J Biochem Technol (2021) 12(1): 45-48
https://doi.org/10.51847/pFILL2lk3G
ISSN: 0974-2328

Lasso Model for Morphological Covariation Patterns between *Colossoma macropomum* and *Piaractus orinoquensis* × *Colossoma macropomum* Hybrid

Manuel Milla Pino, Danny Villegas Rivas*, César Osorio Carrera, Nancy Carruitero Avila, Teresita Merino Salazar, Henry Díaz Merino, Carola Calvo Gastañaduy, Ricardo Shimabuku Ysa, Juan De La Cruz Lozado, River Chávez Santos, Dora Calvo Gastañaduy, Lillet Villavicencio Palacios

Received: 17 December 2020 / Received in revised form: 10 March 2021, Accepted: 14 March 2021, Published online: 23 March 2021

Abstract

Currently, great importance has been given to the study of external morphology, especially in fish, when it is used as a means of identifying hybrids. This paper considers a LASSO model based on the truss protocol to compare morphological covarion patterns between specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀)*. In this study, 25 specimens of *C. macropomum* and 20 specimens of the hybrid *P. orinoquensis (♂) × C. macropomum (♀)*, were analyzed, respectively. The method "Truss protocol" or "trusses" was used. LASSO model achieved to reduce the mean squared error. The final model obtained contains only seven covariates. LASSO model fitted on the morphological covariation patterns between specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀)* showed a good fit and allowed to correctly classify most of the specimens. Differences were observed in the area of the head and the anterior part of the fish evidenced in covariates associated with hydrodynamic abilities and with foraging.

Keywords: Morphometry, Truss protocol, Fishes, Lambda, Shrinkage regression

Introduction

The family Characidae is the most diverse family of freshwater fish species in South America (Diachkova *et al.*, 2019; Nurmayanti *et al.*, 2019). The implementation of morphometric analysis in some species provides scientific knowledge that helps genetic improvement. The morphological characters are physical evidence of the expression of the genotype. Therefore, the differences between specific body characteristics can become very important to establish patterns of differentiation and inheritance (Lazzarotto *et al.*, 2017). In continental fish, the morphometric characteristics referring to the anatomical shape have been used to evaluate the productive response in rearing both in natural environments and in captivity. Currently, there are more modern and precise morphometric analysis techniques, such as geometric morphometry (Bookstein *et al.*, 1985), which together with multivariate statistical analysis and means of direct visualization, constitute one of the most useful tools to describe the biological form and its changes.

Generally, these techniques are based on a set of measured distances between identifiable points on the organisms. In most cases, the measurements (distances between homologous points) present a high correlation, which is exploited in the models that are frequently used to compare between species. However, a variable selection model has desirable requirements: accurate predictions and interpretable models, and stability, that is, small changes in the data should not cause large changes in the predictors used. Traditional methods of variable selection, such as ridge regression, all subsets regression, or stepwise regression, fail one or more of the above requirements. The LASSO regression models (Hastie *et al.*, 2015) are based on the multiple linear models and seek to achieve its “regularization”. Although LASSO works successfully on many occasions, it has some limitations, which can be solved with the model known as Elastic Net (Ramos, 2018). In this sense, this paper considers the LASSO model based on the truss protocol to compare morphological covarion patterns between specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀)* when $p > n$, that is, we have more variables than observations using lars and glmnet package in R.

Materials and Methods

Morphological Covariation Patterns between C. macropomum and the hybrid P. orinoquensis (♂) × C. macropomum (♀)

In this study, 25 adult specimens of *C. macropomum* and 20 adult specimens of the hybrid *P. orinoquensis (♂) × C. macropomum (♀)* with an average weight of 600g, from artificial ponds of a fish farm in Portuguesa state, Venezuela, were analyzed. Within the sample of each species, there are mixed male and female

Manuel Milla Pino, Danny Villegas Rivas*
Faculty of Civil Engineering, Universidad Nacional de Jaén, Cajamarca, Perú.

César Osorio Carrera, Nancy Carruitero Avila, Teresita Merino Salazar, Henry Díaz Merino, Carola Calvo Gastañaduy, Juan De La Cruz Lozado, Lillet Villavicencio Palacios
Postgraduate School, César Vallejo University, Peru.

