Spatial Models for Field Trials

María Xosé Rodríguez - Álvarez[1,2], Martin P. Boer[3], Fred A. van Eeuwijk[3], Paul H. C. Eilers[4]

[1] Department of Statistics and Operations Research
and Biomedical Research Centre (CINBIO), University of Vigo
Campus Lagoas-Marcosende s/n, 36310 Vigo, Spain.
mxrodriguez@uvigo.es
[2] BCAM - Basque Center for Applied Mathematics, Bilbao, Spain
[3] Biometris, Wageningen University & Research, Wageningen, the Netherlands
[4] Erasmus University Medical Centre, Rotterdam, the Netherlands

July 29, 2016

Abstract

An important aim of the analysis of agricultural field trials is to obtain good predictions for genotypic performance, by correcting for spatial effects. In practice these corrections turn out to be complicated, since there can be different types of spatial effects; those due to management interventions applied to the field plots and those due to various kinds of erratic spatial trends. This paper presents models for field trials in which the random spatial component consists of tensor product Penalized splines (P-splines). A special ANOVA-type reformulation leads to five smooth additive spatial components, which form the basis of a mixed model with five unknown variance components. On top of this spatial field, effects of genotypes, blocks, replicates, and/or other sources of spatial variation are described by a mixed model in a standard way. We show the relation between several definitions of heritability and the effective dimension or the effective degrees of freedom associated to the genetic component. The approach is illustrated with large-scale field trial experiments. An R-package is provided.

1 Introduction

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 traits. It is necessary to correct for them when estimating genotypic 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. A familiar example are row and column effects, caused by the movements of machines during ploughing, tilling and other procedures. It is relatively easy to add factors to a statistical model to account for them.

Random spatial variation such as for example fertility trends is harder to model. There are no covariates (like row numbers) to relate it to, so it is necessary to include a model

component for a random field. Roughly speaking, there are two main approaches to model spatial trends: one based on spatial variance-covariance structures; and the other based on smoothing techniques. In the first case, a spatially correlated stochastic component is included into the spatial model. However, this is non-trivial, as correlation in two directions, along the rows and columns of the field has to be modeled. To keep the effort manageable, several assumptions need to be made, and it has become standard to consider separability and stationarity (see e.g. Zimmerman and Harville, 1991). Important contributions in this area are the separable autoregressive model proposed by Cullis and Gleeson (1991) and extended in Gilmour et al. (1997), the separable linear variance model discussed in Piepho and Willians (2010), or the Bayesian first-differencing model in rows and columns given in Besag and Hidgon (1999). As a complementary approach, smoothing methods model spatial trend variation explicitly. The use of smoothing techniques in the agricultural context dates back to Green et al. (1985), and it has been further described and extended, among others, by Durban et al. (2003) and Verbyla et al. (1999). To the best of our knowledge, this modeling technique has been mainly approached in the statistical literature in the one dimensional case, i.e., through separate (or additive) smoothed trend effects along the rows and columns of the field. However, while these approaches have proved useful for modeling large-scale dependence (or global trend), they suffer from the limitation of not always being able to capture small-scale dependence (local trend). As a consequence of this limitation, the inclusion of spatially correlated components might still be necessary (Gilmour et al., 1997; Verbyla et al., 1999).

As an alternative, this paper explores the use of two-dimensional smooth surfaces. We propose the use of tensor product Penalized splines (P-splines, Eilers and Marx, 2003) to explicitly model both sources of spatial dependence. P-splines were introduced by Eilers and Marx (1996), as a simplification of a proposal by O'Sullivan (1986). P-splines approach smoothing as penalized regression: a rich B-spline basis is combined with a penalty on (higher order) differences of the B-spline coefficients to avoid overfitting, and estimation is based on penalized ordinary least squares. As it will be seen, the mixed model representation of P-splines (Currie and Durban, 2002; Wand, 2003) provides us with a general framework for the analysis of field trials. It allows the inclusion of both extra fixed and random components, such as genotypic effects or the correction for rows and columns. Besides, using nested B-spline bases (Lee et al., 2013) the computational effort of our approach, which we call SpATS, is moderate, even for large fields trials. Our SpATS model has a number of other attractive properties: (1) an explicit estimate of the spatial trend in the field is obtained; (2) estimation is stable and fast; (3) missing plots, even a large fraction of them, are easily handled; and (4) extension to a non-normal response, along the lines of the generalized linear model, is straightforward.

We should mention that our approach is not completely new in the agricultural literature. In Taye and Njuho (2008) and Robbins et al. (2012) the authors discuss similar approaches in the context of field experiments, and in forest research the topic has been covered by Cappa and Cantet (2008). This paper goes one step further by proposing a fully anisotropic penalized approach framed within the mixed-effects model context. Use is made of the P-spline ANOVA-type (PS-ANOVA) approach presented in Lee et al. (2013), which gives rise to a model with five smooth spatial components each having a clear interpretation. We also

show the link between the generalized definitions of heritability proposed by Cullis et al. (2006) and Oakey et al. (2006) and the notion of effective dimension of model components, a well-known complexity measure in the smoothing context (Hastie and Tibshirani, 1990). Finally, we provide software for the practical application of our proposal in a free and easy-to-use R-package (R Core Team, 2016), called SpATS.

The rest of the paper is structured as follows. We start by motivating our approach in Section 2. Section 3 presents background on B- and P-splines in one and two dimensions, including their representation as mixed models. They form the basis for spatial models, which are presented in Section 4. Simulations comparing our SpATS model and that of Gilmour et al. (1997) can be found in Section 5 and Section 6 presents several applications to large-scale field trials. A Discussion Section closes the paper. Some technical details have been added as Appendices, where we also describe the R-package that accompany this paper.

2   Motivating example

Uniformity field trials are trials in which a single genotype or variety is evaluated. In practice, the interest of such field trials is that its statistical analysis can help understanding the different sources of spatial variation present in a field, and thus serve as guidance for the design and subsequent analyses when genetic effects are to be evaluated. In this section we present a series of analyses of a set of barley uniformity data discussed in the paper by Williams and Luckett (1988). We focus here on presenting the big picture of our approach, leaving the more technical details to subsequent sections.

In this experiment, plots were laid out in a 15 row by 48 column grid, and the phenotypic trait of interest was yield. Figure 1(a) depicts the raw yield data. Note that there is a rather complex spatial pattern, with patches presenting larger/smaller yield values. Let y_i denotes the yield data (in kg per hectare divided by 10) obtained at plot i $(i = 1, \ldots, 720)$, and u_i and v_i the row and column position respectively, both centered and scaled. A common strategy in the analysis of field trial experiments is to use the following statistical model as starting point

$$\boldsymbol{y} = \mathbf{1}_{720}\beta_0 + \boldsymbol{Z}_r\boldsymbol{c}_r + \boldsymbol{Z}_c\boldsymbol{c}_c + \boldsymbol{\varepsilon}, \tag{1}$$

where $\boldsymbol{y} = (y_1, \ldots, y_{720})^t$, $\mathbf{1}_n$ is a column vector of ones of length n, and $\boldsymbol{c}_r = (c_{r1}, \ldots, c_{r15})^t$ and $\boldsymbol{c}_c = (c_{c1}, \ldots, c_{c48})^t$ are, respectively, the random effect coefficients for the rows and columns with $\boldsymbol{c}_r \sim N\left(\boldsymbol{0}, \sigma_r^2 \boldsymbol{I}_{15}\right)$ and $\boldsymbol{c}_c \sim N\left(\boldsymbol{0}, \sigma_c^2 \boldsymbol{I}_{48}\right)$ and associated design matrices $\boldsymbol{Z}_r$ and $\boldsymbol{Z}_c$. Finally, the random error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_{720})^t \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{720}\right)$.

Figure 1(b) depicts the empirical best linear unbiased predictors (BLUPs) for the row and column random factors. As can be observed, and especially for the rows, the BLUPs show a clear evidence of a pattern, indicating that the independent Gaussian distribution assumption does not hold. Besides, the residuals' spatial plot ($\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \mathbf{1}_{720}\widehat{\beta}_0 + \boldsymbol{Z}_r\widehat{\boldsymbol{c}}_r + \boldsymbol{Z}_c\widehat{\boldsymbol{c}}_c$) shown also in Figure 1(b) suggests that the complex spatial pattern has not been completely captured, and thus a more complex statistical analysis is required. To that end, we propose a modeling strategy based on incorporating a model (spatial) component that simultaneously accounts for the spatial trend across both directions of the field. Specifically, a smooth bivariate surface,

jointly defined over the row and column positions, is assumed

$$\boldsymbol{y} = f(\boldsymbol{u}, \boldsymbol{v}) + \boldsymbol{Z}_r \boldsymbol{c}_r + \boldsymbol{Z}_c \boldsymbol{c}_c + \boldsymbol{\varepsilon}, \tag{2}$$

where $\boldsymbol{u} = (u_1, \ldots, u_{720})^t$, $\boldsymbol{v} = (v_1, \ldots, v_{720})^t$, and $f(\boldsymbol{u}, \boldsymbol{v}) = (f(u_1, v_1), \ldots, f(u_{720}, v_{720}))^t$, with $f(\cdot, \cdot)$ representing a smooth bivariate function. Note that the intercept, β_0, is embedded as part of $f(u, v)$. To have a better understanding of the interpretation of $f(\cdot, \cdot)$, we can further decompose it in a nested-type ANOVA structure

$$f(\boldsymbol{u}, \boldsymbol{v}) = \underbrace{\boldsymbol{1}_n \beta_0 + \boldsymbol{u}\beta_1 + \boldsymbol{v}\beta_2 + \boldsymbol{u} \odot \boldsymbol{v}\beta_3}_{\text{Bilinear polynomial}} + \underbrace{f_u(\boldsymbol{u}) + f_v(\boldsymbol{v}) + \boldsymbol{u} \odot h_v(\boldsymbol{v}) + \boldsymbol{v} \odot h_u(\boldsymbol{u}) + f_{u,v}(\boldsymbol{u}, \boldsymbol{v})}_{\text{Smooth part}},$$

where $\odot$ denotes the element-wise vector (matrix) product. There are now two components: the bilinear polynomial and the smooth part. The bilinear (or parametric) component includes the intercept (β_0), the linear trends along the row (β_1) and column (β_2) directions, as well as the linear interaction trend (β_3). In addition, the smooth component is responsible for modeling the deviation from this compound linear trend. Here,

- $f_u(u)$ is a smooth trend along the rows, identical for all columns (i.e., a main smooth effect).

- $f_v(v)$ is a smooth trend along the columns, identical for all rows.

- $vh_u(u)$ and $uh_v(v)$ are linear-by-smooth interaction trends. For instance, $uh_v(v)$ is a varying coefficient surface trend, consisting of functions, linear in the rows, for each column, but with slopes that change smoothly along the columns, $h_v(v)$ (the same holds for $vh_u(u)$).

- $f_{u,v}(u, v)$ is a smooth-by-smooth interaction trend jointly defined over the row and column directions.

The functions f_u, f_v, h_u and h_v are constructed with variations on one-dimensional P-splines, while $f_{u,v}$ is based on tensor product P-splines. It may come as a surprise that six components are introduced to model the surface f in (2) (the bilinear polynomial and the five smooth trends). The reason is that this decomposition translates model (2) directly to a standard mixed model. In fact, for each of the smooth components the desirable amount of smoothing is computed using restricted maximum likelihood (REML, Patterson and Thompson, 1971). This decomposition also allows to reduce the size of the interaction component $f_{u,v}$ for very large fields, to save computation time (Lee et al., 2013). The technical details will be presented in Section 3.

Figure 1(c) shows the estimated spatial trend across the field, i.e. $\widehat{f}(\cdot, \cdot)$, but excluding the intercept. A nice property of our proposal is that it allows depicting the spatial trend in a grid finer than the number of rows and columns, facilitating results interpretation. Note that we recover quite successfully the spatial variation observed in the raw data. The residuals' spatial plot and BLUPs for $\boldsymbol{c}_r$ and $\boldsymbol{c}_c$ shown also in Figure 1(c) suggest that the spatial independence assumption for the error vector $\boldsymbol{\varepsilon}$ might be appropriate, and that no trend is now present
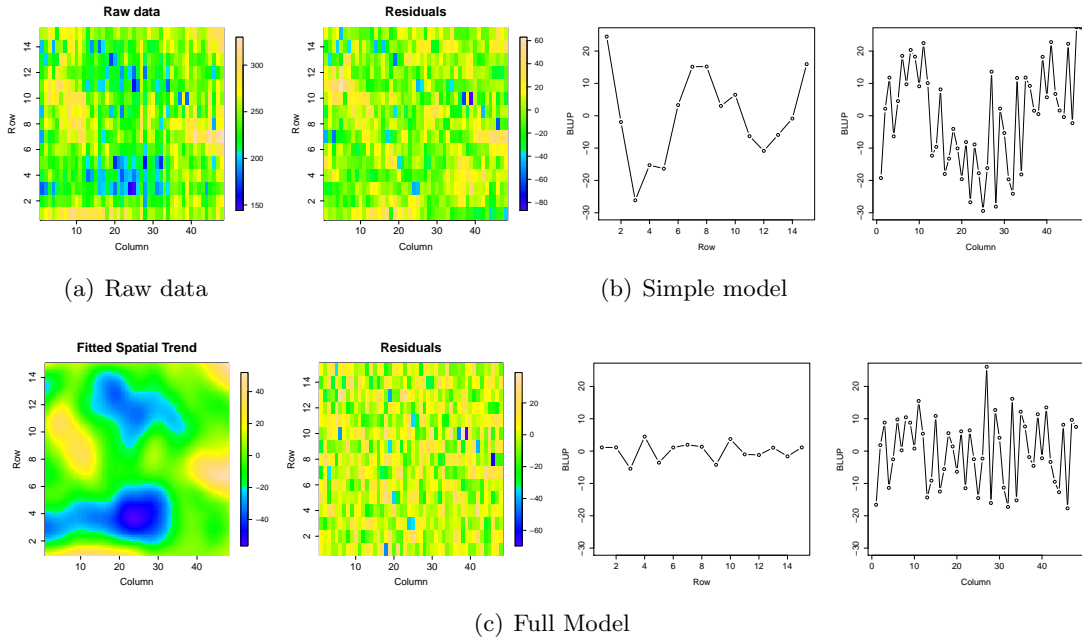
4

(a) Raw data   (b) Simple model

(c) Full Model

Figure 1: For the barley uniformity data: Raw data, residuals' spatial plot, best linear un-biased predictions (BLUPs) for the row and column random factors and contour plot of the estimated spatial trend based on the models including (b) only the row and column random factors (Simple model) (c) the smooth spatial trend (Full model)

in the BLUPs. Figure 2 shows the bilinear and smooth components of the ANOVA-type decomposition discussed above. Note that the estimated smooth functions defined over the rows $(\hat{f}_u)$ and columns $(\hat{f}_v)$ capture the trends observed in the BLUPs for the row and column analysis (model (1) and Figure 1(b)). When we compare the estimated linear-by-smooth interactions trends, we observe that the contribution of $v\hat{h}_u(u)$ to the fitted spatial trend is stronger than that due to $u\hat{h}_v(v)$, but both are mainly responsible for modeling local behaviors at the edges of the field. Finally, the smooth-by-smooth interaction term recovers the local patches observed in the raw data, that the other components would not be able to capture.

