

Using multiple regression, Bayesian networks and artificial neural networks for prediction of total egg production in European quails based on earlier expressed phenotypes

Vivian P. S. Felipe,^{*1} Martinho A. Silva,[†] Bruno D. Valente,^{*} and Guilherme J. M. Rosa^{*}

^{*}Department of Animal Sciences, University of Wisconsin - Madison, Wisconsin 53706; and [†]Department of Animal Sciences, Federal University of Jequitinhonha and Mucuri Valleys, Minas Gerais - Brazil

ABSTRACT The prediction of total egg production (TEP) potential in poultry is an important task to aid optimized management decisions in commercial enterprises. The objective of the present study was to compare different modeling approaches for prediction of TEP in meat type quails (*Coturnix coturnix coturnix*) using phenotypes such as weight, weight gain, egg production and egg quality measurements. Phenotypic data on 30 traits from two lines (L1, n = 180; and L2, n = 205) of quail were modeled to predict TEP. Prediction models included multiple linear regression and artificial neural network (ANN). Moreover, Bayesian network (BN) and a stepwise approach were

used as variable selection methods. BN results showed that TEP is independent from other earlier expressed traits when conditioned on egg production from 35 to 80 days of age (EP1). In addition, the prediction accuracy was much lower when EP1 was not included in the model. The best predictive model was ANN, after feature selection, showing prediction correlations of $r = 0.792$ and $r = 0.714$ for L1 and L2, respectively. In conclusion, machine learning methods may be useful, but reasonable prediction accuracies are obtained only when partial egg production measurements are included in the model.

Key words: egg production, phenotypic network, prediction model, non-parametric model, Bayesian networks.

2015 Poultry Science 94:772–780

<http://dx.doi.org/10.3382/ps/pev031>

INTRODUCTION

Total egg production (TEP) is an important trait in meat-type poultry breeding programs as it is directly associated with the number of hatchable eggs. Early management decisions cannot be based on TEP since this information is accumulated from 35 to 260 days of age. Therefore, prediction of TEP using early expressed traits as explanatory variables would be useful to support decision making in production/breeding systems. For example, a female may be culled earlier if she has a low predicted TEP potential. Some methods that can be used for phenotype prediction are regression analysis, Bayesian Network [BN, Bishop (2006)] and Artificial Neural Network [ANN, Hastie et al. (2009)].

Classical linear regression methods have been extensively applied in scientific research involving a variety of species and traits for many purposes, including phenotypic forecast. Multiple regression analysis is a well-known method that allows predicting a target variable (in this case, TEP) conditionally on a set of covariates (explanatory variables) by fitting a regression model

using least squares (Hastie et al., 2009). Many different models could be fitted by including different subsets of the available observed covariates and interaction terms between those. A challenge in implementing this technique is the selection of the best subset of covariates, as the inclusion of correlated predictors may result in increased standard errors of the regression coefficients which cause predictions to be very sensitive to model changes (Burnham and Anderson, 2002).

BN is an alternative modeling approach, which applies different criteria for model selection. This method consists of a graphical modeling tool that can be used to construct a predictive model parsimoniously by factorizing the posterior density distribution function assuming a set of stable conditional independencies (Jensen, 1996; Lauritzen and Spiegelhalter, 1988; Pearl, 1997).

Another useful approach for prediction is ANN, also known as “universal approximators”. This machine learning tool explores information from the input variables (covariates) and target variable to train a model that maps the relationship function between them, making it useful for future predictions (Bishop, 2006). ANNs are based on the human biological nervous system and consist of a number of interconnected neurons (linear or nonlinear computing elements) organized into layers (Bishop, 2006). The advantage of this model,

compared to the others previously mentioned, is that it can capture (and express) more complex relationships between the covariates and the target variable. This feature provides better predictions when such complexities take place in the modeled signal. Machine learning techniques have recently gained attention in this regard, given their ability to cope with cryptic relationships between variables (e.g., nonlinearities and interactions) (Gianola et al., 2011). Such methods have been applied in animal sciences, and their popularity is increasing due to their superiority in predictive ability for complex systems. ANNs have been applied to the prediction of egg production curves (Ahmad and Golian, 2008; Savegnago et al., 2011; Faridi and Golian, 2011), hatchability (Mehri, 2013), weight and number of eggs (Semsarian et al., 2013), growth curves (Ahmad, 2009) and nutritional requirement estimation (Mehri, 2012) in poultry.

The objective of the present study was to compare the efficiency of multiple (stepwise) regression, BN and ANN to predict individual total egg production of European quails using weekly body weight, partial egg production, and egg quality traits. In addition, we intended to explore BN analysis, which is a tool that has not yet been formally used for prediction of poultry traits, and compare it with methodologies classically used for this purpose.

MATERIALS AND METHODS

The experimental protocol was in agreement with the Ethical Principles in Animal Experimentation adopted by the Committee on Animal Research and Ethics, Federal University of Minas Gerais (UFMG), Brazil.

Data

The data were provided by the European Quail Breeding Program at the Prof. H lio Barbosa Experimental Farm, UFMG, Brazil. The birds were floor-raised on wood shavings in climate-controlled housing until they reached 35 days of age and then transferred to individual laying cages.

Information was recorded on two distinct lines of female European quails (L1 and L2) between October 2010 and July 2011. The traits modeled were: weekly measured body weight from birth to 35 days of age (**BW1** to **BW6**), weight gain from birth to 35 days of age (**WG1**) and from 21 to 35 days of age (**WG2**), age at first egg (**AFE**), number of eggs produced from 35 to 80 days of age (**EP1**), and the number of eggs produced from 35 to 260 days of age (**TEP**). In addition, egg quality traits were measured in four different ages (125, 170, 215 and 260 days), which included egg weight (**Ew1** to **Ew4**), yolk weight (**Y1** to **Y4**), eggshell weight (**ES1** to **ES4**), egg white weight (**EW1** to **EW4**), and egg specific gravity (**DENS1** to **DENS4**). Descriptive statistics are presented in Table 1.

