Master's thesis topic

Biostatistics unit (340c)

Comparison of novel genomic prediction methods with advanced variable selection techniques on synthetic and real datasets

Genomic prediction is an essential tool in breeding of many plant and animal species. It makes it possible to predict the performance of genotypes without physically testing them, and to select the best, thereby increasing genetic gain and reducing workload. The classic genomic prediction tool is a mixed model, but other linear, non-linear and Bayesian methods can be used instead.

In the past seven years, over 25 new single-trait genomic prediction methods have been published, at least half of which use R or Python. The availability of the source codes enables their comparative evaluation. Previous evaluations of these methods were often done for a limited set of conditions or exclusively with real data, and performance is typically assessed relative to established approaches such as Genomic Best Linear Unbiased Prediction (GBLUP), Ridge Regression Best Linear Unbiased Prediction (RRBLUP), or various Bayesian methods (A, B, C, R) as the benchmark or gold standard.

Thus, comprehensive direct comparison of the relative performance of these new methods under a wide variety of scenarios is still lacking but is essential for robust comparative evaluation.

Objectives

1. Review new methods which use advanced techniques for variable selection.
2. Perform cross-validation to comparatively assess method performance for simulated and real datasets.
3. Summarise and report the results.

Contact

For more details please contact Prof. Hans-Peter Piepho (hans-peter.piepho@uni-hohenheim.de) or Maksym Hrachov (maksym.hrachov@uni-hohenheim.de).