Proceedings, 10th World Congress of Genetics Applied to Livestock Production

# Neural Networks to Predict Breeding Values of Egg Production Using Phenotypic Information

*V. A. R. Cruz*[1+], R. P. Savegnago[1*], G. S. Schmidt[2], M. C. Ledur[2], and D. P. Munari[1#]
[1]Univ. Estadual Paulista – FCAV/UNESP, Jaboticabal, Brazil, [2]Embrapa Suínos e Aves, Concórdia-SC, Brazil.

**ABSTRACT:** The objective of this study was to use MLP ANN to learn to predict the breeding values for total egg production (EBV-TEP) with the phenotypic records of traits that presented genetic correlation with the egg production. The EBV-TEP for 1,273 birds were predicted using the BLUP in a single-trait animal model that included hatch as fixed effect and additive genetic and residual as random effects. The inputs of the multilayer perceptron (MLP) neural networks were the phenotypic records of total egg production (TEP), age at first egg (AFE), body weight at 62 weeks of age (BW62) and egg weight at 40 weeks of age (EW40), measured on the. The results suggest that neural networks could be efficient to predict the breeding values for TEP using phenotypic measurements of the birds and the family information, without using the pedigree information.
**Key words:** body weight; breeding value; egg production; multilayer perceptron

## Introduction

In genetic breeding programs for laying hens, the breeding goals are to improve the egg production and other egg quality traits. The simultaneous selection for many traits is possible due to the genetic correlation among then, like as age at first egg, body weight and egg weight (Besbes et al., 1992; Dana et al., 2011; Savegnago et al., 2011). Artificial neural networks (ANNs) have been applied to genomic evaluation of animals and plants using the information of genetic markers (Long et al. 2010; Gianola et al., 2011; González-Camacho et al., 2012; Pérez-Rodríguez et al., 2012; Neves et al., 2012; Okut et al., 2013; Tusell et al., 2013). But, many breeding programs do not use yet genomic information to do the genetic evaluation. Thus, maybe it is possible to use ANNs to map the breeding values of the total egg production obtained by traditional pedigree-based evaluations with phenotypic records of genetic correlated traits, without using the pedigree information. Multilayer perceptron (MLP) is one of the most used ANN with supervised learning and is capable of mapping input to output variables using many kinds of non-linear functions.

The objective of this study was to use MLP ANN to learn to predict the breeding values for total egg production (EBV-TEP) with the phenotypic records of traits that presented genetic correlation with the egg production.

## Materials and Methods

**Phenotype Measurement.** The dataset was composed of phenotypic records measured on hens from a White Leghorn population named "CC", developed and maintained under multi-trait selection for seven generations by Embrapa (Empresa Brasileira de Pesquisa Agropecuária; Brazilian Agricultural Research Corporation) Pigs and Poultry National Research Center, Concórdia, Santa Catarina, Brazil. It was used records of 1273 laying hens from three hatches for percentage of total egg production (TEP) from 17 to 70 weeks of age (the eggs gathered over five days of the week); age at first egg (AFE), measured as the number of days that elapsed until the first egg was laid; body weight at 62 weeks of age (BW62) in grams; and egg weight at 40 weeks of age (EW40) in grams. The pedigree contained 12,132 birds from seven generations.

**Genetic Analysis.** The variance components of the total egg production were estimated using a single-trait animal model by the REML methods implemented in MTDF program (Boldman et al., 1995). The animal model included hatching as a fixed effect and additive genetic and residual as random effects. The breeding values for that trait (EBV-TEP) were predicted using the BLUP.

**Neural Network:** The MLP was used to approximate the breeding values for total egg production with the phenotypic records of the other correlated traits. Eight scenarios were defined according to the input information used to predict the breeding values (Table 1). The inputs of the MLP were the phenotypes of AFE, BW62 and EW40, and the average phenotypic records per sire (PSP) and dam (PDP) of each progeny, for each trait. The EBV-TEP from BLUP was used in the output to training the MLP to predict them. The inputs of the MLP were connected to each hidden neuron via synaptic weights ($w_{kj}$, k = 1, 2, …, n hidden neuron). The weighted input of each k hidden neurons ($a_k$) was calculated by $a_k = b_k + \sum_{j=1}^{n} w_{kj} x_j$, in which $b_k$ was the bias term and $\sum_{j=1}^{n} w_{kj} x_j$ was each input weighted by its synaptic weight. The $a_k$ was transformed using the logistic and the hyperbolic tangent activation functions g(.). The logistic function was described as follow:

$$g(a_k) = \frac{1}{1 + e^{-a_k}}$$

where *e* is the Euler number.

