

# Tensor P-Spline Smoothing for Spatial Analysis of Plant Breeding Trials

Hans-Peter Piepho<sup>a,\*</sup>, Martin P. Boer<sup>b</sup>, Emlyn R. Williams<sup>c</sup>

<sup>a</sup> Institute of Crop Science, University of Hohenheim, 70 593 Stuttgart, Germany

<sup>b</sup> Biometris, Wageningen University & Research, Wageningen, The Netherlands

<sup>c</sup> Statistical Consulting Unit, Australian National University, Canberra, ACT 2601, Australia

\* Corresponding author.

E-mail addresses: [hans-peter.piepho@uni-hohenheim.de](mailto:hans-peter.piepho@uni-hohenheim.de) (H.P. Piepho), [martin.boer@wur.nl](mailto:martin.boer@wur.nl) (M.P. Boer), [emlyn@alphastat.net](mailto:emlyn@alphastat.net) (E.R. Williams).

## ABSTRACT

Large agricultural field trials may display irregular spatial trends that cannot be fully captured by a purely randomization-based analysis. For this reason, paralleling the development of analysis-of-variance procedures for randomized field trials, there is a long history of spatial modelling for field trials, starting with the early work of Papadakis on nearest neighbour analysis, which can be cast in terms of first or second differences among neighbouring plot values. This kind of spatial modelling is amenable to a natural extension using P-splines, as has been demonstrated in recent publications in the field. Here, we consider the P-spline framework, focussing on model options that are easy to implement in linear mixed model packages. Two examples serve to illustrate and evaluate the methods. A key conclusion is that first differences are rather competitive with second differences. A further key observation is that second differences require special attention regarding the representation of the null space of the smooth terms for spatial interaction, and that an unstructured variance-covariance structure is required to ensure invariance to translation and rotation of eigenvectors associated with that null space. We develop a strategy that permits fitting this model with ease, but the approach is more demanding than that needed for fitting models using first differences. Hence, even though in other areas second differences are very commonly used in the application of P-splines, our main conclusion is that with field trials first differences have advantages for routine use.

**Keywords:** Kronecker product; Tensor product; Penalized regression; Intrinsic autoregression; Null space; Agricultural Field Trial.

## 1. Introduction

Designed field trials play a central role in plant breeding and variety testing for most agricultural crops species. The experimental units of such trials (plots) are usually arranged on a rectangular grid and typically show spatial covariance among adjacent plots. There is an ample literature on analysis of field trials using spatial models. One of the earliest papers is Papadakis (1937), who introduced nearest neighbour (NN) methods, an idea taken up later in the seminal papers by Bartlett (1978) and Wilkinson et al. (1983) and subsequently developed further by other authors to cover a wide spectrum of spatial variance-covariance models in

common usage today. Spatial models allow making adjustments for irregular spatial heterogeneity and have been shown to improve efficiency in cases where a purely randomization-based mixed model cannot adequately capture the major heterogeneity across the trial area.

One of the earliest papers applying P-splines in field trials, though only in one dimension, is Currie and Durbán (2002). Recently, the use of P-splines (Eilers and Marx, 1996) has been suggested for two-dimensional spatial analysis of field trials (Rodríguez-Álvarez et al., 2018). In this framework, spatial covariance translates into smooth spatial trend. P-splines have a close connection with linear variance and random walk models (Boer et al., 2020). They use B-spline bases for regression on covariates and provide a very flexible framework for smoothing of trends in multiple dimensions, and some of the most prominent applications for this type of modelling are spatial and spatio-temporal smoothing for large environmental datasets (Lee, 2010; Lee and Durbán, 2011; Lee et al., 2013). The simplest approach to spatial smoothing is by geosadditive models (Ruppert et al., 2003, p. 258), where each spatial dimension, i.e. geographical longitude and latitude, is smoothed by separate terms in an additive fashion. Often, however, non-additive extensions are needed to allow for interaction. A key challenge with environmental data is that measurements may be irregularly spaced and hence special care must be taken when modelling the interaction between dimensions. In field trials, matters are considerably simplified by the fact that plots are usually arranged on a regular grid (Appendix A).