3   Background on P-splines

This section provides background information on P-splines, their tensor products and equivalent mixed model formulations. We refer the interested reader to Eilers and Marx (2010) and Eilers et al. (2015) for an extensive account of many aspects of P-splines.
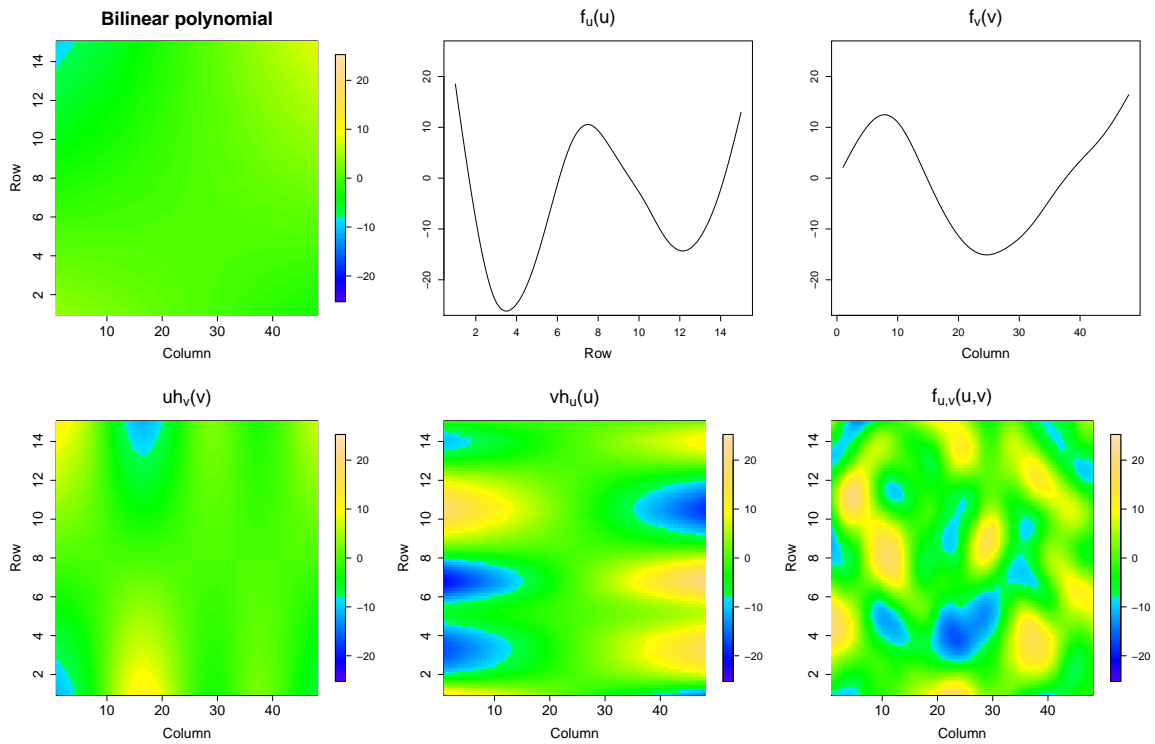
Figure 2: Bilinear and smooth components of the ANOVA-type decomposition of the estimated spatial trend for the barley uniformity data.

3.1 Modeling surfaces by means of P-splines

Suppose we are given n data triples (u_i, v_i, y_i), with u_i and v_i for positions, e.g, rows and columns, and y_i for a response variable, e.g a phenotype in a trial, and that we are interested in the following model

$$y_i = f(u_i, v_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2). \tag{3}$$

In the P-spline framework, the smooth bivariate function $f(u_i, v_i)$ is approximated by the tensor product of B-spline basis (Dierckx, 1993). The idea is simple: we form two B-spline bases, $\hat{\boldsymbol{B}}$, with $\hat{b}_{il} = \hat{B}_l(u_i)$ and $\breve{\boldsymbol{B}}$, with $\breve{b}_{ip} = \breve{B}_p(v_i)$, where $\hat{B}_l(u_i)$ is the l-th B-spline evaluated at u_i (and the same holds for $\breve{B}_p(v_i)$), and take

$$f(u_i, v_i) = \sum_{l=1}^{L} \sum_{p=1}^{P} \hat{B}_l(u_i) \breve{B}_p(v_i) \alpha_{lp},$$

where $\boldsymbol{\alpha} = (\alpha_{11}, \ldots, \alpha_{1P}, \ldots, \alpha_{LP})^t$ is a vector of unknown regression coefficients of dimension $(LP \times 1)$. Here $B_{lp}(u, v) = \hat{B}_l(u)\breve{B}_p(v)$ implicitly defines a bivariate B-spline basis function as the tensor product of two univariate B-splines. Under this representation, model (3) can be expressed in matrix notation as

$$\boldsymbol{y} = \boldsymbol{B}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{4}$$

with

$$\boldsymbol{B} = \breve{\boldsymbol{B}} \square \hat{\boldsymbol{B}} = \left(\breve{\boldsymbol{B}} \otimes \mathbf{1}_L^t \right) \odot \left(\mathbf{1}_P^t \otimes \hat{\boldsymbol{B}} \right),$$

where $\otimes$ denotes the Kronecker product and $\square$ the 'row-wise' Kronecker product (Eilers et al., 2006). Note that model (4) is purely parametric and can thus be estimated by minimizing the residual sum of squares (with the explicit solution $\widehat{\boldsymbol{\alpha}} = (\boldsymbol{B}^t\boldsymbol{B})^{-1}\boldsymbol{B}^t\boldsymbol{y}$.). To prevent overfitting, Eilers and Marx (1996) propose to modify the least squares criterion by incorporating a discrete penalty on the coefficient associated to adjacent B-splines. For the two-dimensional case, the vector $\boldsymbol{\alpha}$ can be seen as an $(L \times P)$ matrix of coefficients, $\boldsymbol{A} = [\alpha_{lp}]$. Now the rows and columns of $\boldsymbol{A}$ correspond to the regression coefficients in the v and u direction, respectively. In anisotropic P-splines, a different amount of smoothing is assumed along the u and v directions. It leads to two penalties: one on all rows of $\boldsymbol{A}$, the other on all of its columns; and the penalized least squares objective function becomes (Eilers and Marx, 2003)

$$S^* = ||\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}||^2 + \hat{\lambda}||\hat{\boldsymbol{D}}\boldsymbol{A}||_F^2 + \breve{\lambda}||\boldsymbol{A}\breve{\boldsymbol{D}}^t||_F^2 = ||\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\alpha}||^2 + \boldsymbol{\alpha}^t\boldsymbol{P}\boldsymbol{\alpha}, \tag{5}$$

where $\boldsymbol{P} = \hat{\lambda}(\boldsymbol{I}_P \otimes \hat{\boldsymbol{D}}^t\hat{\boldsymbol{D}}) + \breve{\lambda}(\breve{\boldsymbol{D}}^t\breve{\boldsymbol{D}} \otimes \boldsymbol{I}_L)$ is the penalty matrix, $\hat{\lambda}$ and $\breve{\lambda}$ are the smoothing parameters acting, respectively, on the columns and rows of $\boldsymbol{A}$, and $\hat{\boldsymbol{D}}$ and $\breve{\boldsymbol{D}}$ are matrices that form differences of order d_u and d_v respectively (in this paper we use second order differences, i.e., $d_u = d_v = 2$). Finally, $||\boldsymbol{M}||_F^2$ denotes the Frobenius norm, i.e., the sum of the squares of the elements of the matrix $\boldsymbol{M}$. The minimizer of (5), given $\hat{\lambda}$ and $\breve{\lambda}$, is

$$\widehat{\boldsymbol{\alpha}} = (\boldsymbol{B}^t\boldsymbol{B} + \boldsymbol{P})^{-1}\boldsymbol{B}^t\boldsymbol{y}. \tag{6}$$

The only tuning mechanism for smoothness is now the strength of the penalty, i.e the value of the smoothing parameters $\breve{\lambda}$ and $\hat{\lambda}$. The number of B-splines will be purposely chosen so large as to get over-fitting without a penalty. Accordingly, a critical issue is setting the right value for the smoothing parameters, which we like to see determined by the data. We discuss this point in detail in Section 3.2. The P-spline principle asks for a generous number of B-splines in the bases. (However, we cannot be too generous, as the size of the systems of equations is LP, the product of the sizes of $\hat{B}$ and $\breve{B}$). Appendix A contains a proposal to use smaller, nested, B-spline bases (Lee et al., 2013) that strongly reduce the computational effort.

3.2   Mixed model-based smoothing parameter selection

As said before, an important aspect when fitting a P-spline model is choosing appropriate values for the smoothing parameters. Eilers and Marx (1996) showed how to use several classical criteria, like cross-validation and Akaike's Information Criterion (AIC). For our application, however, we exploit the formal similarity between P-splines and mixed models (Currie and Durban, 2002; Wand, 2003), as it provides a general framework for the analysis of field trials. In this approach, the smooth functions are treated as sums of fixed and random components, and the smoothing parameters are replaced by ratios of variances which are estimated by REML. We present here the main ideas, and refer to Lee (2010); Lee and Durban (2011) for a more detailed description.

Note that, for given $\breve{\lambda}$ and $\hat{\lambda}$, the solution to the penalized least squares problem (5) given in (6) corresponds to the BLUPs for the $(LP \times 1)$ vector $\boldsymbol{\alpha}$ under the assumption that $\boldsymbol{\alpha} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{P}^+)$. Here, $\boldsymbol{P}^+$ denotes the Moore-Penrose pseudoinverse of the penalty matrix $\boldsymbol{P}$ given in (5). Given that both $\hat{\boldsymbol{D}}^t \hat{\boldsymbol{D}}$ and $\breve{\boldsymbol{D}}^t \breve{\boldsymbol{D}}$ are rank-deficient, $\boldsymbol{P}$ is also rank-deficient, and thus $N(\mathbf{0}, \sigma^2 \boldsymbol{P}^+)$ is a degenerate distribution. This causes numerical instability when applying mixed model estimation techniques. Moreover, it also implies that, for second order differences, the bilinear polynomial discussed in Section 2 remains unpenalized.

To obtain a full rank penalty matrix or precision matrix, the key is to write $\boldsymbol{B}\boldsymbol{\alpha} = \boldsymbol{X}_s \boldsymbol{\beta}_s + \boldsymbol{Z}_s \boldsymbol{c}_s$. There are now two bases, $\boldsymbol{X}_s$, with coefficients that are not penalized at all, and $\boldsymbol{Z}_s$, with a size penalty on its coefficients. There are different ways to decompose $\boldsymbol{B}$ (see, e.g., Ruppert et al., 2003; Eilers, 1999; Currie et al., 2006). In this paper we follow the proposal by Lee and Durban (2011), based on the eigenvalue decomposition (EVD) of $\hat{\boldsymbol{D}}^t \hat{\boldsymbol{D}}$ and $\breve{\boldsymbol{D}}^t \breve{\boldsymbol{D}}$. This approach gives rise to a diagonal penalty matrix, making it very appealing from a computational point of view.

Let $\hat{\boldsymbol{D}}^t \hat{\boldsymbol{D}} = \boldsymbol{U}_u \boldsymbol{E}_u \boldsymbol{U}_u^t$ and $\breve{\boldsymbol{D}}^t \breve{\boldsymbol{D}} = \boldsymbol{U}_v \boldsymbol{E}_v \boldsymbol{U}_v^t$ be the EVD of $\hat{\boldsymbol{D}}^t \hat{\boldsymbol{D}}$ and $\breve{\boldsymbol{D}}^t \breve{\boldsymbol{D}}$ respectively. Here $\boldsymbol{U}_j$ denotes the matrix of eigenvectors and $\boldsymbol{E}_j$ the diagonal matrix of eigenvalues ($j = u, v$). If we use second order differences, $\boldsymbol{E}_u$ and $\boldsymbol{E}_v$ contain $L-2$ and $P-2$ non-zero values respectively. Let us also denote by $\widetilde{\boldsymbol{U}}_j$ and $\widetilde{\boldsymbol{E}}_j$ the sub-matrices corresponding to the non-zero eigenvalues. Setting

$$\boldsymbol{X}_s = [\mathbf{1}_n, \boldsymbol{u}, \boldsymbol{v}, \boldsymbol{u} \odot \boldsymbol{v}] \quad \text{and} \quad \boldsymbol{Z}_s = [\boldsymbol{Z}_v, \boldsymbol{Z}_u, \boldsymbol{Z}_v \square \boldsymbol{u}, \boldsymbol{v} \square \boldsymbol{Z}_u, \boldsymbol{Z}_v \square \boldsymbol{Z}_u] , \tag{7}$$

where $\boldsymbol{Z}_u = \breve{\boldsymbol{B}} \widetilde{\boldsymbol{U}}_u$ and $\boldsymbol{Z}_v = \hat{\boldsymbol{B}} \widetilde{\boldsymbol{U}}_v$, the penalized least squares problem (5) becomes

$$S^* = ||\boldsymbol{y} - \boldsymbol{X}_s \boldsymbol{\beta}_s - \boldsymbol{Z}_s \boldsymbol{c}_s||^2 + \boldsymbol{c}_s^t \widetilde{\boldsymbol{P}} \boldsymbol{c}_s, \tag{8}$$

8

with

$$\widetilde{\boldsymbol{P}} = \text{blockdiag}\left(\check{\lambda}\widetilde{\boldsymbol{E}}_v, \hat{\lambda}\widetilde{\boldsymbol{E}}_u, \check{\lambda}\widetilde{\boldsymbol{E}}_v, \hat{\lambda}\widetilde{\boldsymbol{E}}_u, \check{\lambda}\widetilde{\boldsymbol{E}}_v \otimes \boldsymbol{I}_{L-2} + \hat{\lambda}\boldsymbol{I}_{P-2} \otimes \widetilde{\boldsymbol{E}}_u\right).$$

Each block in $\widetilde{\boldsymbol{P}}$ corresponds to each block in $\boldsymbol{Z}_s$ (see (7)). The solution to (8), for given $\check{\lambda}$ and $\hat{\lambda}$, corresponds to the empirical best linear unbiased estimator (BLUE) for the (4×1) vector $\boldsymbol{\beta}_s$, and the BLUPs for the $((LP - 4) \times 1)$ vector $\boldsymbol{c}_s$ under the linear mixed model

$$\boldsymbol{y} = \boldsymbol{X}_s\boldsymbol{\beta}_s + \boldsymbol{Z}_s\boldsymbol{c}_s + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n\right) \text{ and } \boldsymbol{c}_s \sim N\left(\boldsymbol{0}, \boldsymbol{G}_s\right),$$

where $\boldsymbol{G}_s = \sigma^2\widetilde{\boldsymbol{P}}^{-1}$. Denoting as $\check{\sigma}^2 = \sigma^2/\check{\lambda}$ and $\hat{\sigma}^2 = \sigma^2/\hat{\lambda}$ the variance parameters involved in $\boldsymbol{G}_s$, it follows

$$\boldsymbol{G}_s^{-1} = \text{blockdiag}\left(\frac{1}{\check{\sigma}^2}\widetilde{\boldsymbol{E}}_v, \frac{1}{\hat{\sigma}^2}\widetilde{\boldsymbol{E}}_u, \frac{1}{\check{\sigma}^2}\widetilde{\boldsymbol{E}}_v, \frac{1}{\hat{\sigma}^2}\widetilde{\boldsymbol{E}}_u, \frac{1}{\check{\sigma}^2}\widetilde{\boldsymbol{E}}_v \otimes \boldsymbol{I}_{L-2} + \frac{1}{\hat{\sigma}^2}\boldsymbol{I}_{P-2} \otimes \widetilde{\boldsymbol{E}}_u\right). \quad (9)$$

Besides the possibility of selecting the smoothing parameters ($\check{\lambda} = \sigma^2/\check{\sigma}^2$ and $\hat{\lambda} = \sigma^2/\hat{\sigma}^2$) by (RE)ML, the mixed model representation of the tensor-product P-spline directly provides the interesting ANOVA-type decomposition discussed in Section 2 (Lee and Durban, 2011; Lee et al., 2013). The block-structure of both $\boldsymbol{X}_s$ and $\boldsymbol{Z}_s$ (see (7)) implies

$$\begin{aligned}
f(\boldsymbol{u}, \boldsymbol{v}) &= \boldsymbol{X}_s\boldsymbol{\beta}_s + \boldsymbol{Z}_s\boldsymbol{c}_s \\
&= \boldsymbol{1}_n\beta_0 + \boldsymbol{u}\beta_1 + \boldsymbol{v}\beta_2 + \boldsymbol{u} \odot \boldsymbol{v}\beta_3 \\
&\quad + \underbrace{f_v(\boldsymbol{v})}_{\boldsymbol{Z}_v\boldsymbol{c}_{s1}} + \underbrace{f_u(\boldsymbol{u})}_{\boldsymbol{Z}_u\boldsymbol{c}_{s2}} + \underbrace{\boldsymbol{u} \odot h_v(\boldsymbol{v})}_{[\boldsymbol{Z}_v\square\boldsymbol{u}]\boldsymbol{c}_{s3}} + \underbrace{\boldsymbol{v} \odot h_u(\boldsymbol{u})}_{[\boldsymbol{v}\square\boldsymbol{Z}_u]\boldsymbol{c}_{s4}} + \underbrace{f_{u,v}(\boldsymbol{u}, \boldsymbol{v})}_{[\boldsymbol{Z}_v\square\boldsymbol{Z}_u]\boldsymbol{c}_{s5}},
\end{aligned}$$

where $\boldsymbol{c}_{sk}$ $(k = 1, \ldots, 5)$ contains the elements of $\boldsymbol{c}_s$ that correspond to the k-th block of $\boldsymbol{Z}_s$, i.e., $\boldsymbol{c}_s = \left(\boldsymbol{c}_{s1}^t, \ldots, \boldsymbol{c}_{s5}^t\right)^t$. The dimension of each $\boldsymbol{c}_{sk}$ will depend on the basis dimensions, L and P, used for the tensor product (eqn. (4)). It is easy to show that $\boldsymbol{c}_{s1}$ and $\boldsymbol{c}_{s3}$ are $((P - 2) \times 1)$ vectors, $\boldsymbol{c}_{s2}$ and $\boldsymbol{c}_{s4}$ are $((L - 2) \times 1)$ vectors, and $\boldsymbol{c}_{s5}$ is a $((L - 2)(P - 2) \times 1)$ vector. The nested bases presented in Appendix A allow reducing the dimension associated to the smooth-by-smooth interaction term, and, as a consequence, the computational effort.

