

Dear Editor,

We would be pleased if you consider our manuscript "Lost in translation: On the impact of data coding on penalized regression with interactions" for publication in Genetics Selection Evolution.

In recent scientific work on genomic prediction, a tendency towards models incorporating gene by gene or gene by environment interactions can be observed [1, 2, 3, 4]. It has been shown that the additive ridge regression best linear unbiased prediction is not affected by data coding [5, 6], but that models with products of marker values as additional predictor variables show a sensitivity to changes of the marker coding [7, 6].

In our work, we investigate the cause of this loss of invariance to changes in the marker coding, in a general setup of penalized regression on models defined by polynomials in the predictor variables. Our results are not limited to ridge regression, but for instance also apply to LASSO or to any other penalized regression, and to any type of interaction (GxG or GxE) modelled by products of predictor variables or by covariance matrices given by Hadamard products.

We acknowledge that our considerations are brief, and mainly of theoretical nature. However we are confident that a journal with focus on quantitative genetics as GSE is the right place to present our work, whether as a full length paper, or as a short communication, whichever suits best to the Journal.

Best regards,  
Johannes Martini  
Francisco Rosales

The following persons are suggestions for potential reviewers:

- Daniel Gianola, Department of Animal Science, University of Wisconsin-Madison, gianola@ansci.wisc.edu
- Gustavo de los Campos, Department of Epidemiology and Biostatistics, Michigan State University, gdeloscamos@epi.msu.edu

- [1] U. Ober, W. Huang, M. Magwire, M. Schlather, H. Simianer, T. F. C. Mackay, Accounting for genetic architecture improves sequence based genomic prediction for a *Drosophila* fitness trait, PLOS ONE 10 (2015) e0126880. doi:10.1371/journal.pone.0126880.
- [2] Y. Jiang, J. C. Reif, Modeling epistasis in genomic selection, Genetics 201 (2015) 759–768. doi:10.1534/genetics.115.177907.
- [3] J. W. R. Martini, V. Wimmer, M. Erbe, H. Simianer, Epistasis and covariance: how gene interaction translates into genomic relationship, Theor Appl Genet 129 (2016) 963–976. doi:10.1007/s00122-016-2675-5.
- [4] P. Pérez-Rodríguez, J. Crossa, J. Rutkoski, J. Poland, R. Singh, A. Legarra, E. Autrique, G. d. l. Campos, J. Burgueño, S. Dreisigacker, Single-step genomic and pedigree genotype  $\times$  environment interaction models for predicting wheat lines in international environments, The plant genome 10 (2).
- [5] I. Strandén, O. F. Christensen, Allele coding in genomic evaluation, Genetics Selection Evolution 43 (1) (2011) 25.
- [6] J. W. R. Martini, N. Gao, D. F. Cardoso, V. Wimmer, M. Erbe, R. J. C. Cantet, H. Simianer, Genomic prediction with epistasis models: on the marker-coding-dependent performance of the extended gblup and properties of the categorical epistasis model (CE), BMC Bioinformatics 18 (2017) 3. doi:10.1186/s12859-016-1439-1.
- [7] D. He, L. Parida, Does encoding matter? a novel view on the quantitative genetic trait prediction problem, BMC bioinformatics 17 (9) (2016) 272.