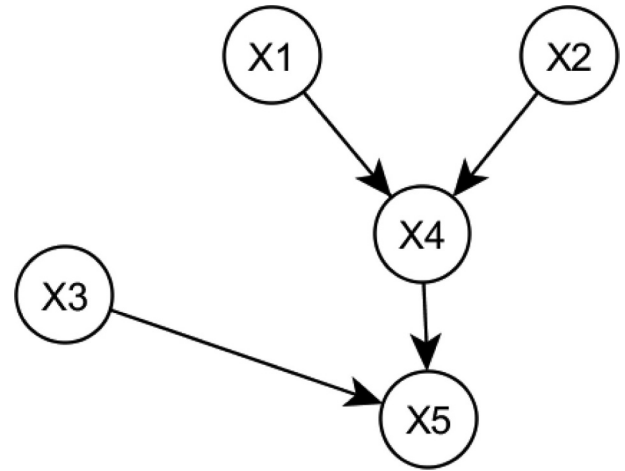


Figure 1. DAG representing a network with 5 variables (X_1 , X_2 , X_3 , X_4 and X_5).

Prediction Models

For this study a linear regression model and ANN were used as prediction tools. Also, a graphical model (BN) was implemented to infer the statistical dependence structure among production traits for selection of TEP predictors.

A BN can be seen as a graphical representation of a probability distribution over a set of variables (Margaritis, 2003). One output of such modeling approach is a Directed Acyclic Graph (DAG), which consists of a group of nodes (representing variables) connected by directed edges. A generic example of a DAG is depicted in Figure 1. This graph characterizes local features of the joint probability distribution, which brings scalability benefits due to factorization (Aliferis et al., 2010).

Given a set of variables $\{X_1, X_2, \dots, X_p\}$ with joint distribution $\Pr(X_1, X_2, \dots, X_p)$ and a DAG \mathbf{D} that is compatible with this joint distribution (Pearl, 2000), the following factorization can be performed:

$$\Pr(X_1, X_2, \dots, X_p) = \prod_{i=1}^p \Pr(X_i | \mathbf{Pa}_i)$$

in which \mathbf{Pa}_i are the parents of X_i in \mathbf{D} , i.e. all variables with arrows directed towards X_i . BN analysis involves searching for a structure that is compatible with the joint distribution of the data. The selected structure can be interpreted as causal under a set of axioms described by Pearl (1997), but for the present study the learned structure is used only in the context of prediction. Figure 1 depicts a network formed by 5 variables and will be used to illustrate some important concepts related to the structure of a BN:

1. Parents: The parents of a given node are connected to that node by an edge directed towards it (e.g., X_1 and X_2 are the parents of X_4);

Table 1. Descriptive statistics of data sets from European Quail lines L1 and L2.