A close look at (9) shows that, despite the five smooth components, only two variance parameters (or smoothing parameters) control their smoothness: $\check{\sigma}^2$ and $\hat{\sigma}^2$. In fact, the same variance parameter apply to both the main effects and the interaction terms. In Lee et al. (2013) the ANOVA-type decomposition is further exploited, and the authors propose to use a different variance component for each smooth component, i.e., each block in (9) will have its own variance component. For ease of notation, let $\boldsymbol{\Lambda}_{s1}^{-1} = \boldsymbol{\Lambda}_{s3}^{-1} = \widetilde{\boldsymbol{E}}_v$, $\boldsymbol{\Lambda}_{s2}^{-1} = \boldsymbol{\Lambda}_{s4}^{-1} = \widetilde{\boldsymbol{E}}_u$, and $\boldsymbol{\Lambda}_{s5}^{-1} = \widetilde{\boldsymbol{E}}_v \otimes \boldsymbol{I}_{L-2} + \boldsymbol{I}_{P-2} \otimes \widetilde{\boldsymbol{E}}_u$. For the PS-ANOVA model the precision matrix is then defined as

$$\boldsymbol{G}_s^{-1} = \text{blockdiag}\left(\frac{1}{\sigma_{s1}^2}\boldsymbol{\Lambda}_{s1}^{-1}, \frac{1}{\sigma_{s2}^2}\boldsymbol{\Lambda}_{s2}^{-1}, \frac{1}{\sigma_{s3}^2}\boldsymbol{\Lambda}_{s3}^{-1}, \frac{1}{\sigma_{s4}^2}\boldsymbol{\Lambda}_{s4}^{-1}, \frac{1}{\sigma_{s5}^2}\boldsymbol{\Lambda}_{s5}^{-1}\right),$$

and thus the variance-covariance matrix $\boldsymbol{G}_s$ is a linear function of variance parameters

$$\boldsymbol{G}_s = \bigoplus_{k=1}^{5} \sigma_{sk}^2\boldsymbol{\Lambda}_{sk} = \bigoplus_{k=1}^{5} \boldsymbol{G}_{sk} = \text{blockdiag}\left(\boldsymbol{G}_{s1}, \boldsymbol{G}_{s2}, \boldsymbol{G}_{s3}, \boldsymbol{G}_{s4}, \boldsymbol{G}_{s5}\right), \quad (10)$$

where $\boldsymbol{G}_{sk} = \sigma_{sk}^2 \boldsymbol{\Lambda}_{sk}$ $(k = 1, \ldots, 5)$. In other words, here the tensor product P-spline mixed model is represented as the sum of 5 sets of mutually independent Gaussian random factors $\boldsymbol{c}_{sk}$ each depending on one variance σ_{sk}^2 $(k = 1, \ldots, 5)$.

4 Spatial models for field trials

The tensor product P-spline presented in Section 3 constitutes the base for the analysis of agricultural field trials. As said, it allows the modeling of the random spatial variation typically presented in a field, providing an explicit estimate of the spatial random field. However, on top of this spatial field, we need to build up more complex models in order to account for the genetic variation, the presence of block and/or replication effects, or other sources of spatial variation as those due to the way the field was prepared. From now on, we therefore consider the following linear mixed model

$$\boldsymbol{y} = \underbrace{\boldsymbol{X}_s \boldsymbol{\beta}_s + \boldsymbol{Z}_s \boldsymbol{c}_s}_{f(\boldsymbol{u}, \boldsymbol{v})} + \boldsymbol{X}_d \boldsymbol{\beta}_d + \boldsymbol{Z}_d \boldsymbol{c}_d + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{c}_s \sim N\left(\boldsymbol{0}, \boldsymbol{G}_s\right) \text{ and } \boldsymbol{c}_d \sim N\left(\boldsymbol{0}, \boldsymbol{G}_d\right), \quad (11)$$

where $\boldsymbol{X}_s$, $\boldsymbol{Z}_s$ and $\boldsymbol{G}_s$ have been defined in (7) and (10). $\boldsymbol{X}_d$ and $\boldsymbol{Z}_d$ represent column-partitioned matrices, associated respectively with extra fixed and random components, as for instance, row, column, replicate and/or genotypic effects. We assume that $\boldsymbol{X}_d$ has full rank, $\boldsymbol{Z}_d = [\boldsymbol{Z}_{d1}, \ldots, \boldsymbol{Z}_{db}]$, and $\boldsymbol{c}_d = \left(\boldsymbol{c}_{d1}^t, \ldots, \boldsymbol{c}_{db}^t\right)^t$. Each $\boldsymbol{Z}_{dk}$ corresponds to the design matrix of the k-th random factor $\boldsymbol{c}_{dk}$, with $\boldsymbol{c}_{dk}$ being a $(m_{dk} \times 1)$ vector $(k = 1, \ldots, b)$. We assume further that $\boldsymbol{c}_s$ and $\boldsymbol{c}_d$ are independent, and that the b components of $\boldsymbol{c}_d$ are mutually independent with diagonal variance-covariance matrices $\sigma_{dk}^2 \boldsymbol{\Lambda}_{dk}$, i.e., $\boldsymbol{G}_d = \bigoplus_{k=1}^b \boldsymbol{G}_{dk} = \bigoplus_{k=1}^b \sigma_{dk}^2 \boldsymbol{\Lambda}_{dk}$. In spite of this restriction, the estimation of rather complex mixed models can be accommodated, as it will be shown in Section 6. In order to keep the notation as simple as possible, we rewrite model (11) in a more compact way as follows

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{Z} \boldsymbol{c} + \boldsymbol{\varepsilon}, \text{ with } \boldsymbol{c} \sim N\left(\boldsymbol{0}, \boldsymbol{G}\right) \text{ and } \boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n\right), \quad (12)$$

where $\boldsymbol{X} = [\boldsymbol{X}_s, \boldsymbol{X}_d]$, $\boldsymbol{Z} = [\boldsymbol{Z}_s, \boldsymbol{Z}_d] = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_q]$ $(q = 5 + b)$, and

$$\boldsymbol{G} = \text{blockdiag}\left(\boldsymbol{G}_s, \boldsymbol{G}_d\right) = \bigoplus_{k=1}^q \boldsymbol{G}_k = \bigoplus_{k=1}^q \sigma_k^2 \boldsymbol{\Lambda}_k. \quad (13)$$

As far as estimation of model (12) is concerned, estimates of the fixed and random effect coefficients, for given values of the variance components, follow from standard mixed-model theory (see Appendix B), and variance components can be obtained, as usual, by maximizing the REML log-likelihood function

$$l = -\frac{1}{2} \log |\boldsymbol{V}| - \frac{1}{2} \log |\boldsymbol{X}^t \boldsymbol{V}^{-1} \boldsymbol{X}| - \frac{1}{2}(\boldsymbol{y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}})^t \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}). \quad (14)$$

Given that $\boldsymbol{G}$ is a linear function of variance components, estimation can be accommodated using standard mixed model procedures, as, e.g., those implemented in the R-packages asreml-R,

nlme and lme4, or the PROC MIXED procedure in SAS®. In next Section we present the numerical procedure implemented in the R-package SpATS that accompany this paper. The procedure presents many appealing features, which make it a good candidate for the analysis of field trials: (a) it is fast and stable; (b) it is robust (i.e., it converges from almost any starting values); and (c) it always provides positive estimates of the variance components, although it is possible to obtain values very close to zero. For all these reasons it has been our choice.

4.1 Variance component estimation

As said, REML estimates of the variance components are obtained by maximizing (14). Taking derivatives with respect to the variance components σ_k^2 ($k = 1, \ldots, q$), we obtain (see e.g., Rodríguez-Álvarez et al., 2015; Johnson and Thompson, 1995)

$$\frac{\partial l}{\partial \sigma_k^2} = -\frac{1}{2} trace\left(\boldsymbol{Z}^t \boldsymbol{Q} \boldsymbol{Z} \boldsymbol{G} \frac{\partial \boldsymbol{G}^{-1}}{\partial \sigma_k^2} \boldsymbol{G} \right) + \frac{1}{2} \widehat{\boldsymbol{c}}^t \frac{\partial \boldsymbol{G}^{-1}}{\partial \sigma_k^2} \widehat{\boldsymbol{c}},$$

where $\boldsymbol{Q} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1} \boldsymbol{X} \left(\boldsymbol{X}^t \boldsymbol{V}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^t \boldsymbol{V}^{-1}$ with $\boldsymbol{V} = \boldsymbol{R} + \boldsymbol{Z} \boldsymbol{G} \boldsymbol{Z}^t$ and $\boldsymbol{R} = \sigma^2 \boldsymbol{I}_n$. By (13), it is easy to show that the former derivatives can be expressed as

$$2\frac{\partial l}{\partial \sigma_k^2} = -\frac{1}{\sigma_k^2} trace\left(\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k \boldsymbol{G}_k \right) + \frac{1}{\sigma_k^4} \widehat{\boldsymbol{c}}_k^t \boldsymbol{\Lambda}_k^{-1} \widehat{\boldsymbol{c}}_k.$$

Then, REML estimates of the variance components are found by equating the former expression to zero, which gives

$$\widehat{\sigma}_k^2 = \frac{\widehat{\boldsymbol{c}}_k^t \boldsymbol{\Lambda}_k^{-1} \widehat{\boldsymbol{c}}_k}{\mathrm{ED}_k}, k = 1, \ldots, q, \tag{15}$$

with

$$\mathrm{ED}_k = \mathrm{trace}\left(\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k \boldsymbol{G}_k \right). \tag{16}$$

An estimate of σ^2 can also be easily obtained following the same reasoning. Specifically, in this case we have

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}^t \widehat{\boldsymbol{\varepsilon}}}{\mathrm{ED}_{\boldsymbol{\varepsilon}}}, \tag{17}$$

where $\widehat{\boldsymbol{\varepsilon}} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\widehat{\boldsymbol{c}}$ and

$$\mathrm{ED}_{\boldsymbol{\varepsilon}} = \mathrm{trace}\left(\boldsymbol{R} \boldsymbol{Q} \right) = n - \mathrm{rank}(\boldsymbol{X}) - \sum_{k=1}^{q} \mathrm{ED}_k. \tag{18}$$

As can be seen, the right-hand side of eqn. (16) depends on the unknown variance components. Hence, eqns. (15) and (17) need to be solved with an iterative procedure. Given some starting values for the variance components, estimation of model (12) is thus obtained by iterating, until convergence, among (a) estimating the fixed and random effect coefficients (linear system (24)); (b) evaluating the right-hand side of eqn. (16); and (c) updating the variances by means of eqns. (15) and (17). In this work, the REML-deviance was used as the convergence criterion.

To the best of our knowledge, this iterative algorithm was originally proposed by Henderson in an unpublished manuscript, and discussed in detail by Harville (1977) and Engel (1990), among others. Schall (1991) further extended the algorithm for the estimation of generalized linear mixed models. In the P-spline literature, the algorithm has been also successfully used (e.g., Schnabel and Eilers, 2009; Lee et al., 2013), and usually referred to as Schall's algorithm.

From a computational point of view, the traces in (16) may involve the computation and manipulations of several large matrices. However, there are several ways this computation can be relaxed. For instance, using some results on mixed models (see, e.g., Johnson and Thompson, 1995) we have that

$$\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k \boldsymbol{G}_k = \boldsymbol{I}_{m_k} - \boldsymbol{G}_k^{-1} \boldsymbol{C}_{kk}^{-1} = \boldsymbol{I}_{m_k} - \frac{1}{\sigma_k^2} \boldsymbol{\Lambda}_k^{-1} \boldsymbol{C}_{kk}^{-1}, \tag{19}$$

where m_k is the number of random coefficients in the vector $\boldsymbol{c}_k$ and $\boldsymbol{C}_{kk}^{-1}$ is that partition of the inverse of $\boldsymbol{C}$ in (24) corresponding to $\boldsymbol{c}_k$. An alternative approach would be to use the result given in eqn. (5.3) in Harville (1977)

$$\boldsymbol{Z}^t \boldsymbol{Q} \boldsymbol{Z} = \boldsymbol{G}^{-1} \boldsymbol{C}_m^{-1} \left[\boldsymbol{X}, \boldsymbol{Z} \right]^t \boldsymbol{R}^{-1} \boldsymbol{Z},$$

where $\boldsymbol{C}_m^{-1}$ denotes the matrix formed by the last m rows of the inverse of $\boldsymbol{C}$ (with $m = \sum_{k=1}^{q} m_k$). Here, the block-diagonal elements of $\boldsymbol{Z}^t \boldsymbol{Q} \boldsymbol{Z}$ correspond to $\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k$. Despite the apparent complexity of this expression, in order to compute (16) only the diagonal of $\boldsymbol{Z}^t \boldsymbol{Q} \boldsymbol{Z}$ needs to be explicitly obtained, since $\boldsymbol{G}$ is a diagonal matrix (see Rodríguez-Álvarez et al., 2015, for further details).

4.2   Effective dimensions

Denoting the denominator of (15) and (17) as $\mathrm{ED}_{\{\cdot\}}$ (from "effective dimension") has not been done without purpose. In the smoothing context, the notion of effective dimension or effective degrees of freedom is well known (see, e.g., Hastie and Tibshirani, 1990). The effective dimension (denoted by ED) of a "smooth" model is defined as the trace of the so-called "hat" matrix $\boldsymbol{H}$, defined as $\widehat{\boldsymbol{y}} = \boldsymbol{H} \boldsymbol{y}$. In this setting, ED can be interpreted as a measure of the complexity of the model: the larger the ED, the more complex (or the less smooth) the model (see also Ye, 1998; Eilers et al., 2015).

In recent years, several new definitions and generalizations of the concept of effective dimension that are applicable to (generalized) linear mixed models have been proposed in the statistical literature (see, e.g., You et al., 2016, and references therein). In almost all cases, the aim has been to provide a complexity measure that allows models' comparison and selection (via, for instance, the AIC). For our application, however, we are more interested in obtaining a separate complexity measure for each component in model (12), that can give us insights about the contribution of that effect when explaining the response (phenotypic) variation. This issue has been already discussed by Cui et al. (2010). In that paper the authors define the effective dimension of a model's component as the trace of the ratio of the component modeled variance matrix to the total variance matrix. This new definition can be interpreted

as the fraction of response variation attributed to individual components. Besides, it allows explaining how components compete with one another to explain that variation.

