



Mixed Model B-splines for high dimensional phenotypic data in plant breeding

Martin P. Boer

Biometris WUR, Wageningen, The Netherlands - martin.boer@wur.nl

One of the aims in plant breeding is to develop new varieties which perform well under stress conditions, such as severe drought and high temperatures. To detect the underlying genetic reasons why certain varieties perform better under stress conditions, it is important to follow the development of the plants over time, for example by automatic measuring plant growth in greenhouses or by remote sensing techniques. This automatic measuring of plant development, called high-throughput phenotyping, has developed rapidly over the last fifteen years. One of the big statistical challenges is to develop efficient models for this type of big datasets: typically hundreds of genotypes are continuously measured over a long time range, with thousands of time points per genotype, resulting in millions of observations for the full data set.

For the analysis of high-throughput phenotyping data, a fast and efficient mixed model is presented. The development over time for each genotype is modelled by B-splines. B-splines are computationally attractive, because of their local support. It is shown that sparseness of the corresponding mixed model equations is a key element of the new algorithm. In existing mixed model approaches based on B-splines, the corresponding linear system of mixed model equations is dense, which makes it computationally very difficult to solve.

One of the important elements of the new method is the modelling of the interaction between genotype and environment over time, since one of the main questions to answer is how the different varieties react on environmental conditions, such as for example the change of temperature during the day.

Keywords: Sparse Matrices; REML; Drought stress; Genotype by Environment interaction