Trait*	L1				L2			
	Training set (n = 90)		Testing set (n = 90)		Training set (n = 102)		Testing set (n = 103)	
	Min-Max	Mean \pm SD	Min-Max	Mean \pm SD	Min-Max	Mean \pm SD	Min-Max	Mean \pm SD
BW1(g)	7.5–12.3	9.6 \pm 0.1	7.4–12.3	9.53 \pm 0.1	7.5–12.4	9.59 \pm 0.09	7.40–14.30	9.63 \pm 0.09
BW2(g)	29.1–52.7	40.6 \pm 0.5	25.9–51.0	39.44 \pm 0.57	23.3–49.5	39.97 \pm 0.56	28.6–51.3	40.01 \pm 0.5
BW3(g)	71.6–115.9	95.4 \pm 0.9	68.3–118.8	94.30 \pm 1.04	61.9–116.7	97.78 \pm 1.01	69.8–113.9	97.65 \pm 0.71
BW4(g)	122.7–202.6	164.49 \pm 1.4	132.8–203.3	164.24 \pm 1.48	136.4–198.4	167.33 \pm 1.19	132.4–189.6	166.49 \pm 0.92
BW5(g)	178.4–264.9	233.89 \pm 1.5	189.6–269.6	224.92 \pm 1.68	198.2–280.8	228.72 \pm 1.36	191.9–256.9	225.64 \pm 1.19
BW6(g)	236.5–349.7	273.37 \pm 2.2	224.3–323.9	272.77 \pm 2.15	191.2–383.7	280.32 \pm 2.01	244.9–358.9	278.32 \pm 1.88
AFE(d)	36–53	44.04 \pm 0.4	36–61	43.88 \pm 0.4	38–73	45.97 \pm 0.72	36–98	45.99 \pm 0.72
EP1(count)	2–41	29.8 \pm 0.75	12–41	29.27 \pm 0.68	1–41	29.10 \pm 0.94	0–43	30.26 \pm 0.74
Ew1(g)	10.8–16.6	14.19 \pm 0.1	11.43–15.97	14.13 \pm 0.1	11.50–17.47	14.57 \pm 0.11	12.37–17.03	14.43 \pm 0.09
Y1(g)	3.3–5.6	4.52 \pm 0.04	3.53–5.50	4.5 \pm 0.04	3.60–5.47	4.60 \pm 0.04	3.53–5.60	4.59 \pm 0.04
ES1(g)	0.75–1.43	1.2 \pm 0.01	0.83–1.467	1.19 \pm 0.01	1.0–1.57	1.25 \pm 0.01	0.85–1.53	1.25 \pm 0.01
EW1(g)	5.15–10.2	8.46 \pm 0.07	6.77–10.2	8.44 \pm 0.08	6.80–10.60	8.72 \pm 0.07	6.70–10.05	8.60 \pm 0.07
Dens1	1.020–1.086	1.060 \pm 0.001	1.043–1.092	1.067 \pm 0.001	1.038–1.098	1.07 \pm 0.001	1.050–1.116	1.071 \pm 0.001
Ew2(g)	10.3–16.65	13.92 \pm 0.14	11.25–16.10	13.70 \pm 0.13	12.05–17.40	14.65 \pm 0.10	12.15–17.20	14.62 \pm 0.1
Y2(g)	2.7–5.5	4.38 \pm 0.06	3.0–5.6	4.32 \pm 0.06	3.70–5.60	4.61 \pm 0.04	3.60–5.75	4.63 \pm 0.04
ES2(g)	0.9–1.5	1.2 \pm 0.02	0.85–1.6	1.18 \pm 0.01	0.90–1.55	1.15 \pm 0.01	0.90–1.40	1.14 \pm 0.01
EW2(g)	6.4–10.2	8.34 \pm 0.08	6.05–9.75	8.2 \pm 0.09	7.15–10.90	8.90 \pm 0.07	6.70–11.00	8.85 \pm 0.07
Dens2	1.048–1.090	1.070 \pm 0.001	1.044–1.108	1.072 \pm 0.001	1.038–1.089	1.063 \pm 0.001	1.013–1.088	1.064 \pm 0.001
Ew3(g)	11.55–16.85	14.31 \pm 0.12	11.55–16.8	14.04 \pm 0.11	12.30–17.20	14.57 \pm 0.10	9.95–17.07	14.44 \pm 0.12
Y3(g)	3.5–5.75	4.56 \pm 0.05	3.10–5.65	4.51 \pm 0.04	3.80–5.90	4.61 \pm 0.04	3.55–5.65	4.57 \pm 0.04
ES3(g)	0.85–1.5	1.15 \pm 0.02	0.66–2.45	1.14 \pm 0.02	0.80–1.45	1.12 \pm 0.01	0.70–1.35	1.11 \pm 0.01
EW3(g)	6.95–10.2	8.6 \pm 0.07	5.75–10.29	8.4 \pm 0.08	7.20–10.80	8.83 \pm 0.07	5.70–10.70	8.78 \pm 0.09
Dens3	1.030–1.077	1.06 \pm 0.001	1.044–1.079	1.061 \pm 0.001	1.039–1.076	1.060 \pm 0.001	1.044–1.077	1.061 \pm 0.001
Ew4(g)	10.1–17.4	14.17 \pm 0.13	10.3–16.65	14.07 \pm 0.12	11.50–17.50	14.37 \pm 0.10	10.60–17.05	14.29 \pm 0.11
Y4(g)	3.45–6.4	4.64 \pm 0.06	3.50–5.65	4.58 \pm 0.05	3.30–6.90	4.60 \pm 0.05	2.95–6.20	4.60 \pm 0.05
ES4(g)	0.8–1.45	1.15 \pm 0.01	0.66–1.55	1.13 \pm 0.016	0.80–1.45	1.12 \pm 0.01	0.70–1.45	1.12 \pm 0.01
EW4(g)	2.9–10.7	8.38 \pm 0.10	5.10–9.90	8.36 \pm 0.09	6.75–10.85	8.66 \pm 0.08	4.60–10.80	8.56 \pm 0.09
Dens4	1.037–1.083	1.061 \pm 0.001	1.041–1.080	1.061 \pm 0.001	1.029–1.078	1.061 \pm 0.001	1.032–1.079	1.062 \pm 0.001
WG1(g)	226.5–340.0	263.81 \pm 2.21	215.2–313.6	263.23 \pm 2.11	181.1–374.2	270.73 \pm 2.01	235.3–350.0	268.69 \pm 1.88
WG2(g)	80.6–147.1	108.88 \pm 1.5	72.1–141.6	108.53 \pm 1.45	13.2–212.2	112.98 \pm 1.80	80.6–200.1	111.83 \pm 1.54
TEP(count)	123–216	187.93 \pm 1.96	110–217	186.59 \pm 1.97	122–216	192.35 \pm 1.81	109–216	194.26 \pm 1.55

*BW1-6: weekly body weight measured from 1 to 35 days of age; AFE: age at first egg; EP1: egg production from 35 to 80 days of age; Ew1-4, Y1-4, ES1-4 and EW1-4: egg, egg yolk, egg shell and egg white weights in four different periods, respectively; Dens1-4: egg specific gravity in four different periods; WG1 and WG2: weight gain from 1 to 35 days of age and from 21 to 35 days of age, respectively; TEP: total egg production from 35 to 260 days of age. Units: g(grams) and d(days).

2. Child: The child(ren) of a given node is connected to that node by an edge directed away from it (e.g., X_4 is the child of X_1 and X_2);
3. Spouse: Two nodes are defined as spouses when they share a common child (e.g., X_3 is a spouse of X_4);
4. Markov Blanket (**MB**): a MB of a node is the set of nodes including its parent(s), child(ren) and spouse(s) (e.g., the MB of X_4 is X_1 , X_2 , X_3 and X_5). The MB can also be seen as the minimal set of nodes which d-separates one node (i.e., makes it conditionally independent in the joint distribution) from all the others (e.g., nodes X_3 and X_4 d-separate node X_5 from the remaining nodes).

The DAG in Figure 1 can also be represented algebraically as:

$$\begin{aligned} \Pr(X_1, X_2, X_3, X_4, X_5) &= \Pr(X_1) \Pr(X_2) \Pr(X_3) \\ &\times \Pr(X_4|X_1, X_2) \Pr(X_5|X_3, X_4). \end{aligned} \quad (1)$$

This factorization involves some conditional independencies. For example, X_5 and X_1 are expected to be marginally dependent because the distribution of X_5 changes depending on the value of X_4 and the

distribution of X_4 changes depending on the value of X_1 ; but they become independent conditionally on X_4 . This and other more complex conditional independencies can be identified more easily by applying the d-separation criterion (Pearl, 1997) in **D**.