In line with previous work in the smoothing context (e.g., Hastie and Tibshirani, 1990; Ruppert et al., 2003), this paper considers defining the effective dimension of a model's component as the trace of the corresponding hat matrix. Appendix C shows the equivalence between this definition and that by Cui et al. (2010). Note first that for the linear mixed model (12) we might define two different hat matrices: one for the fixed part of the model and one for the random part. From standard mixed-model theory (see Appendix B for further details) it follows that

$$\boldsymbol{H}\boldsymbol{y} = \widehat{\boldsymbol{y}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\widehat{\boldsymbol{c}} = \boldsymbol{H}_F\boldsymbol{y} + \boldsymbol{H}_R\boldsymbol{y},$$

where $\boldsymbol{H}_F = \boldsymbol{X}\left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{V}^{-1}$ and $\boldsymbol{H}_R = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}^t\boldsymbol{Q}$. Thus,

$$\text{ED} = \text{trace}\left(\boldsymbol{H}\right) = \text{trace}\left(\boldsymbol{H}_F\right) + \text{trace}\left(\boldsymbol{H}_R\right).$$

However, we can even go one step further. The block structure of both the random design matrix $\boldsymbol{Z}$ and the variance-covariance matrix $\boldsymbol{G}$ in (12) implies that (see Appendix B)

$$\boldsymbol{H}_R = \sum_{k=1}^{q} \boldsymbol{H}_k$$

where $\boldsymbol{H}_k = \boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t\boldsymbol{Q}$, with $\boldsymbol{Z}_k$ and $\boldsymbol{G}_k$ denoting the k-th block of $\boldsymbol{Z}$ and $\boldsymbol{G}$ respectively. As shown in Appendix B, each of these hat matrices corresponds to an individual random component in model (12), either coming from the PS-ANOVA spatial field or a "pure" random factor, i.e.,

$$\widehat{f}_v(\boldsymbol{v}) = \boldsymbol{Z}_1\widehat{\boldsymbol{c}}_1 = \boldsymbol{H}_1\boldsymbol{y} \qquad \widehat{f}_u(\boldsymbol{u}) = \boldsymbol{Z}_2\widehat{\boldsymbol{c}}_2 = \boldsymbol{H}_2\boldsymbol{y} \qquad \boldsymbol{u}\odot\widehat{h}_v(\boldsymbol{v}) = \boldsymbol{Z}_3\widehat{\boldsymbol{c}}_3 = \boldsymbol{H}_3\boldsymbol{y},$$

$$\boldsymbol{v}\odot\widehat{h}_u(\boldsymbol{u}) = \boldsymbol{Z}_4\widehat{\boldsymbol{c}}_4 = \boldsymbol{H}_4\boldsymbol{y} \qquad \widehat{f}_{u,v}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{Z}_5\widehat{\boldsymbol{c}}_5 = \boldsymbol{H}_5\boldsymbol{y}.$$

$$\boldsymbol{Z}_k\widehat{\boldsymbol{c}}_k = \boldsymbol{H}_k\boldsymbol{y}, k = 6, \ldots, q.$$

This results suggests defining the effective dimension for $\boldsymbol{c}_k$ as the trace of $\boldsymbol{H}_k$. Using trace properties, we have

$$\text{trace}\left(\boldsymbol{H}_k\right) = \text{trace}\left(\boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t\boldsymbol{Q}\right) = \text{trace}\left(\boldsymbol{Z}_k^t\boldsymbol{Q}\boldsymbol{Z}_k\boldsymbol{G}_k\right) = \text{ED}_k,$$

and the total effective dimension of model (12) is thus decomposed as the sum of independent contributions

$$\begin{aligned}
\text{ED} &= \text{trace}\left(\boldsymbol{H}\right) \\
&= \text{trace}\left(\boldsymbol{H}_F\right) + \text{trace}\left(\boldsymbol{H}_R\right) \\
&= \text{trace}\left(\boldsymbol{H}_F\right) + \sum_{k=1}^{q} \text{trace}\left(\boldsymbol{H}_k\right) \\
&= \text{rank}(\boldsymbol{X}) + \sum_{k=1}^{q} \text{ED}_k.
\end{aligned}$$

Besides, eqn. (27) in Appendix B also suggests defining the effective dimension for the residuals as the trace of $\boldsymbol{H}_{\boldsymbol{\varepsilon}}$ and thus (see eqn. (18))

$$\text{trace}(\boldsymbol{H}_{\varepsilon}) = \text{trace}(\boldsymbol{RQ}) = \text{ED}_{\varepsilon}.$$

This is in concordance with the definition given by Cui et al. (2010) (see Appendix C). As extensively discussed by the authors, $\text{ED}_{\boldsymbol{\varepsilon}}$ avoids the problem that is posed by most of the traditional definitions for the residual effective dimension in the smoothing context, that is to say, that $\text{ED} + \text{ED}_{\boldsymbol{\varepsilon}}$ is not equal to the number of observations n.

For all components ED_k $(k = 1, \ldots, q)$ will vary between 0 and $(m_k - \zeta_k)$, where ζ_k is the number of zero eigenvalues of $\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k \boldsymbol{G}_k$. In Cui et al. (2010) the authors show that this upper bound, $(m_k - \zeta_k)$, can also be expressed as: $\text{rank}([\boldsymbol{X}, \boldsymbol{Z}_k]) - \text{rank}(\boldsymbol{X})$. The signal-to-noise ratio σ_k^2/σ^2 modulates the value of ED_k: when $\sigma_k^2/\sigma^2 \to 0$ then $\text{ED}_k \to 0$; and when $\sigma_k^2/\sigma^2 \to \infty$, then $\text{ED}_k \to (m_k - \zeta_k)$. Arguably, ED_k can be therefore be interpreted as a measure of the complexity of the corresponding component. A value of zero would indicate that this component does not contribute to the response variability.

A nice property of the PS-ANOVA spatial field is that we have a separate effective dimension for each of the five smooth components, which in turns gives a separate measure of the contribution of that component. In this case, ED_k $(k = 1, \ldots, 5)$ can be interpreted as a measure of the smoothness of the corresponding term: the larger the effective dimension, the less smooth the effect. For the other random factors, ED_k $(k = 6, \ldots, q)$ will also give a measure of the complexity, but here, it is worth interpreting it as a measure of the shrinkage. What is more, when the genetic effect is included in model (12) as random, the associated effective dimension corresponds to the generalized heritability proposed by Oakey et al. (2006). A formal derivation and deeper discussion is provided in Section 4.3.

To finish this part we would like to emphasize that, at convergence, the algorithm presented in Section 4.1 explicitly provides an estimate of the effective dimension associated to each random component in model (12). Moreover, as can be seen in eqn. (15), a beautiful result is that the algorithm furnishes variance components estimates that are the ratio of the sum of squares of the BLUPs of the components of $\boldsymbol{c}_k$ (weighted according to their precision, $\boldsymbol{\Lambda}_k^{-1}$) to the individual effective dimension ED_k (and the same applies for the residual variance).

4.2.1 Motivating example revisited

Let us now come back to the uniformity barley data discussed in Section 2. For fitting model (2), we used cubic B-spline bases of dimension $L = 18$ and $P = 51$, jointly with nested basis for the columns with $P_N = 27$. Table 1 shows the model (i.e., the number of coefficients), and effective dimensions associated to the row and column random factors, the smooth spatial field $f(u, v)$, and each of the PS-ANOVA components. If we focus on the random effects for the rows and columns, we have that the estimated effective dimensions are 5.5 (out of 14) and 38.0 (out of 47) respectively. This result suggests that the column effect is stronger than the row effect, and that these two components are probably needed. For the PS-ANOVA spatial field (excluding the intercept), the total effective dimension is 78.7, with the smooth-by-smooth interaction trend being responsible for the strongest contribution, with an effective dimension of 53.1. As can be observed on the graphical results depicted in Section 2, the smooth trend

14

Table 1: Model and effective dimension of the smooth spatial component, and the ANOVA-type decomposition components for the barley uniformity trial.

Dimensions	Random		Spatial components					
			Smooth					
	$\boldsymbol{c}_r$	$\boldsymbol{c}_c$	Global - $f(u,v)$	$f_u(u)$	$f_v(v)$	$vh_u(u)$	$uh_v(v)$	$f_{u,v}(u,v)$
Model (m_k)	48	15	533	16	49	16	49	400
Effective (ED_k)	38.0	5.5	78.7	6.2	3.7	8.2	4.5	53.1

along the rows, $f_u(u)$, is more complex (or rougher) than the one along the columns, $f_v(v)$, and this fact is made evident on the effective dimension related to each of these components, 6.2 and 3.7 respectively. Besides, as could have also been expected, the effective dimension associated to $vh_u(u)$ is also larger than that associated $uh_v(v)$. These results suggest and emphasize the need of modeling spatial trends by means of bivariate surfaces. Here the additive assumption – only based on main smooth effects – would have not been flexible enough to recover the spatial trend variation present in the data.

4.3 Heritability and effective dimension

To introduce the standard definition of heritability, let us start with the classical quantitative genetic model, in which m_g genotypes, each replicated r times, are evaluated and no other model components (either spatial, fixed or random) are considered

$$\boldsymbol{y} = \boldsymbol{1}_n\beta_0 + \boldsymbol{Z}_g\boldsymbol{c}_g + \boldsymbol{\varepsilon}.$$

We assume that the observations are ordered according to the genotypes, i.e., $\boldsymbol{Z}_g = \boldsymbol{I}_{m_g} \otimes \boldsymbol{1}_r$, and that $\boldsymbol{c}_g \sim N(\boldsymbol{0}, \sigma_g^2 \boldsymbol{I}_{m_g})$ and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. Under this model, the standard heritability measure is defined as

$$H_s^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma^2/r)}. \tag{20}$$

That is to say, the standard heritability is the proportion of the total (phenotypic) variation explained by the genetic component.

We turn now to the notion of effective dimension defined above. For the genetic effects $\boldsymbol{c}_g$, the associated effective dimension is $\mathrm{ED}_g = \mathrm{trace}\left(\boldsymbol{Z}_g^t\boldsymbol{Q}\boldsymbol{Z}_g\boldsymbol{G}_g\right)$, where $\boldsymbol{G}_g = \sigma_g^2\boldsymbol{I}_{m_g}$. By eqn. (3.7) in Harville (1977), we have

$$\boldsymbol{Z}_g^t\boldsymbol{Q}\boldsymbol{Z}_g\boldsymbol{G}_g = \left(\frac{1}{\lambda}\boldsymbol{I}_{m_g} + \boldsymbol{Z}_g^T\boldsymbol{S}\boldsymbol{Z}_g\right)^{-1}\boldsymbol{Z}_g^T\boldsymbol{S}\boldsymbol{Z}_g,$$

where $\lambda = \frac{\sigma_g^2}{\sigma^2}$ and $\boldsymbol{S} = \boldsymbol{I}_n - \boldsymbol{1}_n(\boldsymbol{1}_n^t\boldsymbol{1}_n)^{-1}\boldsymbol{1}_n^t$. It can be shown that $\boldsymbol{Z}_g^T\boldsymbol{S}\boldsymbol{Z}_g$ has $(m_g - 1)$ eigenvalues equal to r (the number of replicates), and 1 eigenvalue equal to zero. Moreover,

$\frac{1}{\lambda} \boldsymbol{I}_{m_g}$ has m_g eigenvalues all equal to $\frac{1}{\lambda}$. Using the property that the trace of a matrix is the sum of its eigenvalues, we have

$$\text{ED}_g = \sum_{i=1}^{m_g} \lambda_i = \sum_{i=1}^{(m_g-1)} \left(\frac{1}{\lambda} + r\right)^{-1} r = (m_g - 1)\frac{\sigma_g^2}{(\sigma_g^2 + \sigma^2/r)},$$

where λ_i are the eigenvalues of $\boldsymbol{Z}_g^t \boldsymbol{Q} \boldsymbol{Z}_g \boldsymbol{G}_g$. Hence, there is direct link between the standard heritability measure H_s^2 defined in (20) and the genetic effective dimension, with

$$H_s^2 = \frac{\text{ED}_g}{m_g - 1}.$$

When the statistical analysis of a field trial experiment involves modeling more sources of variation (as, e.g., spatial and/or extraneous variation), the standard definition of heritability given in (20) does not longer apply, and several generalizations have been proposed in the literature (e.g., Cullis et al., 2006; Oakey et al., 2006). For instance, Cullis et al. (2006) present a generalized definition of heritability – applicable whenever $\boldsymbol{G}_g = \sigma_g^2 \boldsymbol{I}_{m_g}$ – based on the pairwise prediction error variance of genetic effects

$$H_c^2 = 1 - \frac{1}{\sigma_g^2} \sum_{i=1}^{m_g} \frac{pev(c_{gi})}{m_g}.$$

where $pev(c_{gi}) = var(\widehat{c}_{gi} - c_{gi})$. Using the equivalence given in (19) and noting that $\boldsymbol{C}_{kk}^{-1}$ corresponds to $\mathbb{C}\text{ov}(\widehat{\boldsymbol{c}}_g - \boldsymbol{c}_g)$ (i.e., the prediction error variance-covariance matrix for the genetic effects), the generalized heritability measure proposed by Cullis et al. (2006) can also be expressed in terms of the genetic effective dimension

$$H_c^2 = 1 - \frac{1}{\sigma_g^2} \sum_{i=1}^{m_g} \frac{pev(c_{gi})}{m_g} = \frac{\text{trace}(\boldsymbol{I}_{m_g} - \frac{1}{\sigma_g^2}\boldsymbol{C}_{kk}^{-1})}{m_g} = \frac{\text{trace}(\boldsymbol{Z}_g^t \boldsymbol{Q} \boldsymbol{Z}_g \boldsymbol{G}_g)}{m_g} = \frac{\text{ED}_g}{m_g}.$$

In Oakey et al. (2006) a more general definition of heritability is presented that can be used regardless of the structure of genetic variance-covariance matrix $\boldsymbol{G}_g$. As can be seen in eqn. (7) of that paper, the definition the authors propose turns out to be the ratio between the genetic effective dimension and the number of genetic effects m_g minus the number of zero eigenvalues of $\boldsymbol{Z}_g^t \boldsymbol{Q} \boldsymbol{Z}_g \boldsymbol{G}_g$, i.e.,

$$H_g^2 = \frac{\text{ED}_g}{m_g - \zeta_g}. \tag{21}$$

As said before, the denominator of (21) represents the upper bound of the genetic effective dimension, and $1 - H_g^2$ can therefore be interpreted as a shrinkage factor. The generalized heritability proposed by Cullis et al. (2006) would be a special case, but ignoring the number of zero eigenvalues.

On the basis of the estimation procedure presented above, an estimate of the generalized heritability can be thus obtained as

$$\widehat{H}_g^2 = \frac{\widehat{\text{ED}}_g}{m_g - \zeta_g}.$$

where $\widehat{\text{ED}}_g$ denotes the estimated effective.

5 Simulation studies

This section is devoted to present the results of several studies performed to evaluate the behavior of our SpATS model under controlled scenarios, and its comparison with the separable autoregressive (AR×AR) model proposed by Gilmour et al. (1997). In the context of single-trial experiments, this proposal has become the standard modeling strategy, specially among applied breeders, and therefore it has been chosen as the benchmark model. In Section 5.1, the uniformity trial presented in Section 2 is used to introduce the proposal by Gilmour et al. (1997), and comparisons between both approaches when including genotypic effects are reported. Section 5.2 presents the results of a simulation study when the underlying simulated model follows an AR×AR process. In both cases, simulations were done using the R-packages SpATS and ASreml-R.

5.1 Barley uniformity trial

As said, the uniformity trial presented in Section 2 was analyzed using the AR×AR model proposed by Gilmour et al. (1997). Model selection was performed by means of the sample variogram and plots of the residuals as suggested by Stefanova et al. (2009). Starting with the simplest model, including only the separable Gaussian AR process of order 1, we further evaluated the need of extra model components. In total, 9 different models were considered, with the best being

$$\boldsymbol{y} = \mathbf{1}_n\beta_0 + \boldsymbol{u}\beta_1 + \boldsymbol{v}\beta_2 + f_u(\boldsymbol{u}) + f_v(\boldsymbol{v}) + \boldsymbol{Z}_c\boldsymbol{c}_c + \boldsymbol{\xi} + \boldsymbol{\varepsilon}. \tag{22}$$

Here $\boldsymbol{\xi}$ is a (720×1) spatially dependent random vector, for which a separable Gaussian AR process of order 1 in the row and column directions is assumed. Accordingly, $cov(\xi_l, \xi_p) = \sigma_s^2 \rho_r^{|u_l - u_p|} \rho_c^{|v_l - v_p|}$, where ρ_r and ρ_c are the autocorrelation parameters for row and column, respectively. As in our approach, $f_u(\cdot)$ and $f_v(\cdot)$ represent smooth-trend functions over the row and column direction respectively. We note that in the geostatistics literature, $\boldsymbol{\varepsilon}$ is usually referred to as measurement error or nugget effect.

As postulated by Gilmour et al. (1997), model (22) accounts for three sources of spatial variation: the global trend variation, the local trend variation; and the so-called extraneous variation. Here, (a) the global trend variation is modeled by the linear effect along the rows (β_1) and columns (β_2) as well as by the smooth-effect functions $f_u(\cdot)$ and $f_v(\cdot)$; and (b) the local trend variation by means of the spatially dependent random error $\boldsymbol{\xi}$. The extraneous variation, related to the experimental procedure, is accounted for by the column random factor $\boldsymbol{c}_c$. Under this framework, we can see our SpATS model (see eqn. (2)) as that based on aggregating both the local and global trend variation in one component, and modeling it by means of a smooth bivariate surface. Table 2 shows the REML estimates of the variance components based on model (2) and model (22). Based on this table, we find it difficult to compare both approaches. Thus, to gain more insights in the performance of these two models, we designed a simulation study in which, on top of the uniformity data, genotypic effects were included, i.e,

$$\boldsymbol{y}^* = \boldsymbol{y} + \boldsymbol{Z}_g\boldsymbol{c}_g,$$

Table 2: REML estimates of the variance parameters for the barley uniformity trial based on both AR×AR and SpATS approaches.