Ricardo Shimabuku Ysa
Faculty of Mechanical and Electrical Engineering, Universidad Nacional de Jaén, Cajamarca, Perú.

River Chávez Santos
Faculty of Economic and Administrative Sciences, Universidad Nacional ‘Toribio Rodríguez de Mendoza’ de Amazonas, Perú.

Dora Calvo Gastañaduy
Faculty of Education and Humanities, Universidad del Santa, Chimbote, Perú.

***E-mail:** danny_villegas1@yahoo.com

individuals. The method "Truss protocol" or "trusses" (Strauss & Bookstein, 1982) was used, which achieves an exhaustive reconstruction of the shape from the distances between the homologous anatomical landmarks **(Table 1)** and **(Figure 1)**. The distances connecting these landmarks form a series of continuous quadrilaterals with their respective internal diagonals **(Figure 1)**, which allows detecting differences in shape in the vertical, horizontal, and oblique directions. The limitation of this study is the number of measures necessary to achieve better efficiency in estimating parameters related to the morphology of these species.

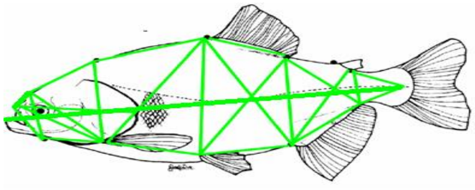

Figure 1. Location of Homologous Points and Distances Measured on the Left Lateral Profile of *C. macropomum* and the Hybrid *P. orinoquensis* (♂) × *C. macropomum* (♀)

Table 1. Truss Measurements from *C. macropomum* and the Hybrid *P. orinoquensis* (♂) × *C. macropomum* (♀) Specimens

Standard length (X_1)
Tip of the snout to end of the epiphyseal sulcus (X_2)
Tip of the snout to insertion of pectoral fin (X_3)
Anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4)
Anterior edge of the epiphyseal sulcus at the insertion of pectoral fin (X_5)
Anterior edge of the epiphyseal sulcus when articulating (X_6)
Articulate to insertion of pectoral fin (X_7)
Posterior edge of epiphyseal sulcus to end of dorsal fin (X_8)
Posterior edge of the epiphyseal sulcus at the insertion of the pelvic fin (X_9)
Posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10})
Posterior edge of the epiphyseal groove when articulating (X_{11})
Insertion of the pectoral fin to insertion of pelvic fin (X_{12})
Dorsal fin base (X_{13})
Anterior edge of dorsal fin to anterior edge of anal fin (X_{14})
Anterior edge of the dorsal fin to insertion of pelvic fin (X_{15})
Anterior edge of the dorsal fin to insertion of pectoral fin (X_{16})
Insertion of the pelvic fin to end of anal fin (X_{17})
Posterior edge of the dorsal fin to the fatty fin (X_{18})
Posterior edge of the dorsal fin to posterior edge of anal fin (X_{19})
Posterior edge of the dorsal fin to anterior edge of anal fin (X_{20})
Posterior edge of the dorsal fin to insertion of pelvic fin (X_{21})
Anal fin base (X_{22})
Posterior edge of the fatty fin to the last scale of the lateral line (X_{23})
Posterior edge of the fatty fin to posterior edge of anal fin (X_{24})
Posterior edge of the fatty fin to the anterior border of the anal fin (X_{25})
Posterior edge of the fatty fin to the anterior border of the anal fin (X_{26})
Eye diameter (X_{27})
Head length (X_{28})
Fat fin base (X_{29})

The morphological covariation patterns between specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀)* were studied using LASSO models in the R package (Team, 2020).

The LASSO Method

This method combines a regression model with a procedure for contracting some parameters towards zero and selecting variables, by imposing a restriction or penalty on the regression coefficients. Below is a formulation of Lasso as an optimization problem (for details see Ramos, 2018):

Suppose we have the data $(x_i, y_i), i = 1, 2, \dots, N$, where $x_i = (x_{i1}, \dots, x_{ip})$ t are the predictor variables and y_i are the responses. We can consider that the x_{ij} are standardized, that is,

$$\sum_i {x_{ij}}/{N} = 0, \tag{1}$$

$$\sum_i {x_{ij}^2}/{N} = 1 \tag{2}$$

or in other words, they have zero mean and variance 1. If the previous condition is not verified, it is enough to classify the variables as part of the preprocessing.