The conditional independencies entailed by the structure of a BN allow generating a more parsimonious representation of the joint distribution of a set of variables, which makes it interesting for prediction purposes. For example, consider that the DAG in Figure 1 is compatible with $\Pr(X_1, X_2, X_3, X_4, X_5)$ (i.e., $\Pr(X_1, X_2, X_3, X_4, X_5)$ can be factorized as in [1]). Further, consider X_5 as the target variable for prediction. The information from the BN indicates that not all variables in the DAG are needed to predict X_5 , but using X_3 and X_4 alone is sufficient given that conditionally on them, X_5 is independent of X_1 and X_2 .

Conditional independencies are not the only information carried by DAG's. For example, it indicates that nodes X_3 and X_4 are marginally independent, but they are dependent conditionally on X_5 . Creating dependencies between two variables upon conditioning on a third variable may take place in the case of unshielded colliders, which consist of two disconnected nodes with a common child in the graph (e.g., $X_3 \rightarrow X_5 \leftarrow X_4$). More

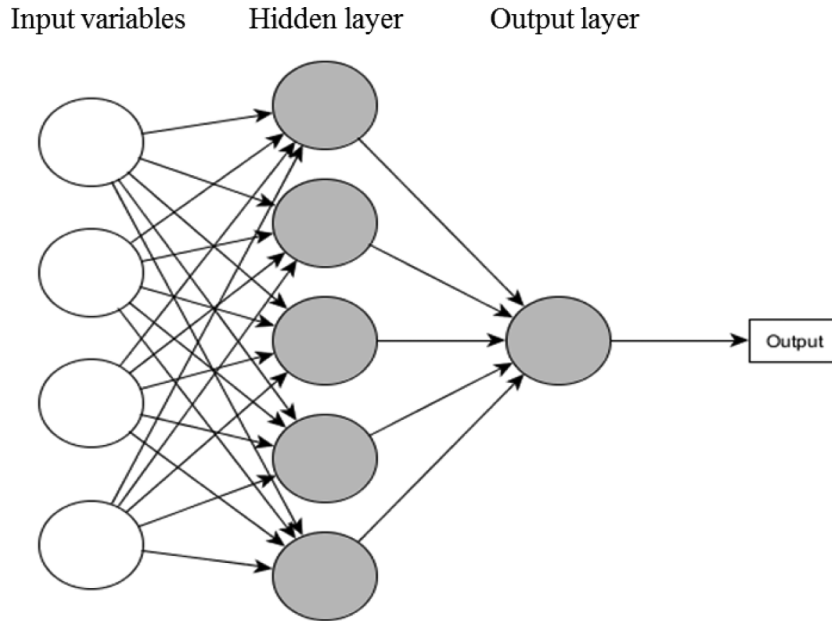


Figure 2. ANN containing one hidden layer and one output layer with 5 neurons and 1 neuron, respectively. The edges represent adaptive weights that are changed with the purpose of achieving the best network configuration for error minimization; biases associated with each active neuron (in grey) are also included in the model. The input variables are, in the case of the present study, observed phenotypes on body weight, age at first egg, partial egg production and egg quality traits; the output is predicted TEP.

specifically, the two parent nodes in an unshielded collider can be d-separated by some subset of remaining nodes, but not if this subset contains their common child. Unshielded colliders are key components for structure learning, especially regarding edges' directions.

For structure learning, two types of algorithms can be applied, which are called “constraint-based” and “search and score”. Constraint-based algorithms (Spirtes et al., 2000) infer from the data whether or not certain conditional independencies between variables hold. The search and score approach attempts to identify the network that maximizes a score function indicating how well the network fits the data (Tsamardinos et al., 2006). For the present study, a Max-Min Hill-Climbing (MMHC) algorithm (Tsamardinos et al., 2006) was used. MMHC is considered a hybrid of the two structure learning approaches. The choice of the algorithm was based on its computational efficiency. Also, the constraint-based methods do not always provide a structure with directed edges because of statistical equivalence of structures.

The model was fitted using the package bnlearn (Scutari, 2010) in R. After the phenotype network was “learned” for L1 and L2 training sets, the MB of the TEP was used as basis to construct a linear regression model for the prediction of this trait. Additional criteria used to construct alternative linear models are discussed below.

The other method applied for phenotype prediction was the ANN. The topology of an ANN (Figure 2), analogous to the human central nervous system, can be defined as a group of interconnected artificial nodes called “neurons”. The neurons are the processing units

and are capable to approximate nonlinear functions of their inputs by supervised learning in which the target variable is provided in order to train the network (Bishop, 2006). There are many types of ANNs and they differ in regard to functions adopted, network architecture, and learning algorithms. The optimal architecture for the analyzed data set was previously tested for the training set. An ANN containing one hidden layer with 3 neurons and one output layer with one neuron was fitted using the Levenberg-Marquadt (LM) algorithm. In addition, a Bayesian Regularized ANN (BRANN) containing one hidden layer with 7 neurons and one output layer with one neuron was used. The advantage of BRANN compared to an ANN without regularization is that they have less overfitting issues (Gianola et al., 2011). ANNs either with or without Bayesian Regularization were fitted using the Neural Networks toolbox available in Matlab (version R2011b).

Two additional standard multiple linear regression models (MLR) were evaluated. One of the models constructed included all the 30 variables available, and can be represented as:

$$Y_i = \mu + \sum_{j=1}^{30} \beta_j x_{ij} + \varepsilon_i, \quad (2)$$

where Y_i represents each observation of TEP, μ is an intercept, β_j 's are the slopes for each covariate ($j = 1, 2, \dots, 30$) and ε_i is a residual term. The second model was constructed by applying the stepwise procedure with backward covariate selection based on the Akaike information criteria (AIC).