Model	Parameter					
	σ_s^2	σ_r^2	σ_c^2	ρ_r	ρ_c	σ^2
AR×AR	265.67	-	144.20	0.383	0.834	173.12
SpATS	-	20.38	145.14	-	-	238.94

where $\boldsymbol{c}_g$ denotes the genotypic effects, with $\boldsymbol{c}_g \sim N(\boldsymbol{0}, \sigma_g^2 \boldsymbol{I}_{m_g})$. For the results reported here, we considered $\sigma_g^2 = 144$, and a total of $m_g = 360$ genotypes, each replicated twice. The 360 genotypes were allocated to the plots following an alpha design, in blocks of size 15 (the number of rows in the field).

For each data set simulated as described above, we fitted our SpATS model (see eqn. (2)), including the genetic random factor. For Gilmour et al. (1997)'s approach we considered model (22) (with and without the nugget) plus the genetic random factor. For comparison purposes, we also fitted a model only with the correction for rows and columns (eqn. (1)). The procedure was repeated a total of $R = 500$ times. For SpATS, we used cubic B-spline bases of dimension $L = 18$ and $P = 51$, jointly with nested basis for the columns with $P_N = 27$.

For measuring models' performance, the discrepancy between the BLUPs for the genotypic effects and the corresponding true (simulated) quantities was measured in terms of the empirical version of the global root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{360} \sum_{i=1}^{360} (\widehat{c}_{gi} - c_{gi})^2}.$$

As far as the REML estimates of σ_g^2 is concerned, the behavior was evaluated in terms of the bias.

Figure 3 shows the boxplots of log10(RMSE) associated to the genotypic BLUPs and the REML estimates of σ_g^2, for each of the four models considered. In terms of the log10(RMSE), and as could have been expected, the worst performance corresponds to the model including only the correction for rows and columns. The remaining three models present a similar behavior, with the best approach being the AR×AR model including the nugget, followed by our SpATS model. However, if we focus on the REML estimates of σ_g^2, we observe that the AR×AR model excluding the nugget tends to overestimate the genetic signal. This behavior can explain the larger heritability (on average) provided by this model in comparison with SpATS or the AR×AR model including the nugget (see Table 4). Surprisingly, the model including only the correction for rows and columns provides good estimates of σ_g^2. However, it also produces the lowest heritabilities. Note that this is the expected behavior, as the variance associated to the measurement error will be inflated in this case.
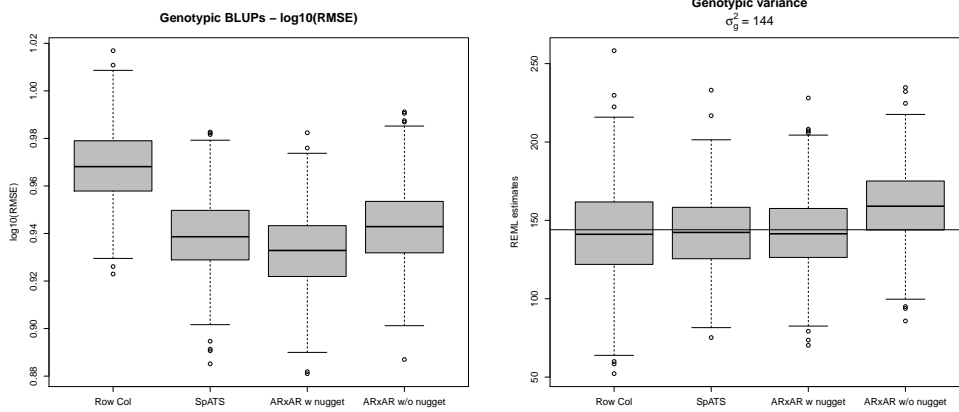
Figure 3: For the barley uniformity data simulation study: Boxplots, based on 500 simulated data sets, of the log10(RMSE) associated to the genotypic random factor and the REML estimates of σ_g^2. "Row Col" corresponds to the model including only the correction for rows and columns, "SpATS" to our proposal, "AR×AR w nugget" and "AR×AR w/o nugget" to the AR×AR model with and without the nugget respectively.

Table 3: Numerical results associated to the simulation study based on the barley uniformity data. For the genotypic BLUPs, the $\log_{10}$(RMSE) is shown. For the REML estimates of σ_g^2 the results show the bias. In all cases, averages and standard deviations over 500 simulated data sets are presented. "Row Col" corresponds to the model including only the correction for rows and columns, "SpATS" to our proposal, "AR×AR w nugget" and "AR×AR w/o nugget" to the AR×AR model with and without the nugget respectively.

| | Model | | | |
	Row Col	SpATS	AR×AR w nugget	AR×AR w/o nugget
Genotypic RMSE	0.968 (0.016)	0.939 (0.016)	0.933 (0.016)	0.943 (0.016)
$\sigma_g^2 = 144$	-2.425 (30.061)	-1.516 (24.538)	-1.563 (24.142)	15.281 (24.391)
H_g^2	0.399 (0.062)	0.481 (0.054)	0.494 (0.053)	0.534 (0.051)

5.2 Separable Gaussian Autoregressive Process

In this study, data was generated assuming a separable Gaussian AR process of order 1 in the row and column directions. Specifically, the following model was considered:

$$\boldsymbol{y} = \boldsymbol{Z}_g \boldsymbol{c}_g + \underbrace{\boldsymbol{\xi} + \boldsymbol{\varepsilon}}_{\boldsymbol{\epsilon}}, \tag{23}$$

where $\boldsymbol{c}_g$ denotes the genotypic effects, with $\boldsymbol{c}_g \sim N(\boldsymbol{0}, \sigma_g^2 \boldsymbol{I}_{m_g})$, $\boldsymbol{\xi} \sim N\left(\boldsymbol{0}, \sigma_s^2 \boldsymbol{\Sigma}\right)$ where $\boldsymbol{\Sigma}_{lp} = \rho_r^{|u_l - u_p|} \rho_c^{|v_l - v_p|}$, $\boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n\right)$, and $\boldsymbol{\epsilon} = \boldsymbol{\xi} + \boldsymbol{\varepsilon}$.

Different levels of genetic variation $\left(\sigma_g^2 \in \{0.25; 1; 4\}\right)$ and autocorrelations $(\rho_r = \rho_c \in \{0.1; 0.5; 0.9\})$ were studied. In all cases, $\sigma^2 = \sigma_s^2 = 1$ and a total of $m_g = 100$ genotypes, each replicated twice, were considered. The 100 genotypes were allocated to the plots following an alpha design, in blocks of size 10. The field layout thus consisted of 10 blocks (rows) and 20 columns ($n = 200$).

For each data set simulated as described above, we fitted our SpATS model, including the PS-ANOVA spatial field and the genetic random factor. For Gilmour et al. (1997)'s approach we considered a model including the genetic random factor, the AR×AR process and the nugget ($\boldsymbol{\varepsilon}$). The procedure was repeated a total of $R = 1000$ times. For SpATS, cubic B-spline bases of dimension 13 and 23 were assumed for the row and column positions, respectively, and nested B-spline bases, with half the dimension, were used.

As for the simulation reported in Section 5.1, models' performance was measured in terms of the RMSE (for the genotypic BLUPs), and the bias for the REML estimates of the variances (σ_g^2, σ^2 and σ_s^2).

Table 4 shows, for those runs for which both SpATS and AR×AR models converged, the results for all levels of genetic variation and autocorrelations considered in the study. The table lists the averages and standard deviations of the RMSEs for the genotypic BLUPs and the bias and associated standard deviation for σ_g^2, σ^2 and σ_s^2. The percentage of fitting models that converged (over 1000 runs) is also shown in the table, jointly with the effective dimension associated to the PS-ANOVA spatial field (excluding the intercept). If we focus on the genotypic random factor, both approaches behave similarly for all scenarios considered. As could have been expected, the RMSE decreases as the genotypic signal increases. A similar performance between both approaches is also observed for σ_g^2. The interpretation of the results for σ^2, σ_s^2 and ED_s requires however a more detailed analysis. First note that when both ρ_r and $\rho_c \to 1$, then $\mathbb{C}\text{ov}\left(\boldsymbol{\xi}\right) \to \sigma_s^2 \mathbb{I}_{200}$, where $\mathbb{I}_n$ is a $n \times n$ matrix of ones. In this case, for each realization of the stochastic process (23), the spatially dependent random error $\boldsymbol{\xi}$ would be a constant vector, and no spatial variation would be therefore present. On the other hand, when both ρ_r and $\rho_c \to 0$ then $\mathbb{C}\text{ov}\left(\boldsymbol{\xi}\right) \to \sigma_s^2 \boldsymbol{I}_{200}$. Thus, $\boldsymbol{\xi}$ is confounded with the nugget or independent error $\boldsymbol{\varepsilon}$, and $\mathbb{C}\text{ov}\left(\boldsymbol{\epsilon}\right) \to \left(\sigma_s^2 + \sigma^2\right) \boldsymbol{I}_{200}$. To some extent, it would also imply that no spatial variation is present and that all is measurement error. As pointed out by Piepho et al. (2015), these two extreme cases may cause convergence problems when fitting the AR×AR model. Our results are in concordance with those previous findings, and we refer to that paper for a more comprehensive discussion. Besides, it could also explain the systematic bias (of opposite sign but similar magnitude) in the estimates of σ_s^2 and σ^2 provided by the AR×AR model. As can be observed, the AR×AR model tends to underestimate σ^2 (negative

Table 4: Numerical results associated to the study when the underlying simulated model follows an AR×AR process. For the REML estimates of σ_g^2, σ^2 and σ_s^2 the results show the bias. For the genotypic BLUPs, the $\log_{10}(\text{RMSE})$ is shown. ED_s denotes the effective dimension associated to the PS-ANOVA spatial field. In all cases, averages and standard deviations are presented. The results correspond to those runs for which both SpATS and AR×AR models converged.

σ_g^2	$\rho_c = \rho_r$	Model	Conv. (%)	Genotypic RMSE	σ_g^2	σ^2	σ_s^2	ED_s
	0.9	SpATS	100.0	-0.375 (0.032)	-0.014 (0.132)	0.021 (0.159)	-	9.10 (2.82)
		AR×AR	68.8	-0.374 (0.032)	-0.017 (0.134)	-0.136 (0.285)	0.199 (1.485)	-
0.25	0.5	SpATS	100.0	-0.355 (0.030)	-0.008 (0.159)	0.306 (0.211)	-	15.27 (5.03)
		AR×AR	96.2	-0.356 (0.030)	-0.018 (0.159)	-0.112 (0.316)	0.168 (0.486)	-
	0.1	SpATS	100.0	-0.341 (0.032)	0.002 (0.190)	0.769 (0.265)	-	8.96 (4.16)
		AR×AR	93.8	-0.341 (0.032)	-0.013 (0.190)	-0.368 (0.632)	0.376 (0.628)	-
	0.9	SpATS	100.0	-0.220 (0.031)	-0.076 (0.192)	0.022 (0.162)	-	8.35 (2.88)
		AR×AR	63.7	-0.220 (0.031)	-0.076 (0.189)	-0.201 (0.329)	0.117 (1.299)	-
1	0.5	SpATS	100.0	-0.180 (0.030)	-0.080 (0.233)	0.327 (0.230)	-	13.54 (4.67)
		AR×AR	94.1	-0.186 (0.030)	-0.085 (0.225)	-0.153 (0.367)	0.231 (0.634)	-
	0.1	SpATS	100.0	-0.152 (0.030)	-0.083 (0.276)	0.797 (0.293)	-	8.10 (3.73)
		AR×AR	90.8	-0.155 (0.030)	-0.085 (0.273)	-0.408 (0.663)	0.442 (0.640)	-
	0.9	SpATS	100.0	-0.143 (0.032)	-0.299 (0.332)	0.021 (0.164)	-	7.89 (2.76)
		AR×AR	61.8	-0.143 (0.033)	-0.305 (0.324)	-0.272 (0.368)	0.187 (1.513)	-
4	0.5	SpATS	100.0	-0.088 (0.031)	-0.308 (0.388)	0.334 (0.235)	-	12.01 (4.22)
		AR×AR	90.7	-0.095 (0.031)	-0.312 (0.375)	-0.217 (0.406)	0.295 (0.548)	-
	0.1	SpATS	100.0	-0.048 (0.031)	-0.315 (0.439)	0.792 (0.298)	-	7.35 (3.36)
		AR×AR	88.5	-0.053 (0.031)	-0.324 (0.446)	-0.492 (0.637)	0.515 (0.635)	-

bias), and this is counteracted by overestimating σ_s^2. This effect is especially remarkable for low autocorrelations. As far as the SpATS model is concerned, for large autocorrelation values the approach performs as expected providing unbiased estimates of σ^2. When the autocorrelations decrease, the bias also increases. However, it is worth remembering that in this case $\mathbb{C}\text{ov}\,(\boldsymbol{\epsilon}) \to \left(\sigma_s^2 + \sigma^2\right)\boldsymbol{I}_{200}$. Thus, our SpATS model is correct when estimating the variance associated to the measurement error. Note that for all simulations $\sigma_s^2 = 1$ was considered, and that the bias associated with σ^2 approaches this value as ρ_r and ρ_c decrease. Finally, the estimated effective dimension for the PS-ANOVA spatial field also reflects the expected performance of the model: for low and large autocorrelations (small spatial variation), the model provides smaller ED_s than for medium values.

6   Applications

To further illustrate our proposal in large-scale experiments, we considered in this Section two data sets, one of a field trial on wheat conducted in Chile and discussed in the paper by Lado et al. (2013), and the other on sugar beet from a big field in France. This material was kindly provided by the breeding company SESVanderHave (Tienen, Belgium).

6.1   Sugar beet data, France

In this section a big field of sugar beet is analyzed. The field experiment was located in France, in the year 2011. Data were recorded for eight traits, and the results will be presented for

the trait α-amino nitrogen in millimol per liter. All material was provided by the breeding company SESVanderHave (Tienen, Belgium).

The field experiment consisted of 31 trials, with a total of 2411 plots and 1095 genotypes. All the trials contained four common checks, the other 1091 genotypes were observed in one of the trials. The trials were different in size, with the most common layout an alpha design with 36 genotypes and two replicates. Four of the trials were unreplicated, 26 trials had an alpha design with two replicates, and one trial had an alpha design with three replicates.

The field layout is shown in Figure 4(a). The 31 trials are represented by a different color. Note that the layout of the field is irregular, that there are no trials in the lower right corner (white areas). Nonetheless, no data manipulation, such as e.g. filling in the missing values to make the layout regular, is required to use our approach.

For this experiment, we assumed a model including random factors for rows (c_r) columns (c_c) and trials (c_t). The genetic lines were also included in the model as random (c_g) but the genetic checks were included as fixed (β_c). To be more precise, and using the same notation as in Section 2, the following mixed model was fitted

$$y = f(u, v) + X_c\beta_c + Z_gc_g + Z_rc_r + Z_cc_c + Z_tc_t + \varepsilon.$$

where X_c is the $(n \times 4)$ design matrix assigning observations to genetic checks and Z_g is the $(n \times 1091)$ design matrix associated to genetic lines, with $c_g = (c_{g1}, \ldots, c_{g1090})^t \sim N\left(0, \sigma_g^2 I_{1090}\right)$. The rows of these two matrices corresponding to genetic lines and genetic checks, respectively, have all their elements equal to zero. For the rest of the random factors, we assumed $c_r \sim N\left(0, \sigma_r^2 I_{26}\right)$, $c_c \sim N\left(0, \sigma_c^2 1_{113}\right)$ and $c_t \sim N\left(0, \sigma_t^2 I_{31}\right)$. For the tensor-product P-spline, a basis dimension of 29 and 53 was chosen for the row and column positions, respectively, and nested bases, with half the dimension, were used. Accordingly, the model has 1789 coefficients to be estimated and 2411 observations. Despite the large dimension, the fitting process needed around 18 seconds.

Figure 4 depicts the raw yield data, the fitted values, the residuals, the fitted spatial trend (i.e., the PS-ANOVA component but excluding the intercept), and the genotypic BLUPs. As can be observed, the fitted spatial trend is successful in recovering the complex spatial pattern across the field, and the residual plot suggest that the spatial independence assumption for the error vector ε might be appropriate. Table 5 shows the model and (estimated) effective dimensions associated to each of the five smooth components of the PS-ANOVA spatial trend and each random factor. Note that the estimation procedure presented in Section 4 provides, in some cases, effective dimensions that are exactly zero, meaning that this model component does not contribute or have an impact on the trait of interest. If we focus on the genetic signal, we have an effective dimension of about 590.5. In this case, it is easy to show that $\text{rank}\left([X, Z_g]\right) - \text{rank}\left(X\right) = 1090$. Equivalently, there is only one eigenvalue equal to zero, which implicitly induces a zero-mean constraint on the BLUPs for c_g. As a consequence, an estimate of the generalized heritability is $\widehat{H}_g = 590.5/(1091 - 1) = 0.54$.