Here we will focus on P-spline approaches that allow the partitioning of total variance in an analysis-of-variance (ANOVA) type fashion (Lee and Durbán, 2011; Wood et al., 2013; Wood, 2017), which is also at the heart of the recently proposed P-spline based approaches for field trials (Rodríguez-Álvarez et al., 2018). We seek to use a P-spline framework that is flexible enough for agricultural field trials, yet simple enough for straightforward implementation using mixed model packages. In fact, it is a key feature of P-splines that they have a mixed model representation, meaning that parameters are amenable to estimation by residual maximum likelihood (REML) (Patterson and Thompson, 1971). This makes them particularly appealing because the mixed model framework allows accommodating other features of plant breeding trial data, including genetic correlation among related breeding lines, which can be modelled using genetic marker data (Meuwissen et al., 2001), and random design effects arising from the randomization layout of the trial, as others have pointed out before us (e.g., Verbyla et al., 2018). Also, embedding in a REML framework allows a likelihood-based comparison with other mixed models accounting for spatial correlation.

We consider B-splines with a penalty term of the form  $\theta u^T D^T D u$ , where  $\theta$  is a penalty parameter,  $u$  is a vector of regression coefficients with corresponding design matrix  $B$  and  $D$  is a differencing matrix (Wood, 2017). In the simplest case, the B-spline basis provides a smooth in just one spatial dimension, but it may also be extended in two dimensions using the Kronecker product of bases for rows and columns (see Appendix A). The penalty determines the smoothness of the coefficients  $u$  along the spatial dimensions, because it strives for small values of  $Du$ . The mixed model representation of P-splines rests on the fact that the singular penalty matrix  $P = D^T D$  is formally equivalent to a precision matrix of a random effect  $u$ . In order to exploit that equivalence when using a mixed model package for fitting P-splines,  $P$  needs to be converted to a suitable variance-covariance matrix. In the mixed model representation this leads to a fixed-effects component, representing an unpenalized part, and a random-effects component, representing the penalized part (Currie and Durbán, 2002; Wand and Omerod, 2008; Lee and Durbán, 2011). The resulting models are very similar to those proposed in Verbyla et al. (2018) for smoothing using two covariates based on an integrated

squared second derivative penalty used in cubic spline smoothing; but, as will be explained, there are slight differences in the way the unpenalized terms are handled. As our main impetus for this paper is to provide a P-spline framework that is conveniently implemented in a general REML-based mixed model package, we closely follow the philosophy set out in Wood et al. (2013), primarily focussing on such penalties that have just a single parameter and upon conversion give rise to a variance-covariance matrix that is linear in the parameters. It is acknowledged that some penalty matrices that have been proposed for tensor-product P-splines do not have this desirable property and therefore are not considered here in detail. In our discussion, we will briefly review a few important examples of such penalties in order to put our framework into perspective.

A particular focus will be on the choice of differencing matrix  $D$ . In the context of field trials, the early development of NN methods was based on second differences (Papadakis, 1937; Wilkinson et al., 1983), but later the focus turned to first differences (Besag and Kempton, 1986; Kempton et al., 1994) and the closely related linear variance (LV) model (Williams, 1986), because they are simpler and often provide a good fit. Second differences were rarely used subsequently (but see Green et al., 1985). The recent proposal to use P-splines (Rodríguez-Álvarez et al., 2018) in field trials, however, has led to a revival of this option. In light of the new options that P-splines provide, a revisit of the long-standing question in field trials whether first or second differences are preferable is in order.

This paper has several objectives: (i) To provide a thorough derivation of the two-dimensional P-spline representation as a mixed model, focusing on an ANOVA-type decomposition and paying particular attention to the separation between penalized and unpenalized terms, (ii) to show how such P-splines are intimately related to other commonly used spatial models, such as LV models, (iii) to compare first and second difference penalties in several field trial datasets, and (iv) to compare P-splines empirically with other spatial models such as the first-order autoregressive (AR1) model (Gilmour et al., 1997). The rest of the paper is structured as follows. Section 2 considers a single column of plots to introduce key concepts and notation. This is extended to smooth marginal (main) effects for rows and columns in Section 3. Section 4 shows how interaction between row and column smooths may be added. The framework will be illustrated using examples in Section 5 and compared to other spatial models. The paper ends with a discussion in Section 6.

## 2. Smoothing along a single column of plots

In this section we will set the stage considering a single column of  $k$  plots. Key elements of the notation that will be introduced step-by-step, also for use throughout subsequent sections, are summarized in Table 1 for ease of reference. Let  $B_r$  be a  $k \times m_r$  matrix of  $q$ -th degree  $[(q+1)$ -th order] B-spline bases (Eilers and Marx, 1996) for the row-coordinates  $h_r$  of the plots (typically the row numbers). Generally, the first interior knot will be placed at the first plot and the last interior knot at the last plot of the column of plots. Often, the remaining interior knots are placed at the other plots. Alternatively, the remaining interior knots can be placed at larger distances between the first and last plot, which reduces the number of random coefficients and hence the dimension of the mixed model equations. The number of B-spline bases,  $m_r$ , will generally be given by the number of interior knots ( $i_r$ ) plus  $(q-1)$ . Thus, if the interior knots are placed at every plot, we will have  $m_r = k + q - 1$ .

