**Master thesis project**

**Deflated preconditioned conjugate gradient method applied to genomic evaluations**

Genetic progress by selection and mating is based on prediction of the ability of the parents to breed the most efficient progeny. Due to recent advances in molecular biology, this process of prediction is now being based on the genome of some animals, in the form of single nucleotide polymorphism (SNP) markers, in addition to phenotypic and pedigree information. An approach, called single-step genomic evaluation, is becoming the method of choice for such prediction, because it analyses simultaneously phenotypic, pedigree, and genomic information within the best linear unbiased prediction (BLUP) framework.

Currently, pedigree and phenotype datasets may contain several millions of records, and genotype datasets may include several thousand SNPs for several hundred thousand animals. A model that fits SNP markers as random effects was recently proposed to tackle this amount of information, and has been tested on large datasets. The associated large and sparse linear systems of equations were solved using the deflated preconditioned conjugate gradient (DPCG) method. The DPCG method is an attractive method, because it allows fast convergence by treating the unfavourable eigenvalues of the system matrix to which it is applied through a second-level preconditioner.

Nevertheless, the DPCG method as implemented in our software can still be improved. Indeed, a naive approach to approximate the same space as the span of the unfavourable eigenvalues is currently implemented. This approach does not account for some data properties, such as the low dimensionality of the genomic information. Furthermore, the current implementation could request too much additional memory for single-step evaluations with many traits of interest.

The aim of this Master thesis project is to investigate the definition of the (second-level) preconditioner used in the DPCG method, with the aim to optimize it for very large genomic evaluations, while maintaining the current performance.

We envision the following aspects in this Master thesis project:

1) literature review on single-step genomic models;

2) literature review on iterative solvers, such as the PCG and DPCG methods;

3) investigation of different definitions of a (second-level) preconditioner;

4) implementation and evaluation of the performance of the developed approaches on different test cases.

**Contact**

Prof. Kees Vuik (TU Delft; c.vuik@tudelft.nl), Jeremie Vandenplas (Wageningen UR; jeremie.vandenplas@wur.nl)

**References**Fernando, R.L., H. Cheng, B.L. Golden, and D.J. Garrick. 2016. Computational strategies for alternative single-step Bayesian regression models with large numbers of genotyped and non-genotyped animals. Genet. Sel. Evol. 48:96.

Vandenplas, J., H. Eding, M.P.L. Calus, and C. Vuik. 2018. Deflated preconditioned conjugate gradient method for solving single-step single nucleotide polymorphism BLUP. Paper 11.25 in Proceedings of the 11th World Congress on Genetics Applied to Livestock Production, Auckland, New Zealand.