6.2   Wheat data, Chile

The study consisted on 384 advanced lines from two breeding programs in Chile and Uruguay (South America). The lines were evaluated in two different environments, Santa Rosa (SR)

(a) Field layout   (b) Raw yield data   (c) Fitted values

(d) Residuals' spatial plot   (e) Fitted spatial trend   (f) Genotypic BLUPs
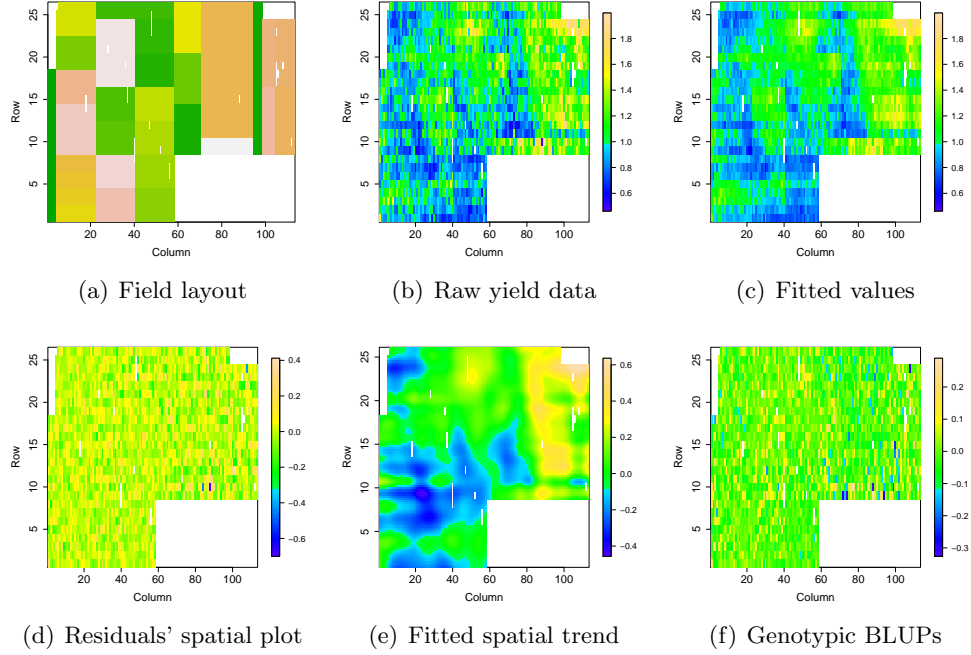
Figure 4: For the sugar beet experiment in France: field layout, raw yield data, fitted values, fitted spatial trend, residuals' spatial plot and genotypic BLUPs.

Table 5: For the sugar beet experiment in France: model and effective dimensions associated to the PS-ANOVA spatial trend and each random factor. The letter u denotes the row position, v the column position, and c_r, c_r, c_t and c_g the row, column, trial and genetic line random factors, respectively.

Dimensions	Model components								
	$f_u(u)$	$f_v(v)$	$vh_u(u)$	$uh_v(v)$	$f_{u,v}(u,v)$	c_r	c_c	c_t	c_g
Model (m_k)	27	51	27	51	364	26	113	31	1091
Effective (ED_k)	0.0	12.4	21.1	0.0	120.1	18.3	13.1	22.8	590.5

and Cauquenes (CQ), under three different levels of water supply and in two consecutive years, 2011 and 2012. More precisely, for both 2011 and 2012, the lines were evaluated in Santa Rosa under mild water stress (MWS) and fully irrigated (FI) conditions, and for 2012 in Cauquenes, a dry region, under severe water stress (WS). Different traits of interest were considered, as grain yield (GY), thousand kernel weight (TKW), number of kernels per spike (NKS) and days to heading (DH). It should be noted that, in 2012, only GY was considered. For each trial, an alpha-lattice design with 20 incomplete blocks, each containing 20 genotypes, was used, and two replicates were sown. Each block comprised 1 column and 40 rows, yielding a total of 800 ($20 \times 40 \times 2$) plots on the field.

Based on these data, the final aim pursued in the paper by Lado et al. (2013) was to perform genomic selection on the basis of genotyping-by-sequencing methods. As noted by the authors, the modeling of the spatial variation of the phenotypic data in each trial has an impact on the prediction accuracy of the subsequent genomic selection. As could have been expected, the better the modeling, the larger the prediction accuracy. In this paper we thus re-analyzed the phenotypic data using our approach. For the sake of simplicity we only present here the results for 2011. For each trait (GY, TKW, NKS and DH) and water condition (MWS and FI) the following model was assumed

$$\boldsymbol{y} = f(\boldsymbol{v}, \boldsymbol{u}) + \boldsymbol{Z}_g \boldsymbol{c}_g + \boldsymbol{Z}_r \boldsymbol{c}_r + \boldsymbol{Z}_c \boldsymbol{c}_c + \boldsymbol{\varepsilon},$$

with $\boldsymbol{c}_g \sim N\left(\boldsymbol{0}, \sigma_g^2 \boldsymbol{I}_{384}\right)$, $\boldsymbol{c}_r \sim N\left(\boldsymbol{0}, \sigma_r^2 \boldsymbol{I}_{40}\right)$ and $\boldsymbol{c}_c \sim N\left(\boldsymbol{0}, \sigma_c^2 \boldsymbol{I}_{20}\right)$. For the tensor-product P-spline, a basis dimension of 43 and 23 was assumed for the row and column positions, respectively, and, as usual, we used nested bases, with half the dimension. Under this representation, each model has about 802 coefficients to be estimated and 800 observations (there are missing values), but the fitting processes needed between 2 and 10 seconds.

Figures 5(a) and 6(a) depict the raw data for each trait and water condition. The fitted spatial trends are shown in Figures 5(b) and 6(b), and Figures 5(c) and 6(c) shows the spatial plot of the residuals. Table 6 shows the effective dimensions related to each of the five smooth components of the PS-ANOVA spatial trend as well as those associated to the row and column random factors. As for the previous example, some ED_k are zero or close to zero. For instance, for the DH and FI condition, the spatial variation is mainly modeled by the smooth trend over the rows and the row random factor, with the rest of components having a rather low or even null impact. All results suggest that for both, the MWS and FI conditions, the GY is the trait presenting the largest spatial variation, and DH the one with the lowest. This is in concordance with the estimated heritability also presented in Table 6, with the largest and the lowest heritability having been obtained for DH and GY, respectively. It should be noted that, in all cases, the heritability estimate obtained using SpATS is larger than the broad sense heritability reported in the paper by Lado et al. (2013).
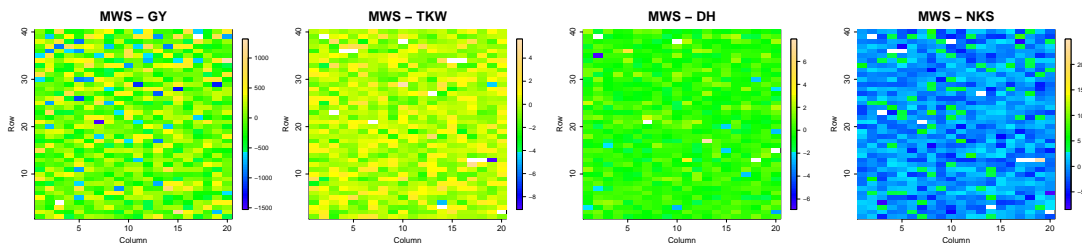
7   Discussion

We have presented a powerful and efficient new approach to the modeling of field trials. Effects of genotypes are described by a mixed model in a standard way, on top of tensor product P-splines to fit the spatial field.
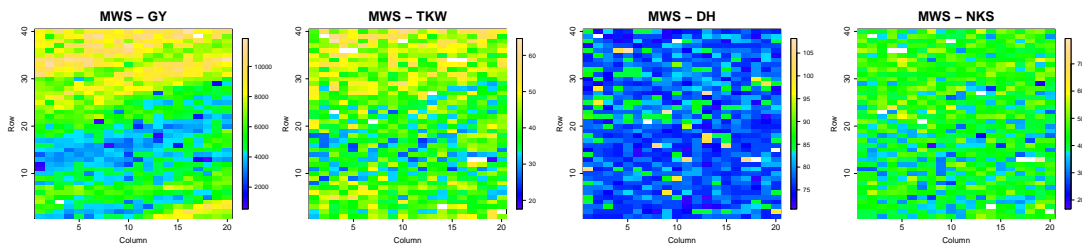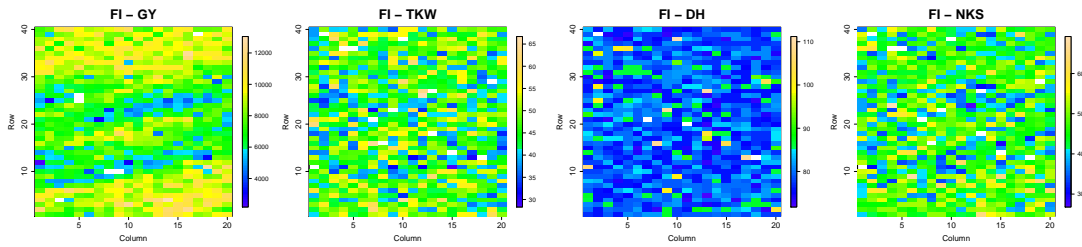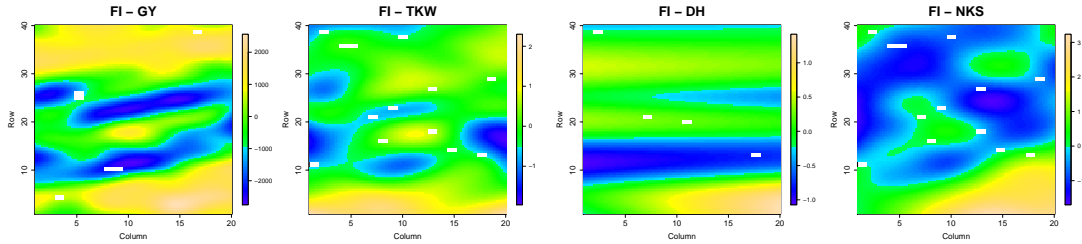
(a) Raw data
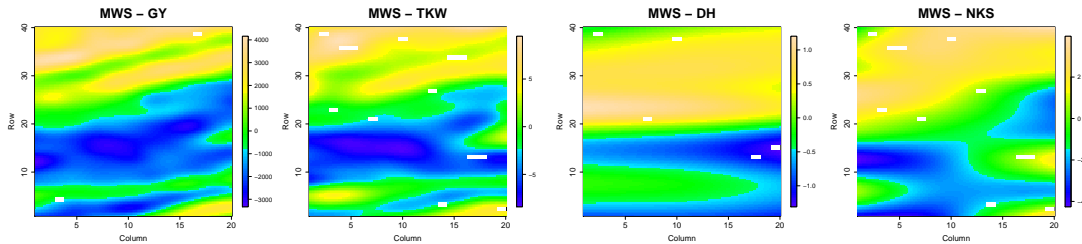
(b) Fitted spatial trend

(c) Residuals' spatial plot

Figure 5: Raw data, fitted spatial trend and residuals' spatial plot for the Chilean wheat data in Santa Rosa, 2011, for each trait (GY: Grain Yield; TKW: thousand kernel weight; DH: days to heading; NKS: number of kernels per spike) and mild water stress (MWS) condition.

25

(a) Raw data

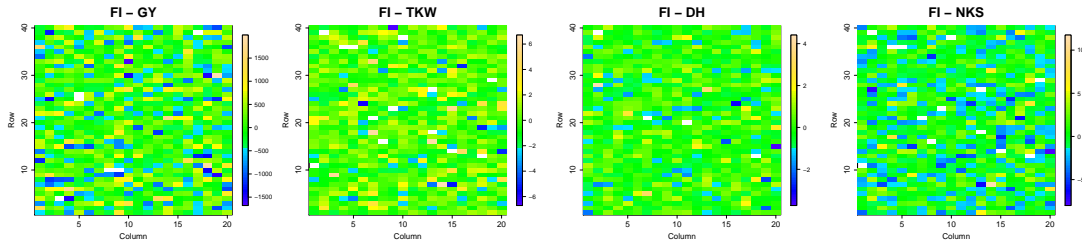(b) Fitted spatial trend

(c) Residuals' spatial plot

Figure 6: Raw data, fitted spatial trend and residuals' spatial plot for the Chilean wheat data in Santa Rosa, 2011, for each trait (GY: Grain Yield; TKW: thousand kernel weight; DH: days to heading; NKS: number of kernels per spike) and fully irrigated (FI) condition.

26

Table 6: For the Chilean wheat data: Effective dimensions associated to the spatial trend and the row and column random factors, and generalized heritability. The letter u denotes the row position, v the column position, and $\boldsymbol{c}_r$ and $\boldsymbol{c}_r$ the row and column random factors, respectively. The results shown are for Santa Rosa in 2011 and for each trait (GY: Grain Yield; TKW: thousand kernel weight; DH: days to heading; NKS: number of kernels per spike) and condition (MWS: Mild water stress; FI: Fully irrigated).

		Condition and Trait							
		MWS				FI			
		GY	TKW	DH	NKS	GY	TKW	DH	NKS
	$f_u(u)$	8.8	8.5	6.6	2.4	10.2	2.9	7.2	1.3
	$f_v(v)$	0.0	0.7	0.5	0.0	0.6	0.7	0.0	0.0
	$vh_{(}u)$	16.8	16.7	1.1	7.6	8.4	0.4	1.4	0.5
Effective dimension	$uh_v(v)$	0.0	0.0	0.0	1.8	1.6	0.0	0.1	0.7
(ED$_k$)	$f_{u,v}(u,v)$	90.6	49.5	0.2	0.6	64.9	17.9	0.0	12.2
	$\boldsymbol{c}_r$	20.2	5.2	3.7	12.5	13.7	17.8	7.7	23.7
	$\boldsymbol{c}_c$	6.9	0.4	5.6	3.6	6.3	0.0	0.0	0.3
	Total	143.3	81.0	17.7	28.5	105.7	39.7	16.4	38.8
H_g^2		0.74	0.94	0.96	0.84	0.69	0.94	0.98	0.86

Our approach breaks with the current tradition to model the spatial component as correlated noise. We believe that this offers large advantages. First, as shown in the simulations, modeling spatial correlation by e.g. autoregressive processes may need a lot of manual ad hoc tuning. In contrast, our SpATS model is very robust and runs without user intervention, as our testing on very many real trials has shown. A second advantage is that model selection is substantially simplified. Furthermore, the applications showed that our method can easily deal with irregular field layouts, and local patches with special behavior become easily visible. We believe that having an explicit estimate of the spatial field variation is very helpful. In many cases it is more complex than an autoregressive process. Besides, the ANOVA-type decomposition of tensor product P-splines provides interesting insights on the field trial being analyzed.

An important result of this paper is the expression of the standard as well as generalized heritability on the basis of the genetic effective dimension. This result provides, on the one hand, a common definition of heritability that can be used for a broad range of statistical models used for the analysis of field trials. On the other hand, we believe that the link between the genetic effective dimension and the (generalized) heritability measure brings new insights to the interpretation of both quantities. To the best of our knowledge, this is the first time that this relation is presented in the literature. The study of the appropriateness of this result to non-Gaussian responses represents an interesting area of research.

What has been said here about genetic variation (and the associated heritability measure) also applies to other random effects incorporated into a mixed model. This suggests moving the interpretation of the results obtained by fitting a mixed model from estimated variances

components to estimated effective dimensions. Moreover, evaluating the results in terms of effective dimensions also furnishes a common scale that will allow determining those random model components that contribute the most to explaining the observed phenotypic variation.

A current line of research is focused on the extension of our SpATS model to the analysis of multi-environment trials (METs, trials carried out in multiple environments or contexts, see, e.g., Smith et al., 2009). In the analysis of METs we are faced with several challenges. First, the modeling of a different spatial trend for each environment, and second the need to account for the (possible) interaction between the genotype and the environment. The first challenge may be approached by including an interaction between a factor (the environment) and a smooth surface (the tensor product P-spline). The second challenge is related to the inclusion of correlated random effects (whenever the genotype effect is treated as random). This possibility will allow assuming a different genetic variance for each environment, and, at the same time, modeling the (within environments) correlation between genotypes sharing the same environment and the (across environments) correlation of genotype effects along the different environments. Further work is warranted to develop computationally-efficient estimation procedures in this setting.

