P-spline models for agricultural field trials

María Xosé Rodríguez-Álvarez[1], *Paul H.C. Eilers*[2], *Martin Boer*[3], *Fred van Eeuwijk*[3]

[1]mxrodriguez@uvigo.es, Department of Statistics and OR, University of Vigo, Spain
[2]p.eilers@erasmusmc.nl, Erasmus Medical Center, Rotterdam, The Netherlands
[3]martin.boer@wur.nl & fred.vaneeuwijk@wur.nl, Biometris, University of Wageningen, The Netherlands

Spatial variation is common in agricultural field trials. Many factors combine to generate micro-environments that differ from plot to plot, strongly influencing yield and other phenotypes. It is necessary to correct for them when estimating genotype effects. A part of the spatial variation can be attributed to systematic effects, caused by the way the field was prepared before and during sowing or planting. It is relatively easy to add factors to a statistical model to account for them. Random spatial variation is harder. There are no covariates to relate it to, so it is necessary to include a model component for a random field. A common approach is to describe spatial effects as correlated random noise (Gilmour et al., 1997) by assuming a separable autoregressive (AR) process. However, modeling spatial effects as correlated noise presents some drawbacks. The separable AR model can be limited in its flexibility, and estimation for large fields is slow and frequently unstable. Moreover, the interpretation of the results is not easy.

As an alternative, in this talk we explore modeling the spatial trend variation explicitly by means of tensor product B-splines. In the spirit of P-splines, we use anisotropic penalties to tune smoothness (Eilers and Marx, 2003). As far as estimation is concerned, we propose the use of the *Separation of Anisotropic Penalties* (SAP) algorithm recently proposed by Rodríguez-Álvarez et al. (2014). The algorithm possess a number of valuable properties: (1) it is fast and stable; (2) it converges in few iterations from any starting value; (3) it always provides positive estimates of the variance components; and (4) it allows to build up more complex models by including both fixed and random effects. All these characteristics make the SAP algorithm a good candidate for the analysis of agricultural field trials. On top of the spatial field, effects of genotypes and families of varieties are included in the model, as well as block and/or replication effects or other sources of spatial variation. We illustrate our proposal with challenging experimental data.

Keywords: Field Trials, Spatial Variation, Tensor-product P-splines, SAP algorithm.