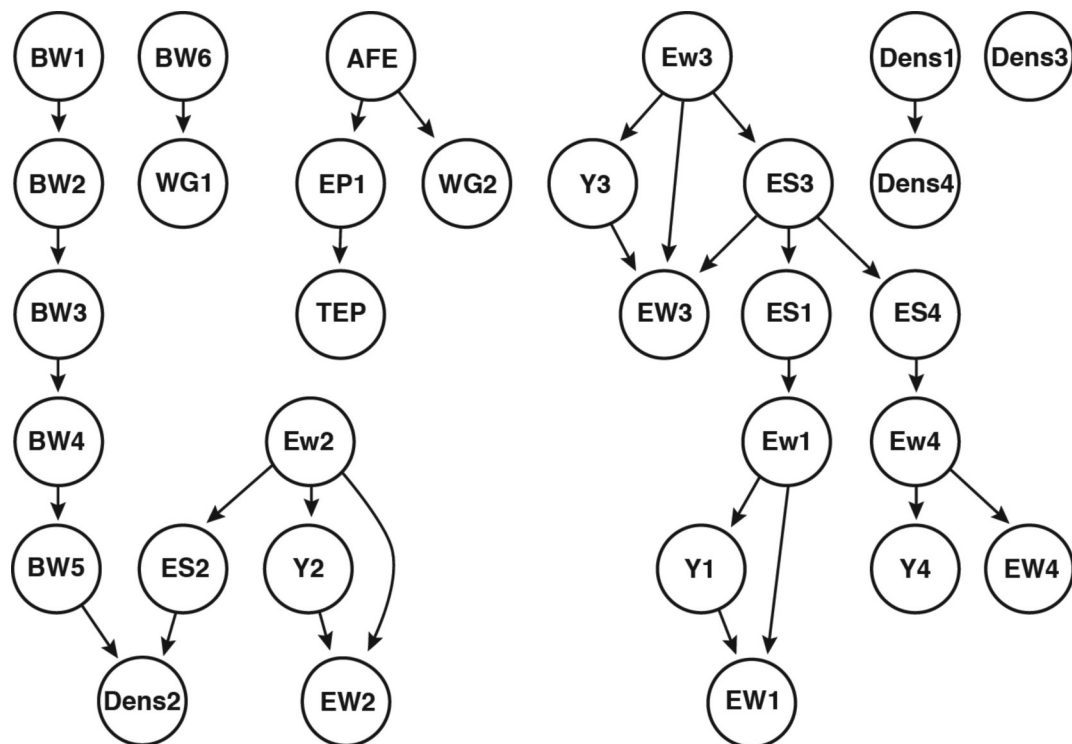


Figure 3. L1 phenotype structure considering weekly measured body weight from birth to 35 days of age (BW1 to BW6), weight gain from birth to 35 days of age (WG1) and from 21 to 35 days of age (WG2), age at first egg (AFE), number of eggs produced from 35 to 80 days of age (EP1), total egg production (TEP) and egg quality traits measured in four different life stages of the bird (125, 170, 215 and 260 days of age) (egg weight (Ew1 to Ew4), yolk weight (Y1 to Y4), egg shell weight (ES1 to ES4), egg white weight (EW1 to EW4) and egg specific gravity (Dens1 to Dens4)).

As previously mentioned, the conditional independencies provided by the DAG, learned by BN analyses, were used to construct a predictive model for TEP. In this case, as will be seen, either EP1 or AFE were considered as predictors for a least squares regression. Lastly, an ANN model was fitted considering the training sets comprising of all 30 variables as well as the input set with pre-selected variables by either the stepwise approach or BN methods.

For validation purposes, data sets on L1 and L2 were each split into training and testing sets. As ANN cannot handle genetic relationships among animals, these sets were generated such that the genetic covariances among individuals in the same set were minimized. More specifically, birds in the same set were not from the same family, that is, they had no parents in common. L1 training and testing sets consisted of 90 birds each. L2 training and testing populations comprised of 103 and 102 birds, respectively. The models were trained for both lines and validated within the line and across lines using defined testing populations. The predictive ability of the different models was assessed by computing the correlation (r) between observed and predicted TEP, and by the predicted mean squared error (PMSE) in the testing sets, both within and between lines.

RESULTS

Bayesian Network Structure

The network structures obtained, comprising the 31 phenotypic traits (TEP and 30 covariates), for L1 and L2 are presented in Figures 3 and 4, respectively. For the L1 line (Figure 3), results indicate that TEP is directly connected only to partial egg production (EP1). The remaining observed traits are not expected to contribute to predicting the TEP in the presence of (i.e., conditionally on) EP1. In further analysis (graphic not shown), in which EP1 was removed from the data set, AFE became directly connected to TEP. Excluding both EP1 and AFE from the covariates set, the selected DAG presented TEP as completely disconnected from the other nodes, which implies independence of the TEP from other traits considered in this specific set for L1. An interesting aspect of the result obtained is the path BW1→BW2→BW3→BW4→BW5, for which the statistical consequence seems to fit what is expected given growth curve behavior and timeline.

For L2 (Figure 4), the DAG also shows that TEP is directly dependent only on EP1. Similar to L1, results indicate that, conditionally on EP1, no other earlier observed trait (such as body weight) brings additional information about the TEP potential of a bird. Also,

