (2019). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. https://www.Rproject.org/
26. Milborrow S. (2019). Derived from mda:mars by Trevor Hastie and Rob Tibshirani. Uses Alan Miller's Fortran.
27. Eyduran E. (2020). ehaGoF: Calculates Goodness of Fit Statistics. R package version 0.1.0. https://CRAN.R-project.org/package= ehaGoF
28. Lerner I. M. (1951). Natural selection and egg size in poultry. The American Naturalist, 85: 365-372.
29. Hogsett M.L., and Nordskog, A.W. (1958). Genetic-Economic Value in Selecting for Egg Production Rate, Body Weight and Egg Weight. Poultry Science, 37, 1404-1419.
30. Durmuş İ., Alkan S., Narinç D., Karabağ K., Karslı T. (2017). Effects of mass selection on egg production on some reproductive traits in Japanese quail. European Poultry Science, 81: doi.org/10.1399/eps.2017.168.
31. Scanes C.G., Butler L.D. and Kidd M.T. (2020). Reproductive management of poultry. In Animal Agriculture (Ed. Bazer F.W., Lamb G.C., Wu G.). Academic Press, p.349-366.
32. Aggrey S.E., Nichols C.R., Cheng K.M. (1993). Multiphasic Analysis of Egg Production in Japanese Quail. Poultry Science, 72: 2185-2192.
33. Aktan S., Koşkan Ö., Özsoy A.N., Akçadağ H. (2003). Some egg production characteristics and phenotypic correlations in Japanese quails (Coturnix coturnix japonica). Journal of Animal Research, 13 (1-2): 57-59.