+CNPq-PDJ Fellowship 150285/2014-0; *FAPESP Fellowship 2013/20091-0; #CNPq fellowship

The hyperbolic tangent activation function was as follow:

$$g(a_k) = \frac{e^{a_k} - e^{-a_k}}{e^{a_k} + e^{-a_k}}$$

The activated input of the hidden neurons were send to the output layer as $a_l = \sum_{k=1}^{m} w_k g_k \left( b_k + \sum_{j=1}^{n} w_{kj} x_j \right) + b_l$, where $a_l$ is the activated input of output neuron and $b_l$ is the bias of the output neuron. After, $a_l$ was activated using a linear function. For each scenario, 1000 MLP were constructed. The MLP ANNs were trained by means of the BFGS (Broyden Fletcher Goldfarb Shanno) backpropagation algorithm. The MLPs were constructed using the Automated Network Search function of the Statistica software 8.0 (StatSoft Inc., 2008), which optimizes the construction of neural networks and selects those that had better predictive performance according to the "intelligent problem solver".

**Table 1. Traits used as inputs of MLPs in each scenario.**

| Source of record | Traits[4] | Scenario[5] 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Bird[1] | TEP | x | x | x | x | x | x | x | x |
| | AFE | | x | | x | | x | | x |
| | BW62 | | x | | x | | x | | x |
| | EW40 | | x | | x | | x | | x |
| PDP[2] | PDP-TEP | | | x | x | | | x | x |
| | PDP-AFE | | | | x | | | | x |
| | PDP-BW62 | | | | x | | | | x |
| | PDP-EW40 | | | | x | | | | x |
| PSP[3] | PSP-TEP | | | | | x | x | x | x |
| | PSP-AFE | | | | | | x | | x |
| | PSP-BW62 | | | | | | x | | x |
| | PSP-EW40 | | | | | | x | | x |

[1]Records of phenotypic traits on birds.
[2]Average of records of phenotypic traits of dams' progenies.
[3]Average of records of phenotypic traits of sires' progenies.
[4]TEP = percentage of total egg production; AFE = age at first egg; BW62 = body weight at 62 weeks of age; EW40 = egg weight at 40 weeks of age.
[5] Records of traits used in each of the eight scenarios.

**Training, testing and validation sets:** The dataset was divided into three subsets: 60% of the data were used to training the MLPs, 20% to test and 20% to validate them. The training set allowed the MLP to learn the relationship between the input and output datasets. The test and validation phases ensured that the trained neural network would have good ability to generalize, thereby avoiding the overfitting.

**Predictions' accuracy:** The best MLP ANNs for predicting the EBV-TEP in each scenario were selected based on the accuracy of the models measured in each phase. The accuracy of the models was measured using the coefficient of determination ($R^2$), mean square error (MSE) and Spearman rank correlation between EBV-TEP from BLUP (desired output) and EBV-TEP from MLP (predicted output). The MSE was calculated using the following expression:

$$MSE_i = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} | y_{ij} - \hat{y}_{ij} |^2}{n}$$

which $\mathbf{y_{ij}}$ is the observed record **i** of bird **j**, $\hat{y}_{ij}$ is the estimated record **i** of bird **j** and **n** is the number of observations. Smaller values of MSE indicate better fitting of the models.

The Spearman's rank correlation was calculated between the EBV-TEP obtained by BLUP and those predicted by the MLPs.

## Results and Discussion

The $R^2$ increased from the first to the second scenarios, when AFE, BW62 and EW40 were added as input variables together with TEP (Table 2). The same results was observed from the 3rd to the 4th, from the 5th to the 6th and from the 7th to the 8th scenario (Table 2), when the traits related to information of the birds` families (PDP and PSP for TEP, AFE, BW62 and EW40) were added as inputs of the MLPs. Higher $R^2$ values were also observed from the 2nd to the 4th, from the 4th to the 6th and from the 6th to the 8th scenarios, while the MSE decreased.