If we denote $\hat{\beta} = \left(\hat{\beta}_1, \dots, \hat{\beta}_p\right)^{\mathrm{T}}$, the estimate of lasso $\left(\hat{\alpha}, \hat{\beta}\right)$ is defined as the optimal solution to the optimization problem:

$$\text{mín}_{\alpha,\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \alpha - \sum_j \beta_j x_{ij} \right)^2 \right\} \\ \text{subject to } \sum_j \left|\beta_j\right| \leq t \tag{3}$$

where $t \geq 0$ is a fitting parameter.

Fixed β that satisfies $\sum_j \left|\beta_j\right| \leq t$, optimize in is a differentiable optimization problem in a variable, whose optimality condition is gradient equal to zero.

Prediction and Estimation of the Parameter t

We estimate the prediction error for the LASSO using cross-validation with k-folds.

If we call

$$s = \frac{t}{\sum_{i=1}^{p} \hat{\beta}_j^o}, \tag{4}$$

where $\hat{\beta}_j^o$ are the least-squares estimators, and we vary s in a sufficiently small interval, between 0 and 1, for each value of s or respectively of t, we obtain by cross-validation an estimator $\hat{e}\ (t)$, of mean square prediction error. We thus determine t^*, the value of *t* with smaller $\hat{e}\ (t)$, and this is the parameter considered.

Algorithms to Find Solutions

Once we have obtained an estimate of *t*, which we will call t^*, we proceed to solve the optimization problem;

$$\text{mín}_{\alpha,\beta} \sum_{i=1}^{N} (y_i - x_i^t \beta)^2 \\ \text{subject to } \sum_{i=1}^{p} \left|\beta_j\right| \leq t^* \tag{5}$$

We observe that the previous problem has p variables, since $\beta \in \mathbb{R}^p$, and a constraint; we can transform this restriction into 2^p linear restrictions:

$$\|\beta\|_1 \leq t^* \tag{6}$$

$$\sum_{i=1}^{p} |\beta_j| \leq t^* \tag{7}$$

$$\sum_{i=1}^{p} \beta_i^+ + \beta_i^- \leq t^* \tag{8}$$

$$\sum_{i=1}^{p} u_i\beta_i \leq t^* \ \forall (u_1, \dots, u_p) \in \{-1,1\}^p \tag{9}$$

The previous problem is a convex quadratic optimization problem with 2^p linear constraints. It is possible to obtain an equivalent formulation with a linear number in p of constraints, expanding the number of variables. For this, we make the change:

$$\beta = \beta_i^+ - \beta_i^-, \tag{10}$$

considering that β_i can be expressed as

$$\beta_i = \beta_i^+ - \beta_i^-, \tag{11}$$

With

$$\beta_i^+, \beta_i^- \geq 0. \tag{12}$$

from where

$$|\beta_i| = \beta_i^+ + \beta_i^-. \tag{13}$$

Therefore,

$$\begin{gathered} \text{mín}_{\beta^+,\beta^-} \sum_{i=1}^{N} \left(y_i - x_i^t(\beta_i^+ - \beta_i^-)\right)^2 \\ \text{subject to } \textstyle\sum_{i=1}^{p}(\beta_i^+ + \beta_i^-) \leq t^* \\ \beta^+, \beta^- \geq 0 \end{gathered} \tag{14}$$

This problem has 2^p variables since $\beta^+, \beta^- \in \mathbb{R}^p$, and $2p + 1$ constraints.

The Lars Package in R

Computes the prediction error of cross-validated K-fold mean squared for Forward Stagewise, LASSO, or LARS. For details, see Hastie and Efron (2013).

