

Improving genomic prediction using secondary phenotypes

Bader Arouisse¹, Fred A. van Eeuwijk¹, Willem Kruijer¹

¹Biometris, Wageningen University and Research (Wageningen, Netherlands)

E-mail for correspondence: willem.kruijer@wur.nl

Abstract:

In the past decades, genomic prediction has had a large impact on plant and animal breeding. Given the current advances of high-throughput phenotyping, it is increasingly common to measure a large number of traits, in addition to the target trait of interest. This raises the important question whether these 'secondary' traits or 'co-data' can be used to improve genomic prediction for the target trait. With only a small number of secondary traits, this is known to be the case, given sufficiently high heritability and genetic correlations. Here we focus on the more challenging situation when many secondary traits are available. Secondary traits are then usually modelled using separate ridge penalties (Van De Wiel et al. (2016)), or, equivalently, through additional kernels (relatedness matrices). This approach is however infeasible when secondary traits are not measured on the test set, and cannot distinguish between genetic and residual correlations. Here we discuss three approaches that could potentially overcome one or both of these limitations. Our first approach relies on a dimension reduction, achieved using either random forests or penalized regression of the target on the secondary traits, while the second approach reduces the dimension using penalized selection indices (Lopez-Cruz et al. (2019)). In both cases, we use the bivariate GBLUP, with the fitted values as secondary trait. Finally, we extend existing multi-kernel approaches by replacing secondary traits by their genomic predictions, which we denote GM-BLUP. For most of our simulated datasets, prediction using selection indices was most accurate, while on real maize and arabidopsis data, the dimension reduction using random forests performed best. GM-BLUP performed only slightly better than existing multi-kernel methods, however being also applicable when secondary traits are not measured on the test set.

Key words: GBLUP, secondary traits, selection indices, penalized regression, random forests

Marco Lopez-Cruz, Eric Olson, Gabriel Rovere, Jose Crossa, Susanne Dreisigacker, Suchismita Mondal, et al. (2020). Regularized selection indices for breeding value prediction using hyper-spectral image data. *Sci Rep* 10, 8195.

Mark A. van de Wiel, Tonje G. Lien, Wina Verlaat, Wessel N. van, Wieringen and Saskia M. Wilting (2015). Better prediction by use of co-data: adaptive group-regularized ridge regression. *Statistics in Medicine*.