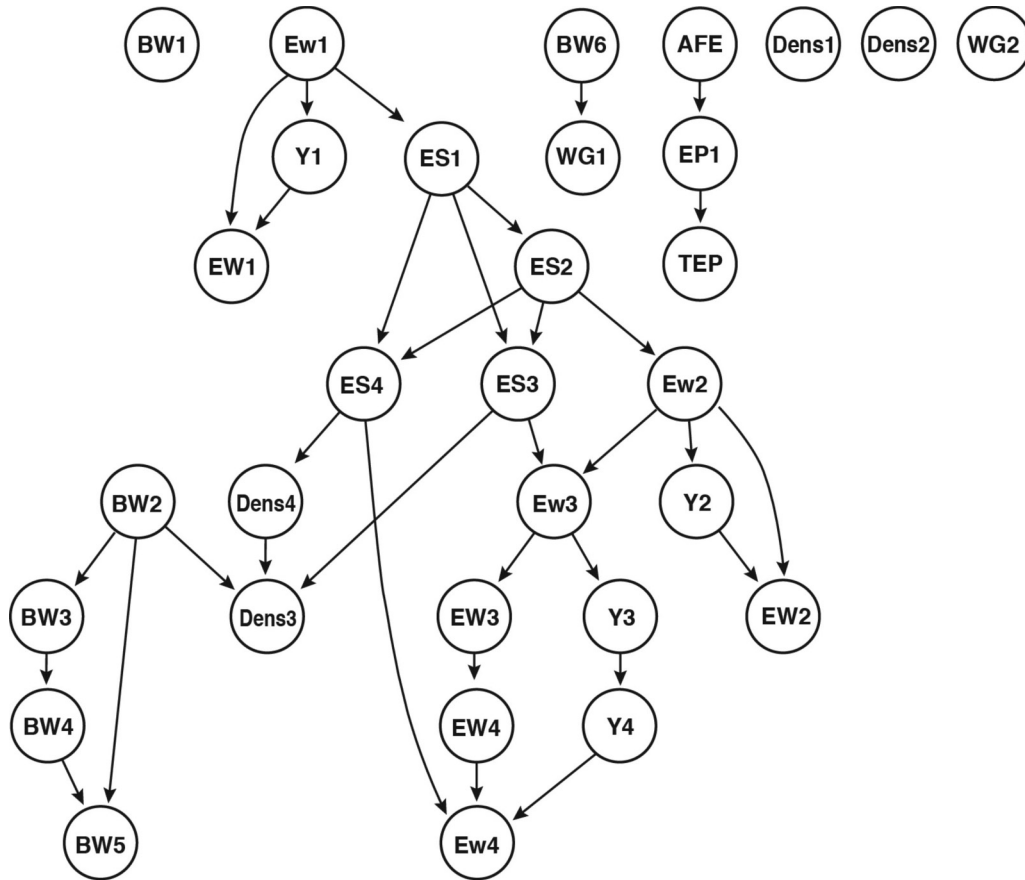


Figure 4. L2 phenotype structure considering weekly measured body weight from birth to 35 days of age (BW1 to BW6), weight gain from birth to 35 days of age (WG1) and from 21 to 35 days of age (WG2), age at first egg (AFE), number of eggs produced from 35 to 80 days of age (EP1), total egg production (TEP) and egg quality traits measured in four different life stages of the bird (125, 170, 215 and 260 days of age) (egg weight (Ew1 to Ew4), yolk weight (Y1 to Y4), egg shell weight (ES1 to ES4), egg white weight (EW1 to EW4) and egg specific gravity (Dens1 to Dens4)).

similar to the results obtained for L1, excluding EP1 and AFE from the L2 training set resulted in an output DAG in which the TEP was disconnected from the other traits.

The structures learned for L1 and L2 present a different number of directed and undirected edges, including 14 edges present exclusively in the L1 DAG, 18 edges exclusively in the L2 DAG, and 15 edges in common to both structures. However, in the vicinities of TEP, both DAGs are equivalent, meaning that the inferred MB for this node is the same for both lines. This result implies that among the variables studied, EP1 can be considered the only predictor for TEP regardless of the line. Moreover, AFE can be considered a predictor trait in the absence of partial egg production measurement as TEP and AFE are independent conditionally solely on EP1.

Predictive Ability

The results for correlation between observed and predicted TEP and model PMSE for MLR and BN are presented in Table 2. The predictive ability for each model

with or without a variable selection step for within and across lines prediction scenarios is showed. Comparing predictions within lines, the predictive ability was higher in L2 than in L1 for most part of prediction models/methods.

Comparing the predictive ability across models, choosing covariates using the BN (Parent: EP1) yielded better results in terms of correlation and PMSE compared to MLR (with or without stepwise approach). However, when AFE was considered a predictor (scenario in which EP1 was removed from the observed data) the predictive ability was reduced by 50% in all BN prediction scenarios. MLR resulted in lower correlation between observed and predicted TEP both within and across lines. Using stepwise variable selection approach did not result in dramatic improvement over including all possible covariates, especially for L2. One exception was for predictive ability across lines when L1 was used as training set. Results on across lines predictive ability showed that the stepwise procedure greatly improved prediction of L2 TEP ($r = 0.645$ and 0.038 for with and without stepwise, respectively). Interestingly, stepwise variable selection in L2 gave lower correlation than in L1 when predicting L2 ($r = 0.590$ and 0.645 ,

Table 2. Correlation (r) between predicted and measured TEP for training and testing sets and model predicted mean squared error (PMSE) within and across lines, using Multiple Regression (with all 30 variables as predictors or applying stepwise procedure) and Bayesian Networks (considering partial egg production from 35 to 80 days (EP1) or age at first egg (AFE) as predictor variables).

	Multiple Regression (MLR)			
	Line 1		Line 2	
	30 variables	Stepwise*	30 variables	Stepwise**
r_{training}	0.705	0.658	0.871	0.861
$r_{\text{testing (within line)}}$	0.332	0.445	0.603	0.590
PMSE (within line)	382.57	303.58	167.57	168.53
$r_{\text{testing (across lines)}}$	0.362	0.354	0.038	0.645
PMSE (across lines)	359.42	372.66	3830.44	179.24
Bayesian Networks (BN)				
	Line 1		Line 2	
	Parent:EP1	Parent:AFE	Parent:EP1	Parent:AFE
r_{training}	0.530	0.269	0.759	0.462
$r_{\text{testing (within line)}}$	0.554	0.299	0.677	0.377
PMSE (within line)	241.86	317.31	132.06	219.36
$r_{\text{testing (across lines)}}$	0.677	0.376	0.554	0.299
PMSE (across lines)	164.46	312.91	276.5	383.49

*Variables: EP1, DENS2, ES3, DENS3, EW4, Y4, DENS4.