The best predictions of EBV-TEP by the MLPs occurred when the information on family performance (PDP-TEP, PDP-AFE, PDP-BW62, PDP-EW40, PSP-TEP, PSP-AFE, PSP-BW62 and PSP-EW40) was included to the models (scenarios 4, 6 and 8 compared with scenario 2; and scenarios 3, 5 and 7 compared with scenario 1 - Table 2). The reduction in MSE occurred in scenario 8 (in which were used input variables from the birds, PDP and PSP) in relation to the scenario 2, in which was used only the traits of the birds (TEP, AFE, BW62 and EW40). Same results for the MSE reduction were found comparing scenarios 8 and 4 and scenarios 8 and 6, in which scenarios 4 and 6 were used less input traits than in scenario 8. So, the predictions of MLP for EBV-TEP became better as the information regarding the birds' families (PSP and PDP traits) were included in the input layer of the MLPs.

**Table 2. Adjusted coefficient of determination ($R^2$), Spearman's correlation (r) between the EBV-TEP obtained using BLUP and MLP, and mean square error (MSE) and in each set (training, testing and validation).**

| Scenario | Set | $R^2$ | r | MSE |
|---|---|---|---|---|
| 1 | Training | 0.29 | 0.48 | 0.0012 |
| | Testing | 0.31 | 0.51 | 0.0009 |
| | Validation | 0.39 | 0.54 | 0.0009 |
| 2 | Training | 0.56 | 0.72 | 0.0006 |
| | Testing | 0.51 | 0.68 | 0.0007 |

| | | | | |
|---|---|---|---|---|
| | Validation | 0.61 | 0.73 | 0.0006 |
| 3 | Training | 0.42 | 0.57 | 0.0009 |
| | Testing | 0.38 | 0.58 | 0.0009 |
| | Validation | 0.46 | 0.61 | 0.0009 |
| 4 | Training | 0.75 | 0.85 | 0.0004 |
| | Testing | 0.70 | 0.81 | 0.0004 |
| | Validation | 0.74 | 0.81 | 0.0004 |
| 5 | Training | 0.43 | 0.61 | 0.0009 |
| | Testing | 0.41 | 0.62 | 0.0008 |
| | Validation | 0.48 | 0.64 | 0.0008 |
| 6 | Training | 0.85 | 0.92 | 0.0002 |
| | Testing | 0.80 | 0.89 | 0.0003 |
| | Validation | 0.81 | 0.89 | 0.0003 |
| 7 | Training | 0.48 | 0.64 | 0.0008 |
| | Testing | 0.46 | 0.64 | 0.0008 |
| | Validation | 0.50 | 0.64 | 0.0009 |
| 8 | Training | 0.91 | 0.96 | 0.0001 |
| | Testing | 0.88 | 0.93 | 0.0002 |
| | Validation | 0.87 | 0.92 | 0.0002 |

Spearman's correlation (Table 2) between EBV-TEP obtained using BLUP and MLP ranged from 0.48 (training set in scenario 1) to 0.96 (validation set in scenario 8). This result indicated that the traits that were genetically correlated with TEP and measured on birds and its relatives (scenario 8), were important for the MLP to predict EBV-TEP.

## Conclusion

The neural networks were efficient to predict the breeding values for total egg production using phenotypic measurements of the birds and the family information, without using the pedigree information.

## Literature Cited


Besbes, B., Ducrocq, V., Foulley, J.L. et al. (1992). Genet. Sel. Evol. 24:539-552.

Boldman, K.G., Kriese, L.A.,Van Vleck L. D. et al. (1995).

Dana, N., Van Der Waaij, E.H., Van Arendonk, J.A.M. (2011). Trop. Anim. Health. Prod. 43:21-28.

Gianola, D., Okut, H., Weigel K.A., et al. (2011). BMC Genet. 12:87-100.

González-Camacho, J.M., de los Campos, G., Pérez-Rodríguez, P., et al. (2012). Theor. Appl. Genet. 125:759-771.

Long, N., Gianola, D., Rosa, G.J.M., et al. (2010). Genet. Res. 92:209-225.

Neves, H.H., Carvalheiro, R., Queiroz, S.A. (2012). BMC Genet. 8;13:100.

Okut, H., Wu, X.L., Rosa, G.J.M., et al. (2013). Genet. Sel. Evol. 11;45:34.

Pérez-Rodríguez, P., Gianola, D., González-Camacho, J.M., et al. (2012). G3 2:1595-1605.

Savegnago, R.P., Caetano, S.L., Ramos, S.B. et al. (2011). Poult. Sci. 90:2174-2188.

StatSoft Inc. (2008). Available in: <http://www.statsoft.com/support>.

Tusell, L., Pérez-Rodríguez, P., Forni, S., et al. (2013). Animal. 7:1739-1749.