Results and Discussion

Table 2 and **Figure 2** show the fit of the LASSO model on patterns of morphological covariation between *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀).*, where the covariates (landmarks distances): eye diameter (X_{27}), anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4), posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10}), insertion of the pectoral fin to insertion of pelvic fin (X_{12}), posterior edge of the dorsal fin to the fatty fin (X_{18}), anterior edge of the dorsal fin to insertion of pectoral fin (X_{16}) and posterior edge of the dorsal fin to the posterior edge of anal fin (X_{19}) were included in the model, suggesting there are characteristics associated with the morphological covariation patterns that allow differentiation between redundant specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂) × C. macropomum (♀).* These covariates are associated with morphological covariation patterns that make a difference in the head area and the anterior part of the fish. These covariates are characteristics associated with hydrodynamic abilities and the foraging for food. **Figure 3** shows how lasso achieves, using their respective optimal values of λ, to reduce the MSE. The advantage of the final model obtained by lasso is that it is much simpler since it contains only seven covariates. These results coincide with those reported by Perdomo *et al.* (2017) who compared the morphometry of two continental fish species raised in Trujillo state, Venezuela, and those reported by Villegas *et al.* (2020a) in a multivariate analysis that allowed a morphometric comparison of a hybrid originated from *C. macropomum and P. orinoquensis*, and those reported by Villegas *et al.* (2020b) when studying the redundancy in morphological covariation patterns between *C. macropomum and P. orinoquensis*. However, the results differ from those indicated by Villegas *et al.* (2020c) when using a multiple logistic model to study the morphological covariation patterns between the mentioned species. The foregoing reveals what was indicated by Porras-Rivera and Rodríguez-Pulido (2019) and Conte-Grand *et al.* (2015), who pointed out that external morphology is not always reliable when used as the only means of identification, particularly for hybrid individuals beyond the first generation.

Table 2. LASSO Model Fitted on Morphological Covariation Patterns between *C. macropomum* and the Hybrid *P. orinoquensis (♂) × C. macropomum (♀).*

Land Marks Distance	LASSO model Coefficients
Intercept	3.7965292496
Anterior edge of the epiphyseal sulcus to the end of the epiphyseal sulcus (X_4)	-0.0116715619
Posterior edge of the epiphyseal sulcus to the insertion of the pectoral fin (X_{10})	0.0001085698
Insertion of the pectoral fin to insertion of pelvic fin (X_{12})	0.0159858221
Anterior edge of the dorsal fin to insertion of pectoral fin (X_{16})	0.0015759937
Posterior edge of the dorsal fin to posterior edge of anal fin (X_{19})	0.0011496300
Eye diameter (X_{27})	-0.1496373453
Fat fin base (X_{29})	-0.0200950546
% Deviance	87.13
Optimum Lambda (λ)	0.02401

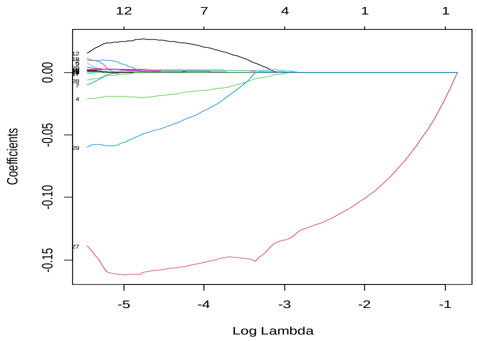

Figure 2. LASSO Adjustment on Morphological Covariation Patterns between *C. macropomum* and the Hybrid *P. orinoquensis (♂) × C. macropomum (♀).*

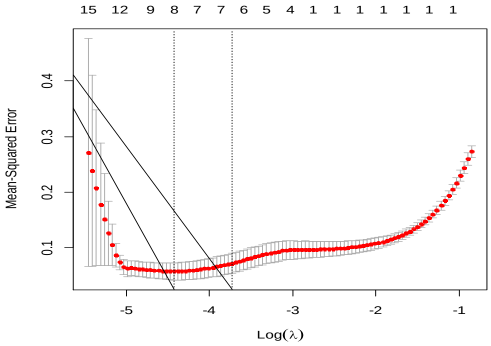

Figure 3. Mean Squared Error for Lambda (λ) in a LASSO Model on Morphological Covariation Patterns between *C. macropomum* and the Hybrid *P. orinoquensis (♂)* × *C. macropomum (♀)*

Conclusion

LASSO model achieved, using their respective optimal values of λ, to reduce the mean squared error. The final model obtained by LASSO it was much simpler since it contains only seven covariates. LASSO model fitted on the morphological covariation patterns between specimens of *C. macropomum* and the hybrid *P. orinoquensis (♂)* × *C. macropomum (♀)* showed a good fit and allowed to correctly classify most of the specimens. Differences were observed in the area of the head and the anterior part of the fish between the hybrid and its parent. The morphological differences between these two species were evidenced in covariates associated with hydrodynamic abilities and with foraging. Finally, the results of this research suggest the use of the LASSO model to compare morphological covariation patterns between the hybrid *P. orinoquensis (♂)* × *C. macropomum (♀)* and *P. orinoquensis* when the sample size is less than the number of landmarks (*n* < *p*).