**Variables: AFE, EP1, Ew1, Y1, EW1, Dens1, Ew2, Y2, Dens2, Ew3, Y3, EW3, Dens4.

Table 3. Correlation (r) between predicted and measured total egg production for training and testing sets and predicted mean squared error (PMSE) of Artificial Neural Networks with or without variable selection (Stepwise, BN) and two types of learning algorithms (Levenberg-Marquadt, Bayesian Regularized Neural Networks) within and across lines.

	Levenberg-Marquadt algorithm (LM)					
	Line 1			Line 2		
	30 variables	Stepwise*	BN***	30 variables	Stepwise**	BN***
r_{training}	0.659	0.683	0.554	0.765	0.785	0.739
$r_{\text{testing (within line)}}$	0.671	0.792	0.752	0.529	0.597	0.714
PMSE (within line)	442.89	322.01	279.51	218.32	186.76	156.23
$r_{\text{testing (across lines)}}$	0.181	0.424	0.597	0.248	0.282	0.580
PMSE (across lines)	1071.14	485.87	211.74	497.61	509.58	300.46
Bayesian Regularized Neural Networks (BRANN)						
	Line 1			Line 2		
	30 variables	Stepwise*	BN***	30 variables	Stepwise**	BN***
r_{training}	0.674	0.691	0.559	0.873	0.879	0.759
$r_{\text{testing (within line)}}$	0.432	0.453	0.487	0.639	0.579	0.685
PMSE (within line)	464.02	426.76	367.15	214.24	332.63	158.24
$r_{\text{testing (across lines)}}$	0.496	0.500	0.617	0.427	0.389	0.559
PMSE (across lines)	450.40	547.23	284.97	380.93	440.70	273.89

*Variables: EP1, DENS2, ES3, DENS3, EW4, Y4, DENS4;

**Variables: AFE, EP1, Ew1, Y1, EW1, Dens1, Ew2, Y2, Dens2, Ew3, Y3, EW3, Dens4;

***EP1.

respectively). To investigate these results further, we re-trained a model in L2 with the covariates selected in L1 with the stepwise procedure. The predictive correlation increased to 0.660. This indicates that L1 was not a better training set for L2, but instead the covariates selected by the stepwise procedure on L1 training set provided a better subset of predictors for the target trait.

In summary, the results showed that using BN (with EP1 in the set of predictors) to construct regression

models provided better predictions of TEP and superior generalization ability within and across quail lines when compared to the standard approaches used with MLR. Using the same data, ANN was fitted for prediction of TEP using two different learning algorithms and three distinct covariate sets (including all variables or the ones selected when using stepwise procedure or BN) (Table 3). The learning algorithm influenced ANN predictive ability. The LM algorithm worked better than the BRANN for prediction within the same line

(correlations of 0.752 and 0.714 against 0.487 and 0.685 for L1 and L2, respectively), and showed quite similar results across lines.

The traits considered as input variables, as expected, also affected the ANN predictive ability, and the pre-selection of variables was advantageous for prediction using ANN. Lower correlations were obtained when all 30 variables were considered in the model. In general, BN was the best method to select variables for prediction of TEP, except within line for L1.

The ANN seemed to be a better predictive machine than the traditional linear model, especially for L1 (in which the correlation was almost 24% higher), and pre-selection of variables improved both model prediction and generalization. However, despite of the fact that ANN provided higher correlation values, the use of BN for prediction resulted in lower PMSE.

DISCUSSION

The aim of this study was to compare different approaches for prediction of TEP in quails using phenotypes expressed early in life as predictor variables. Initially, BN analysis was performed to obtain a phenotypic network that was compatible with the joint distribution of the traits, and therefore make explicit the conditional independencies for this distribution. This information described which nodes comprise the MB of TEP. This is important information for the prediction of TEP given that the remaining nodes do not contribute to the prediction conditionally on the MB set. Also, such data-driven analysis is interesting to verify that the statistical consequences of the generated DAG were consistent with prior beliefs about the observed biological system. An example is the path involving the body weight traits, which is the expected structure from biological system knowledge.

DAGs are an intuitive way to understand direct and indirect relationships among variables and to foresee how model likelihood and predictive ability can change according to the variables that are added to or excluded from the models. For example, when a convergent structure (e.g., $X3 \rightarrow X5 \leftarrow X4$) is present in a DAG, if the child is not inserted in the model (e.g. a model such as $X3 = f(X4) + e$), the predictive ability is expected to be poor since the two parents would be marginally independent (considering that there is no other path between them). However, when the child is included in the model (e.g. $X3 = f(X4, X5) + e$), accounting for the value of one parent can actually contribute to the prediction about the other, since they are dependent when both are conditioned on the common child.

Another advantage of the BN approach is that it makes it easier to evaluate local features of the joint distribution. For example, it can be easily verified that the local structure involving the target trait was the same for L1 and L2. For both lines, TEP was independent of all other observed variables given EP1, indicating that

using this variable alone to construct regression models for prediction of TEP would be sufficient regardless of the line. This dependence between TEP and EP1 was expected given that the second is a component of the first. Results also showed that generalization was improved in across line predictions when BN was used as a variable selection tool. It is usually the case that models with fewer variables also contain fewer nuisance variables and have greater generality (Ginzburg and Jensen, 2004).

However, knowing which variables are present in the MB of a given trait does not tell us per se how good a predictor these variables are. For example, AFE was in the MB of TEP, but using AFE as predictor had a negative impact on model predictive ability compared with using EP1 as predictor. Our results show that management decisions should be based only on traits relative to partial egg production because only EP1 and AFE satisfactorily predict TEP. Another alternative would be to use genetic marker information for prediction (Fulton, 2012) of this trait, but this type of information was not available.

In the context of linear model prediction, selecting covariates is a key step. The drawbacks of stepwise multiple regression, such as bias in parameter estimation and inconsistencies among model selection algorithms, are well known (Hurvich and Tsai, 1990; Burnham and Anderson, 2002; Whittingham et al., 2006). In this study, using a linear model with the stepwise procedure to predict TEP did not find an optimal set of predictors. Compared to the stepwise procedure, BN was approximately 10% and 7% more accurate for within line prediction in L1 and L2, respectively.