Our calculations were done with the R-package SpATS that can be freely downloaded from https://cran.r-project.org/package=SpATS. However, it is worth remembering that estimation can also be accommodated using any standard mixed model software.

Acknowledgements

Appendix A   Nested bases

With large data sets the computation of $\boldsymbol{Z}_s$ in (7), as well as its inner product, can demand a lot of time, especially for large values of L and P. Lee et al. (2013) propose to speed up computation by using nested bases. The idea is to reduce the dimension of the marginal B-spline bases $\hat{\boldsymbol{B}}$ and $\breve{\boldsymbol{B}}$ (and therefore the associated number of coefficients to be estimated), but only for the smooth-by-smooth interaction term , i.e., $f_{u,v}$. As pointed out by the authors, this simplification can be justified by the fact that the main effects, f_u and f_v, and the smooth varying coefficient terms, h_v and h_u, would in fact explain most of the structure (or spatial trend) presented in the data, and so a less rich representation of the interaction term could be needed.

Let $\hat{\boldsymbol{B}}_N$ and $\breve{\boldsymbol{B}}_N$ be two reduced marginal B-spline basis of dimension $n \times L_N$ ($L_N < L$) and $n \times P_N$ ($P_N < P$) with associated penalty matrices $\hat{\boldsymbol{D}}_N^t \hat{\boldsymbol{D}}_N$ and $\breve{\boldsymbol{D}}_N^t \breve{\boldsymbol{D}}_N$, respectively.
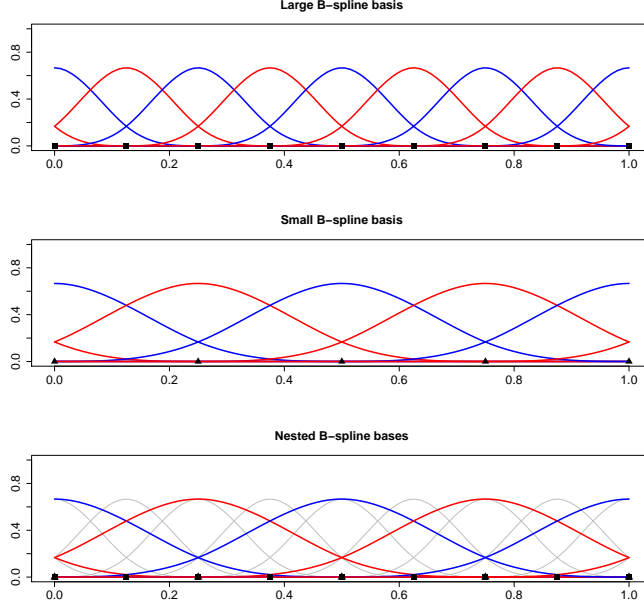
Figure A1: Example of nested B-spline bases.

Then, the reduced mixed model matrix $\boldsymbol{Z}_s$ for the PS-ANOVA model is constructed as follows

$$\boldsymbol{Z}_s = \left[\boldsymbol{Z}_v, \boldsymbol{Z}_u, \boldsymbol{Z}_v \square \boldsymbol{u}, \boldsymbol{v} \square \boldsymbol{Z}_u, \widetilde{\boldsymbol{Z}}_v \square \widetilde{\boldsymbol{Z}}_u \right],$$

where $\widetilde{\boldsymbol{Z}}_v = \hat{\boldsymbol{B}}_N \widetilde{\boldsymbol{U}}_v^N$ and $\widetilde{\boldsymbol{Z}}_u = \breve{\boldsymbol{B}}_N \widetilde{\boldsymbol{U}}_v^N$, with $\widetilde{\boldsymbol{U}}_v^N$ and $\widetilde{\boldsymbol{U}}_v^N$ being the matrices containing the eigenvectors associated to the non-zero eigenvalues of $\hat{\boldsymbol{D}}_N^t \hat{\boldsymbol{D}}_N$ and $\breve{\boldsymbol{D}}_N^t \breve{\boldsymbol{D}}_N$, respectively.

In order to ensure that the reduced model is in fact nested in the model including only the main effects, Lee et al. (2013) showed that the number of segments that define $\hat{\boldsymbol{B}}_N$ and $\breve{\boldsymbol{B}}_N$ should be a divisor of the number of segments used in the original bases $\hat{\boldsymbol{B}}$ and $\breve{\boldsymbol{B}}$. The reasoning behind is graphically illustrated in Figure A1, that has been taken from the paper by Lee et al. (2013). The two top plots depict two cubic B-spline bases of dimension 11 and 7, respectively. The squares and triangles denote the breakpoints (knots) that define each segment. As can be observed, the knots of the small basis correspond to a subset of the knots of the large basis. This implies that the space spanned by the small basis is a subset of the space spanned by the large one. This can be seen on the plot at the bottom, where both bases overlap.

Note that the use of nested bases reduces the number of coefficients associated to $f_{u,v}$ from $(L-2)(P-2)$ to $(L_N-2)(P_N-2)$. This reduction allows being generous with the number of B-spline basis functions used for the main effects and the smooth varying coefficient terms. Our experience suggest using (a) as many segments for $\hat{\boldsymbol{B}}$ and $\breve{\boldsymbol{B}}$ as number of rows and columns in the field, respectively; and (b) half the number of segments for the nested basis.

Appendix B   Some results on mixed models and hat matrices

For given values of the variance components (σ^2 and σ_k^2, $k = 1, \ldots, q$), BLUEs for $\boldsymbol{\beta}$ and BLUPs for $\boldsymbol{c}$ can be obtained as the solution to the linear system of equations (Henderson, 1963)

$$\underbrace{\begin{bmatrix} \boldsymbol{X}^t \boldsymbol{R}^{-1} \boldsymbol{X} & \boldsymbol{X}^t \boldsymbol{R}^{-1} \boldsymbol{Z} \\ \boldsymbol{Z}^t \boldsymbol{R}^{-1} \boldsymbol{X} & \boldsymbol{G}^{-1} + \boldsymbol{Z}^t \boldsymbol{R}^{-1} \boldsymbol{Z} \end{bmatrix}}_{\boldsymbol{C}} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{c}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}^t \boldsymbol{R}^{-1} \boldsymbol{y} \\ \boldsymbol{Z}^t \boldsymbol{R}^{-1} \boldsymbol{y} \end{bmatrix} \tag{24}$$

which gives rise to closed-form expressions

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^t \boldsymbol{V}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^t \boldsymbol{V}^{-1} \boldsymbol{y}, \tag{25}$$

$$\widehat{\boldsymbol{c}} = \boldsymbol{G} \boldsymbol{Z}^t \boldsymbol{V}^{-1} (\boldsymbol{y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}) = \boldsymbol{G} \boldsymbol{Z}^t \boldsymbol{Q} \boldsymbol{y}, \tag{26}$$

where $\boldsymbol{V}$ and $\boldsymbol{Q}$ have been defined in Section 4.1. As it will be seen, the last equivalence in (26) plays an important role in our approach, and it is simply obtained by substituting $\widehat{\boldsymbol{\beta}}$ by (25). The previous expressions reveal that the hat matrices associated to the fixed, $\boldsymbol{\beta}$, and random, $\boldsymbol{c}$, effects are $\boldsymbol{H}_F = \boldsymbol{X} \left(\boldsymbol{X}^t \boldsymbol{V}^{-1} \boldsymbol{X} \right)^{-1} \boldsymbol{X}^t \boldsymbol{V}^{-1}$ and $\boldsymbol{H}_R = \boldsymbol{Z} \boldsymbol{G} \boldsymbol{Z}^t \boldsymbol{Q}$ respectively (as already discussed in Section 4.2).

It is worth remembering that in this paper we assume that $\boldsymbol{Z} = [\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_q]$, where each $\boldsymbol{Z}_k$ represents the design matrix associated to the k-th random factor $\boldsymbol{c}_k$, with $\boldsymbol{c} = \left(\boldsymbol{c}_1^t, \ldots, \boldsymbol{c}_q^t \right)^t$, and that the variance-covariance $\boldsymbol{G} = \bigoplus_{k=1}^q \boldsymbol{G}_k$. As a consequence

$$\widehat{\boldsymbol{c}} = \begin{bmatrix} \widehat{\boldsymbol{c}}_1 \\ \widehat{\boldsymbol{c}}_2 \\ \vdots \\ \widehat{\boldsymbol{c}}_q \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{G}_2 & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{G}_q \end{bmatrix} \begin{bmatrix} \boldsymbol{Z}_1^t \\ \boldsymbol{Z}_2^t \\ \vdots \\ \boldsymbol{Z}_q^t \end{bmatrix} \boldsymbol{Q} \boldsymbol{y}$$

$$= \begin{bmatrix} \boldsymbol{G}_1 \boldsymbol{Z}_1^t \boldsymbol{Q} \boldsymbol{y} \\ \boldsymbol{G}_2 \boldsymbol{Z}_2^t \boldsymbol{Q} \boldsymbol{y} \\ \vdots \\ \boldsymbol{G}_q \boldsymbol{Z}_q^t \boldsymbol{Q} \boldsymbol{y}, \end{bmatrix}.$$

This result implies that

$$\boldsymbol{H}_R \boldsymbol{y} = \boldsymbol{Z} \widehat{\boldsymbol{c}} = \sum_{k=1}^q \boldsymbol{Z}_k \widehat{\boldsymbol{c}}_k = \sum_{k=1}^q \boldsymbol{Z}_k \boldsymbol{G}_k \boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{y} = \sum_{k=1}^q \boldsymbol{H}_k \boldsymbol{y},$$

where $\boldsymbol{H}_k = \boldsymbol{Z}_k \boldsymbol{G}_k \boldsymbol{Z}_k^t \boldsymbol{Q}$. Accordingly, the hat matrix associated to the random part of model (12) can be decomposed as a sum of independent hat matrices, each related to a specific random factor in the model, i.e.,

$$\boldsymbol{H}_R = \sum_{k=1}^q \boldsymbol{H}_k.$$

Finally, it can also be shown (see, e.g., eqn. (9d) and Appendix 1 in Johnson and Thompson, 1995) that

$$\widehat{\varepsilon} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\widehat{\boldsymbol{c}} = \boldsymbol{R}\boldsymbol{Q}\boldsymbol{y}, \tag{27}$$

and the residuals' hat matrix is thus $\boldsymbol{H}_{\varepsilon} = \boldsymbol{R}\boldsymbol{Q}$.

Appendix C   Equivalence between effective dimensions definitions

This section presents the equivalence between the definition given by Cui et al. (2010) of the effective dimension associated to a model's component and ED_k. Specifically, Cui et al. (2010) define the effective dimension of a random factor $\boldsymbol{c}_k$ as

$$\mathrm{ED}\left(\boldsymbol{Z}_k\right) = \lim_{\varsigma \to +\infty} \mathrm{trace}\left(\boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t\left(\boldsymbol{V} + \varsigma\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^t\right)^+\right) \tag{28}$$

$$= \mathrm{trace}\left(\boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t\left[(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V}\left(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}\right)\right]^+\right), \tag{29}$$

where $\boldsymbol{\Gamma}^+$ denotes the Moore-Penrose pseudoinverse of $\boldsymbol{\Gamma}$, $\boldsymbol{W}$ is a positive definite matrix, ς is a positive scalar, and $\boldsymbol{P}_{\boldsymbol{X}} = \boldsymbol{X}\left(\boldsymbol{X}^t\boldsymbol{X}\right)^{-1}$. The expression given above defines the effective dimension of a model's component as the trace of the ratio of that "component's modeled variance matrix" $\left(\boldsymbol{Z}_k\boldsymbol{G}_k\boldsymbol{Z}_k^t\right)$ to "total variance matrix" $\left(\boldsymbol{V} + \varsigma\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^t\right)$. Note that this definition treats $\boldsymbol{\beta}$ as a random vector with variance-covariance $\varsigma\boldsymbol{W}$. However, as pointed out by the authors, a fixed effect can be viewed as the limiting case of a random effect for which the variance-covariance matrix goes to infinity (i.e., when $\varsigma \to +\infty$). Similarly to $\mathrm{ED}\left(\boldsymbol{Z}_k\right)$, Cui et al. (2010) define the effective dimension for the error term as

$$\mathrm{ED}\left(\boldsymbol{\varepsilon}\right) = \lim_{\varsigma \to +\infty} \mathrm{trace}\left(\boldsymbol{R}\left(\boldsymbol{V} + \varsigma\boldsymbol{X}\boldsymbol{W}\boldsymbol{X}^t\right)^+\right) \tag{30}$$

$$= \mathrm{trace}\left(\boldsymbol{R}\left[(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V}\left(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}\right)\right]^+\right),$$

and show that

$$n = \mathrm{ED}(\boldsymbol{X}) + \sum_{k=1}^q \mathrm{ED}(\boldsymbol{Z}_k) + \mathrm{ED}\left(\boldsymbol{\varepsilon}\right)$$

$$= \mathrm{rank}\left(\boldsymbol{X}\right) + \sum_{k=1}^q \mathrm{ED}(\boldsymbol{Z}_k) + \mathrm{ED}\left(\boldsymbol{\varepsilon}\right).$$

The definition given in Cui et al. (2010) thus partitions the number of observations n into independent effective dimensions for the model's components and error. In author's words, this result jointly with (28) and (30), suggests interpreting the effective dimension of a model's component $\mathrm{ED}\left(\boldsymbol{Z}_k\right)$ as the fraction of response variation attributed to that individual effect, and the same applies to the error term.

To show that $\mathrm{ED}_k = \mathrm{ED}\left(\boldsymbol{Z}_k\right)$, we use results derived in the paper by Hoog et al. (1990). Given that (see identity (1) in Hoog et al., 1990)

$$\boldsymbol{Q} = \boldsymbol{V}^{-1} - \boldsymbol{V}^{-1}\boldsymbol{X}\left(\boldsymbol{X}^t\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^t\boldsymbol{V}^{-1} = \left[(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\boldsymbol{V}\left(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}}\right)\right]^+,$$

we have

$$
\begin{aligned}
\mathrm{ED}_k &= \mathrm{trace}\left(\boldsymbol{Z}_k^t \boldsymbol{Q} \boldsymbol{Z}_k \boldsymbol{G}\right) \\
&= \mathrm{trace}\left(\boldsymbol{Z}_k \boldsymbol{G} \boldsymbol{Z}_k^t \boldsymbol{Q}\right) \\
&= \mathrm{trace}\left(\boldsymbol{Z}_k \boldsymbol{G} \boldsymbol{Z}_k^t \left[(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{V}\,(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\right]^+\right) \\
&= \mathrm{ED}\left(\boldsymbol{Z}_k\right).
\end{aligned}
$$

We would like to note that the equivalence between (28) and (29) can also be proved using results of Hoog et al. (1990). Furthermore, we also have that

$$
\begin{aligned}
\mathrm{ED}_{\boldsymbol{\varepsilon}} &= \mathrm{trace}\left(\boldsymbol{R}\boldsymbol{Q}\right) \\
&= \mathrm{trace}\left(\boldsymbol{R}\left[(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{V}\,(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{X}})\right]^+\right) \\
&= \mathrm{ED}\left(\boldsymbol{\varepsilon}\right).
\end{aligned}
$$

Appendix D   SpATS package

This section contains a brief description of the developed R-package associated to this paper. The package can be freely downloaded from https://cran.r-project.org/package=SpATS, where a more detailed depiction of it use can be found. The main function of the package is SpATS(), which fits the spatial model presented in Section 4. Numerical and graphical summaries of the fitted spatial model can be obtained, as usual in R, by using summary.SpATS(), variogram.SpATS(), predict.SpATS() and plot.SpATS(). In the implementation of the package, the sparse structure of the design matrix associated with the genotype has been taken into account, which, in combination with the estimation procedure presented in Section 4.1 and the possible use of nested B-spline bases, makes the package computational efficient, allowing the analysis of very large datasets.