- Table 1 about here -

The trend down the column may be modelled by the regression term  $B_r u_r$ , where  $u_r$  is a vector of  $m_r$  regression coefficients. To penalize these, we employ the  $(m_r - p) \times m_r$  matrix  $D_r$  of  $p$ -th differences, from which the penalty term is  $\theta_r u_r^T P_r u_r$ , where  $\theta_r$  is a penalty parameter and  $P_r = D_r^T D_r$  is the penalty matrix. Classical spline smoothing adds this penalty to the residual sum of squares (Green et al., 1985). If a mixed model representation of P-splines is used, the penalty is added to the (residual) log-likelihood (Paul Eilers, in discussion of Verbyla et al., 1999). The penalty is recognized to be equivalent to the term in the log-likelihood for normal random effects, when  $\theta_r = \sigma_r^{-2}$ , where  $\sigma_r^2$  is a variance parameter, and hence the precision matrix is  $\sigma_r^{-2} P_r$ . To fit this with a mixed model package, the precision matrix must usually be converted to a variance-covariance matrix. This conversion is not unique because the precision matrix is singular. To resolve this issue, one may use the spectral decomposition  $P_r = U_r \text{diag}(d_r) U_r^T$ , where the columns of  $U_r$  are the eigenvectors and  $d_r$  is the vector of eigenvalues. Denote the subvector of  $d_r$  containing the  $m_r - p$  positive eigenvalues by  $d_{+r}$  and the corresponding eigenvectors by the columns of  $U_{+r}$ . Further, denote the  $p$  eigenvectors corresponding to the zero eigenvalues in  $d_r$  by  $U_{0r}$ . Then observing that  $U_r U_r^T = I_{m_r}$ , the regression term may be reparameterized by a one-to-one transformation as

$$B_r u_r = B_r U_r U_r^T u_r = (B_r U_{0r}) (U_{0r}^T u_r) + (B_r U_{+r}) (U_{+r}^T u_r) = X_r \beta_r + Z_r w_r \quad (1)$$

where  $X_r = B_r U_{0r}$ ,  $\beta_r = U_{0r}^T u_r$ ,  $Z_r = B_r U_{+r}$ , and  $w_r = U_{+r}^T u_r$ . Finally, using the fact that  $P_r = U_{+r} \text{diag}(d_{+r}) U_{+r}^T$  and replacing  $U_{+r}^T u_r$  by  $w_r$ , the penalty becomes

$$\theta_r u_r^T P_r u_r = \theta_r u_r^T U_{+r} \text{diag}(d_{+r}) U_{+r}^T u_r = \theta_r w_r^T \text{diag}(d_{+r}) w_r. \quad (2)$$

Thus, in the transformed model, only  $w_r$  is penalized and is formally equivalent to a random effect, whereas  $\beta_r$ , representing the null space of the penalty, is unpenalized and therefore is formally equivalent to a fixed effect in a mixed model. There are several ways to utilize these key results to fit the P-splines using a mixed model package. In all of these, as a direct consequence of the derivation in (1), it is required to fit the unpenalized fixed-effect term  $X_r \beta_r$  (Wood et al., 2013; Lee et al., 2021). As regards the random term, we may explicitly fit  $Z_r w_r$  with variance-covariance matrix  $\text{var}(w_r) = \sigma_r^2 \text{diag}(d_{+r})$ . This is equivalent to fitting  $B_r u_r$  as random with singular variance-covariance matrix  $\text{var}(u_r) = \sigma_r^2 P_r^+$ , where  $P_r^+ = U_{+r} \text{diag}(d_{+r}^{-1}) U_{+r}^T$  is the Moore-Penrose inverse of  $P_r$  (Boer et al., 2020), because  $u_r^T \{\text{var}(u_r)\}^+ u_r = \theta_r w_r^T \text{diag}(d_{+r}) w_r$  as in (2). In fact, it is proved in Lee et al. (2021) that we may use any generalized inverse of the precision matrix as our variance-covariance matrix and that this provides a unique fit if we simultaneously fit  $X_r \beta_r$ , representing the null space or unpenalized space, of the penalty matrix. For example, if a positive-definite variance-covariance matrix is required by the mixed model package, one could use the g-inverse  $P_r^- = P_r^+ + U_{0r} U_{0r}^T$ . In what follows, we will use the Moore-Penrose inverse  $P_r^+$ , computed from the spectral decomposition, for convenience and refer to a *unique fit* whenever the null space of  $P_r$  is fully represented by fixed effects. It is pointed out here that for  $p > 1$ ,  $U_{0r}$ , and