Acknowledgments: We thank all the authors for their contribution to this research.

Conflict of interest: Authors have declared that no competing interests exist.

Financial support: The research was financed with own resources

Ethics statement: As per international standards or university standard, ethical approval has been collected and preserved by the authors.

References

Bookstein, F. L., Chernoff, B., Elder, R. L., Humphries, J. M., Smith, G., & Strauss, R. E. (1985). Morphometrics in evolutionary biology. The Academy of Natural Sciences of Philadelphia, Michigan.

Conte-Grand, C., Sommer, J., Ortí, G., & Cussac, V. (2015). Populations of odontesthes (teleostei: atheriniformes) in the Andean region of Southern South America: body shape and hybrid individuals. *Neotropical Ichthyology*, *13*(1), 137-150.

Diachkova, A., Tikhonov, S., & Tikhonova, N. (2019). The Effect of High Pressure Processing on the Shelf Life of Chilled Meat and Fish. *International Journal of Pharmaceutical Research & Allied Sciences*, *8*(3), 98-108.

Hastie, T. (2013). Efron B. Paquete larsenr. *CRAN. Rproject. org/package= Lars*.

Hastie, T., Tibshiriani, R., & Wainwright, M. (2015). *Statistical Learning with Sparsity. The Lasso and Generalizations*. Florida: Chapman & Hall/CRC.

Lazzarotto, H., Barros, T., Louvise, J., & Caramaschi, É. P. (2017). Morphological variation among populations of Hemigrammus coeruleus (Characiformes: Characidae) in a Negro River tributary, Brazilian Amazon. *Neotropical Ichthyology*, *15*(1), e160152.

Nurmayanti, I., Diantini, A., & Milanda, T. (2019). Measurement of knowledge risk factors of Lung Cancer disease in salted-fish-traders at Pangandaran Indonesia. *Journal of Advanced Pharmacy Education & Research*, *9*(4), 54-59.

Perdomo, D., Castellanos, K., Maffei-Valero, M., Gechele, J., Corredor, Z., Piña, J., Martínez, M., & Naranjo, A. (2017). Morphometric and meat yield comparison of two continental fish species raised in Trujillo state, Venezuela. *Academu Journal*, *16*(37), 83-95.

Porras-Rivera, G., & Rodríguez-Pulido, J. A. (2019). Morphometric comparison and characterization of the hybrid (Pseudoplatystoma metaense x Leiarius marmoratus) and its parental lines (Siluriformes: Pimelodidae). *International Journal of Morphology*, *37*(4), 1409-1415.

Ramos, L. (2018). LASSO regression. Facultad de Matemáticas. Universidad de Sevilla. España. 61 p.

Strauss, R. E., & Bookstein, F. L. (1982). The truss: body form reconstructions in morphometrics. *Systematic Biology*, *31*(2), 113-135.

Team, R. C. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Villegas, D., Milla, M., Castillo, O., & Durant, K. (2020a). Multivariate analysis in the morphometric comparison of the hybrid Colossoma macropomum X Piaractus brachypomus) and its parents. *Revista de Investigación en Agroproducción Sustentable*, *4*(1), 29-36.

Villegas, D., Milla, M., Pérez, Y., Villegas, S., Garrido, Z., Delgado, E., Ruiz, W., Velasquez, Y., & De Souza, B. (2020b). Redundancy in morphological covariation patterns between Colossoma macropomum and Piaractus orinoquensis. *Uttar Pradesh Journal of Zoology*, *41*(14), 37-46.

Villegas, D., Milla, M., Garrido, Z., Grados, M., Osorio, C., Delgado, E., Velasquez, Y., Ruiz, W., Shimabuku, R., Paredes, J., et al. (2020c). On a logistic model for morphological covariation patterns between colossoma macropomum and the hybrid colossoma macropomum (♀) x piaractus orinoquensis (♂). *Uttar Pradesh Journal of Zoology*, *41*(18), 28-34.