By way of example, we present here the syntax for the Australian wheat trial example discussed in the paper by Gilmour et al. (1997). The aim of this trial was the evaluation of advance breeding lines and commercial varieties. The trial consisted of 107 varieties, which were sown in three replicates, each replicate being a complete block. Each block comprised 5 columns and 22 rows, yielding a total of 330 ($5 \times 22 \times 3$) plots on the field. To meet the 110 plots per replicate, from the 107 varieties, three were sown twice in each of these. For more details about the trial, we refer the readers to the cited paper. On the basis of the results shown in Gilmour et al. (1997), the following statistical model was assumed

$$
\boldsymbol{y} = \boldsymbol{X}_g \boldsymbol{\beta}_g + f\left(\boldsymbol{u}, \boldsymbol{v}\right) + \boldsymbol{Z}_r \boldsymbol{c}_r + \boldsymbol{Z}_c \boldsymbol{c}_c + \boldsymbol{\varepsilon},
$$

where $\boldsymbol{\beta}_g$ is a (106×1) vector of fixed variety (genetic) effects, and $\boldsymbol{X}_g$ is the corresponding (330×106) design matrix. Note that the dimension of the genetic effect is $m_g - 1$ (where $m_g = 107$) since the intercept is included in $f\left(\boldsymbol{u}, \boldsymbol{v}\right)$. Here, $\boldsymbol{c}_r \sim N\left(\boldsymbol{0}, \sigma_r^2 \boldsymbol{I}_{22}\right)$ and $\boldsymbol{c}_c \sim N\left(\boldsymbol{0}, \sigma_c^2 \boldsymbol{I}_{15}\right)$ are vectors of row and column random effects respectively, and $\boldsymbol{\varepsilon} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{330}\right)$.

The dataset can be found in the R-package agridat, under the name gilmour.serpentine. Here there is a brief summary of the data

```
> library(agridat)
> GS <- gilmour.serpentine
> summary(GS)
     col            row          rep              gen           yield
 Min.   : 1   Min.   : 1.0   R1:110   TINCURRIN   :  6   Min.   :194.0
 1st Qu.: 4   1st Qu.: 6.0   R2:110   VF655       :  6   1st Qu.:469.0
 Median : 8   Median :11.5   R3:110   WW1477      :  6   Median :617.5
 Mean   : 8   Mean   :11.5            (WWH*MM)*WR*:  3   Mean   :591.8
 3rd Qu.:12   3rd Qu.:17.0            (WqKPWmH*3Ag:  3   3rd Qu.:713.5
 Max.   :15   Max.   :22.0            AMERY       :  3   Max.   :925.0
                                      (Other)     :303
```

The dataset contains the column and row positions (`col` and `row` variables), the block (variable `rep`), the variety (`gen`) and the yield (`yield`). In order to incorporate in the model the random factors of rows and columns, we need first to create the corresponding factor variables, that we denote as `col_f` and `row_f`, and we then fit the model

```
>  GS$col_f = factor(GS$col)
>  GS$row_f = factor(GS$row)

> fit.SpATS <- SpATS(response = "yield", genotype = "gen", genotype.as.random = FALSE,
+ spatial = ~ PSANOVA(col, row, nseg = c(16,20), degree = 3, nest.div = 2),
+ fixed = NULL, random = ~ row_f + col_f,
+ data = GS, control =  list(tolerance = 1e-03, monitoring = 1))

Timings:
SpATS 0.38 seconds
All process 0.57 seconds
```

Through `response` and `genotype` arguments, users specify the name of the variables in the dataset that contains, respectively, the response variable (phenotype) of interest and the genotype or variety. The genotype can be included in the model either as fixed (default) or random (`genotype.as.random = TRUE`). For modeling the spatial trend, argument `spatial`, we consider 16 segments (`nseg`) for the column position and 20 for the row. This, jointly with the fact we use cubic B-splines, `degree = 3`, gives rise to B-spline bases of dimension $P = 16 + 3 = 19$ and $L = 20 + 3 = 23$ for the columns and rows, respectively. By specifying the argument `nest.div = 2`, we indicate the use of nested bases, with half the number of segments of the original ones (see Appendix A). The fixed and random effects to be included in the model are indicated in `fixed` and `random`, and argument `control` allows to modify some default parameters that control the fitting process. For instance, the tolerance for the convergence criterion for the variance components can be altered using this argument, as well as the maximum number of iterations. Under this representation, the model has a total of 322 coefficients, but it took less than 1 seconds to be fitted. A numerical summary of the fitted model can be obtained by calling the function `summary()`. By indicating the argument `which = "all"`, we obtain both the estimates of the variance components and the effective dimensions

```
> summary(fit.SpATS, which = "all")
[...]

Variance components:
                 Variance           SD      log10(lambda)
row_f           4.397e+02    2.097e+01           0.67320
col_f           4.442e+03    6.665e+01          -0.33128
f(col)          1.245e+04    1.116e+02          -0.77895
f(row)          7.240e+01    8.509e+00           1.45657
f(col):row      7.847e+02    2.801e+01           0.42159
col:f(row)      6.490e-06    2.548e-03           8.50408
f(col):f(row)   2.530e+03    5.030e+01          -0.08684

Residual        2.072e+03    4.552e+01

Dimensions:
                Effective    Model    Nominal    Ratio    Type
gen                 106.0      106        106     1.00       F
Intercept             1.0        1          1     1.00       F
row_f                12.6       22         21     0.60       R
col_f                10.3       15         14     0.74       R
col                   1.0        1          1     1.00       S
row                   1.0        1          1     1.00       S
row:col               1.0        1          1     1.00       S
f(col)                2.3       17         17     0.14       S
f(row)                1.0       21         21     0.05       S
f(col):row            2.6       17         17     0.15       S
col:f(row)            0.0       21         21     0.00       S
f(col):f(row)         7.5       99         99     0.08       S

Total               146.3      322        320     0.46
Residual            183.7
Nobs                 330

Type codes: F 'Fixed'    R 'Random'    S 'Smooth/Semiparametric'
```

In this example, there are seven variance components, five associated to the spatial trend, one associated to the row random effects, one to the column random effects, and the residual variance σ^2. The column `log10(lambda)` shows the logarithm of base 10 of the smoothing parameters, i.e., the ratio between the residual variance and the variance component. As far as the dimensions is concerned, for each component in the model (either fixed, random or spatial), the function returns (a) the `effective` dimension or effective degrees of freedom, (b) the `model` dimension, i.e., the number of parameters to be estimated, (c) the `nominal` dimension, which, for the random components is the model dimension minus one, lost due to

the constraint of zero-mean imposed to them; and (d) the `ratio` between the `effective` and the `nominal` dimension. It is worth remembering that, if the variety had been included as random, the `ratio` for the variety would have provided an estimate of the so-called generalized heritability as proposed by Oakey et al. (2006). If we focus on the random effects for the rows and columns, we have that the effective dimensions are, respectively, about the 60% and the 74% of the nominal dimensions. As discussed in Section 3.2, for the spatial trend we have, in total, 8 components (excluding the intercept). The linear effects for the rows and the columns (`row` and `column`), as well as the linear interaction (`row:col`), represent the fixed or unpenalized part of the tensor-product P-spline. The remaining five components, i.e., the main effects (`f(row)` and `f(col)`), the smooth varying coefficient terms (`f(col):row` and `row:f(col)`); and the smooth-by-smooth interaction component (`f(col):f(row)`) correspond to the penalized or random part, and have been extensively discussed in Sections 3.1 and 4.2. On the basis of the effective dimensions associated to each of these five components, we may inferred that most of the trend has been captured by the main effect and the smooth varying coefficient term along the column position, but also by the smooth-by-smooth interaction term, for which we have an effective dimension of 7.6.

To complement those numerical results, the `SpATS` package furnishes different graphical results that can be used to further explore the fitted model. Specifically, the sample variogram can be obtained using the function `variogram()`, which can also be plotted; and the function `plot()` depicts six different graphics: the raw data, the fitted data, the residuals, the estimated spatial trend (excluding the intercept), the genotypic BLUEs (or BLUPs) and their histogram. Except for the histogram, the plots are depicted in terms of the spatial coordinates (e.g., the rows and columns of the field).

```
> plot(fit.SpATS)
```

```
> plot(variogram(fit.SpATS))
```

The result of the above code is shown in Figures D2 and D3. The spatial plots of the residuals and the genotypic BLUEs ($\widehat{\boldsymbol{\beta}}_g$) do not suggest the presence of any extra spatial pattern that should have been taken into account. A similar conclusion can be drawn from the sample variogram of the residuals shown in Figure D3. Finally, note that the fitted spatial trend takes values between -300 and 200, whereas the residuals vary between -100 and 100. This result highlights what could have been expected based on the raw data, that spatial (plot-to-plot) variation is larger than random (plot-to-plot) variation.

References

Besag, J. and D. Hidgon (1999). Bayesian analysis of agricultural field experiments. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 61*, 691–746.

Cappa, E. P. and R. J. C. Cantet (2008). Bayesian estimation of a surface to account for a spatial trend using penalized splines in an individual-tree mixed model. *Canadian Journal of Forest Research 37*, 2677–2688.
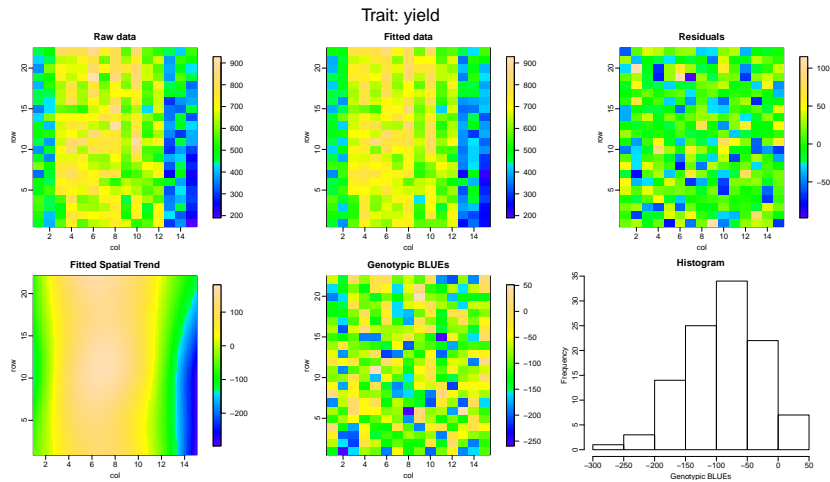
Figure D2: Graphical results provided by the SpATS package for the Australian wheat trial.
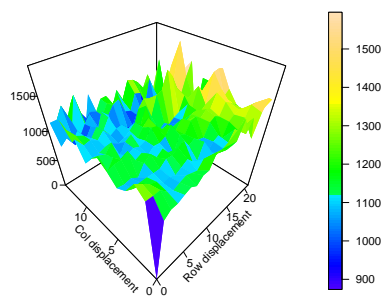

Figure D3: Sample variogram of the residuals provided by the SpATS package for the Australian wheat trial.

Cui, Y., J. S. Hodges, X. Kong, and B. P. Carlin (2010). Partitioning degrees of freedom in hierarchical and other richly-parameterized models. *Technometrics 52*, 124–136.

Cullis, B., A. Smith, and N. Coombes (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics 11*(4), 381–393.

Cullis, B. R. and A. C. Gleeson (1991). Spatial analysis of field experiments - an extension to two dimensions. *Biometrics 47*, 1449–1460.

Currie, I. D. and M. Durban (2002). Flexible smoothing with P-splines: a unified approach. *Statistical Modelling 2*(4), 333–349.

Currie, I. D., M. Durban, and P. H. C. Eilers (2006). Generalized linear array models with applications to multidimensional smoothing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 68*(2), 259–280.

Dierckx, P. (1993). *Curve and Surface Fitting with Splines*. Oxford: Clarendon Press.

Durban, M., C. A. Hackett, J. W. McNicol, A. C. Newton, W. T. B. Thomas, and I. D. Currie (2003). The practical use of semiparametric models in field trials. *Journal of Agricultural, Biological, and Environmental Statistics 8*(1), 48–66.

Eilers, P. H. C. (1999). Discussion of Verbyla et al. *Journal of the Royal Statistical Society, Series C (Applied Statistics) 48*, 300–311.

Eilers, P. H. C., I. D. Currie, and M. Durban (2006). Fast and compact smoothing on large multidimensional grids. *Computational Statistics and Data Analysis 50*, 61–76.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with B-splines and penalties. *Statistical Science 11*(2), 89–121.

Eilers, P. H. C. and B. D. Marx (2003). Multidimensional calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligence Laboratory Systems 66*, 159–174.

Eilers, P. H. C. and B. D. Marx (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics 2*, 637–653,.

Eilers, P. H. C., B. D. Marx, and M. Durban (2015). Twenty years of P-splines. *SORT 39*(2), 149–186.

Engel, B. (1990). The analysis of unbalanced linear models with variance components. *Statistica Neerlandica 44*, 195–219.

Gilmour, A. R., B. R. Cullis, and A. P. Verbyla (1997). Accounting for natural and extraneous variation in the analysis of field experiments. *Journal of Agricultural, Biological and Environmental Statistics 2*, 269–293.

Green, P., C. Jennison, and A. Seheult (1985). Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society, Series B (Statistical Methodology) 47*, 299–315.

Harville, D. A. (1977). Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems. *Journal of the American Statistical Association 72*(358), 320–338.

Hastie, T. J. and R. J. Tibshirani (1990). *Generalized additive models*. London: Chapman & Hall.

Henderson, C. R. (1963). Selection index and expected genetic advance. *Statistical Genetics and Plant Breeding 982*, 141–163.

Hoog, F. R., T. P. Speed, and E. R. Willians (1990). On a matrix identity associated with generalized least squares. *Linear Algebra and its Applications 127*, 449–456.

Johnson, D. L. and R. Thompson (1995). Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *Journal of Dairy Science 78*, 449–456.

Lado, B., I. Matus, A. Rodríguez, L. Inostroza, J. Poland, F. Belzile, A. del Pozo, M. Quincke, M. Castro, and J. von Zitzewitz (2013). Increased genomic prediction accuracy in wheat breeding through spatial adjustment of field trial data. *G3 (Genes, Genomes, Genetics) 3*(12), 2105–2114.

Lee, D.-J. (2010). *Smoothing mixed model for spatial and spatio-temporal data*. PhD thesis, Department of Statistics, Universidad Carlos III de Madrid, Spain.

Lee, D.-J. and M. Durban (2011). P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling 11*(1), 49–69.

Lee, D.-J., M. Durban, and P. H. C. Eilers (2013). Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested basis. *Computational Statistics and Data Analysis 61*, 22–37.

Oakey, H., A. Verbyla, W. Pitchford, B. Cullis, and H. Kuchel (2006). Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics 113*(5), 809–819.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems (with discussion). *Statistical Science 1*, 505–527.

Patterson, H. D. and R. Thompson (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika 58*(3), 545–554.

Piepho, H. P., J. Möhring, M. Pflugfelder, and W. Hermann (2015). Problems in parameter estimation for power and AR(1) models of spatial correlation in designed field experiments. *Communications in Biometry and Crop Science 10*(1), 3–16.

Piepho, H. P. and E. R. Willians (2010). Linear variance models for plant breeding trials. *Plant Breeding 129*, 1–8.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Robbins, K. R., J. E. Backlund, and K. D. Schnelle (2012). Spatial corrections of unreplicated trials using a two-dimensional spline. *Crop Science 52*, 1138–1144.

Rodríguez-Álvarez, M. X., D.-J. Lee, T. Kneib, M. Durban, and P. H. C. Eilers (2015). Fast smoothing parameter separation in multidimensional generalized P-splines: the sap algorithm. *Statistics and Computing 25*, 941–957.

Ruppert, D., M. P. Wand, and R. Carroll (2003). *Semiparametric Regression*. Cambridge University Press.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika 78*(4), 719–727.

Schnabel, S. K. and P. H. C. Eilers (2009). Optimal expectile smoothing. *Computational Statistics and Data Analysis 52*, 4168–4177.

Smith, A., B. Cullis, and R. Thompson (2009). Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. *Biometrics 57*, 1138–1147.

Stefanova, K. T., A. B. Smith, and B. R. Cullis (2009). Enhanced diagnostics for the spatial analysis of field trials. *Journal of Agricultural, Biological, and Environmental Statistics 14*(4), 392–410.

Taye, G. and P. M. Njuho (2008). Smoothing fertility trends in agricultural field experiments. *Statistics 42*, 275–289.

Verbyla, A. P., B. R. Cullis, M. G. Kenward, and S. J. Welham (1999). The analysis of designed experiments and longitudinal data by smoothing splines (with discussion). *Journal of the Royal Statistical Society, Series C (Applied Statistics) 48*, 269–312.

Wand, M. P. (2003). Smoothing and mixed models. *Computational Statistics 18*(2), 223–249.

Williams, E. R. and D. J. Luckett (1988). The use of uniformity data in the design and analysis of cotton and barley variety trials. *Australian Journal of Agricultural Research 39*, 545–554.

Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association 93*(441), 120–131.

You, C., S. Müller, and J. T. Ormerod (2016). On generalized degrees of freedom with application in linear mixed models selection. *Statistics and Computing 26*, 199–210.

Zimmerman, D. L. and D. A. Harville (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics 47*, 223–239.