In terms of predictive ability, ANN performed better than linear models for TEP prediction, probably due to the existence of nonlinear relationships among traits. This result is expected given that non-linear components and potential interactions among predictors were not considered in the linear model. The algorithm of choice for ANN should be LM and a pre-selection of covariates is also recommended. In general, BN variable selection improved prediction accuracy of ANN. Within lines, correlations between observed and predicted TEP were higher than 0.7 (0.752 for L1 and 0.714 for L2), which indicate that partial egg production (EP1) can be a good predictor of TEP. Such predicted TEP could be used for management decisions in production systems. Somewhat different results were obtained by Savegnago et al. (2011) working with chicken layers. The authors found that early partial production periods were not good indicators of total egg production in a breeding context, concluding that birds should be selected based on TEP evaluations. This indicates that model choice and prediction ability are species and trait dependent.

In conclusion, BN is a good tool to describe distributions in a more parsimonious way for improving generalization. Also, BN worked well with ANN for pre-selection of input variables for TEP prediction in European quails. Lastly, growth traits were not good

predictors of TEP. As such, results indicate that to achieve reasonable predictive ability for TEP, some early egg production measurements are necessary.

ACKNOWLEDGMENTS

Financial support by the Wisconsin Agriculture Experiment Station and by COBB-Vantress, Inc. (Siloam Springs, AR) to Vivian Felipe and Guilherme Rosa is acknowledged. Funding for data collection was provided by the National Council of Scientific and Technological Development (CNPq) and the Foundation for Research Support of the State of Minas Gerais (FAPEMIG/APQ-00739-11), Brazil. The authors would like to thank also J. Susko-Parrish for editorial assistance.

REFERENCES

- Ahmad, H. A. 2009. Poultry growth modeling using neural networks and simulated data. *J. Appl. Poult. Res.* 18:440–446.
- Ahmad, H. A., and A. Golian. 2008. Neural network model for egg production curve. *J. Anim. Vet. Adv.* 7:1168–1170.
- Aliferis, C. E., A. Stanikov, I. Tsamardinos, S. Mani, and X. D. Koutsoukos. 2010. Local causal and Markov Blanket induction for causal discovery and feature selection for classification Part I: Algorithms and empirical evaluation. *J. Mach. Learn. Res.* 11:171–234.
- Bishop, C. M. 2006. *Pattern Recognition and Machine learning*. Springer, New York.
- Burnham, K. P., and D. R. Anderson. 2002. *Model Selection and Multimodel Inference: a Practical Information-Theoretic Approach*. 2nd ed. Springer Verlag, New York.
- Faridi, A., and A. Golian. 2011. Use of neural network models to estimate early egg production in broiler breeder hens through dietary nutrient intake. *Poult. Sci.* 90:2897–903.
- Fulton, J. E. 2012. Genomic selection for poultry breeding. *Animal Frontiers*. 2:30–36.
- Gianola, D., H. Okut, K. A. Weigel, and G. J. M. Rosa. 2011. Predicting complex quantitative traits with Bayesian neural networks: a case study with Jersey cows and wheat. *BMC Genet.* 12:87.
- Ginzburg, L. R., and C. X. J. Jensen. 2004. Rules of thumb for judging ecological theories. *Trends Ecol. Evol.* 19:121–126.
- Hastie, T., R. Tibshirani, and J. Friedman. 2009. *The elements of statistical learning*. Springer, New York.
- Hurvich, C. M., and C. L. Tsai. 1990. The impact of model selection on inference and linear regression. *Am. Stat.* 44:214–217.
- Jensen, F. V. 1996. *An introduction to Bayesian Networks*. Springer Verlag, New York.
- Lauritzen, S. L., and D. J. Spiegelhalter. 1988. Local computations with probabilities on graphical structures and their applications to expert systems. *J. R. Stat. Soc.* 50:157–224.
- Margaritis, D. 2003. *Learning Bayesian network model structure from data*. PhD Diss. Carnegie Mellon University, Pittsburgh.
- Mehri, M. 2012. Development of artificial neural network models based on experimental data of response surface methodology to establish the nutritional requirements of digestible lysine, methionine, and threonine in broiler chicks. *Poult. Sci.* 91:3280–3285.
- Mehri, M. 2013. A comparison of neural network models, fuzzy logic, and multiple linear regression for prediction of hatchability. *Poult. Sci.* 92:1138–42.
- Pearl, J. 1997. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 2nd ed. Morgan Kaufmann, San Mateo, CA.
- Pearl, J. 2000. *Causality: Models, Reasoning and Inference*. 2nd ed. Cambridge University Press, New York.
- Savegnago, R. P., B. N. Nunes, S. L. Caetano, A. S. Ferraudo, G. S. Schmidt, M. C. Ledur, and D. P. Munari. 2011. Comparison of logistic and neural network models to fit the egg production curve of White Leghorn hens. *Poult. Sci.* 90:705–11.
- Scutari, M. 2010. Learning Bayesian Networks with the bnlearn R package. *J. Stat. Softw.* 35:1–22.
- Semsarian, S., M. P. E. Nasab, S. Zarehdaran, and A. A. Dehghani. 2013. Prediction of the weight and number of eggs in Mazandaran native fowl using artificial neural network. *International Journal of Advanced Biological and Biomedical Research* 5:532–537.
- Spirites, P., C. Glymour, and R. Scheines. 2000. *Causation, Prediction and Search*. 2nd ed. The MIT Press, Cambridge, Massachusetts.
- Tsamardinos, I., L. E. Brown, and C. F. Aliferis. 2006. The max-min hill-climbing Bayesian network structure learning algorithm. *Machine Learning*. 65:31–78.
- Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behavior? *J. Anim. Ecol.* 75:1182–1189.