hence  $X_r$ , is determined only up to an orthogonal rotation, i.e.  $U_{0r}$  can be replaced by  $U_{0r}R$  for any orthogonal rotation matrix  $R$  (Harville, 1997, p. 538). Statistical packages may differ in the particular form of  $U_{0r}$  obtained in the spectral decomposition of a singular matrix. This indeterminacy is of no consequence, however, because the fit of  $X_r\beta_r$  by fixed effects, as well as the residual likelihood for the random effects, is invariant to the rotation  $R$ . In later sections, we will consider two-dimensional extensions of such models that fit specific components of  $X_r\beta_r$  by random effects for smoothing and in this case the fit is not rotation invariant.

We note in passing that in addition to the smooth plot effect  $B_ru_r$ , we will generally fit an independently distributed residual vector  $e$  with variance-covariance matrix  $\sigma_e^2 I_k$ , corresponding to a 'nugget' variance in geostatistical parlance (Piepho and Williams, 2010). The full mixed model then takes the usual form  $y = X\beta + Zu + e$ , where the null-space term  $X_r\beta_r$  is integrated with the fixed effects  $X\beta$  and the smooth term  $B_ru_r$  is integrated with the random effects  $Zu$ .

### Special cases:

- (i) When  $p=1$  (first differences), then  $\sqrt{m_r}U_{0r} = 1_{m_r}$  and from the properties of B-spline bases  $\sqrt{m_r}X_r = \sqrt{m_r}B_rU_{0r} = 1_k$ , corresponding to the general intercept.
- (ii) When  $p=2$  (second differences),  $X_r = B_rU_{0r}$  can be replaced with  $(1_k:h_k)$ , where  $h_k^T = (1, 2, \dots, k)$ , because  $(1_k:h_k)$  is in the column space of  $X_r$  (Lee, 2010; Lee and Durbán, 2011; Lee et al., 2021). Thus, to achieve a unique smooth, we may add a regression on the serial plot number in the fixed part of the model. While the representation of  $X_r$  as a regression on plot numbers has some intuitive appeal and is therefore very popular, there is no strict need to use this representation, and one can just use the parameterization in terms of the original eigenvectors, i.e.  $X_r = B_rU_{0r}$ . We think that when it comes to a two-dimensional extension of the smoothing approach, it is in fact preferable to stick with this parameterization (see Section 4.1), because  $X_r$  may form part of a term that needs to be smoothed.
- (iii) If  $p=1$  and  $q=1$ , and the interior knots are placed at the plots, we have  $m_r = k + q - 1 = k$  and  $B_r = I_k$ . This model is equivalent to the LV model of Williams (1986) and also to the NN model based on first differences by Besag and Kempton (1986), as well as a first-order random walk model (Boer et al., 2020; Appendix B). Specifically, the spatial variance-covariance matrix of the LV model can be shown to be another generalized inverse of  $P_r$  (see Boer et al., 2020).
- (iv) If  $p=2$  and  $q=1$ , and the interior knots are placed at the plots, we have  $m_r = k + q - 1 = k$  and  $B_r = I_k$ . This model is equivalent to the second-differences NN model of Green et al. (1985). For a revealing link of this case with the second-order random walk model (Durbin and Koopman, 2001, p. 39) see Appendix B.

## 3. Smooth marginal effects for rows and columns on a rectangular grid of plots

Now assume that we have plots placed on a rectangular grid of  $k$  rows and  $s$  columns (possibly with some missing data). In row-column layouts it is sensible to fit effects for rows



and columns. First consider effects for rows. These may be smoothed using random effects. If observations are first ordered by rows and then by columns within rows, this corresponds to fitting random effects  $u_r$  with the same specifications for the penalty as given in Section 2 with design matrix  $B_r \otimes 1_s$ . In order to make the smooth unique, we require fitting the corresponding fixed effects  $X_r \otimes 1_s$ . The same kind of effects may be fitted for columns. Thus, we fit random effects  $u_c$  with design matrix  $1_k \otimes B_c$  and variance-covariance matrix  $\sigma_c^2 P_c^+$ . This will be unique if we add fixed effects  $1_k \otimes X_c$ , where  $X_c = B_c U_{0c}$ . We may interpret the terms  $(B_r \otimes 1_s)u_r$  and  $(1_k \otimes B_c)u_c$  as 'smooth' marginal effects for rows and columns (Wood 2017, p. 232). When  $p = 1$  (first differences), then the two unpenalized terms coincide, i.e.  $X_r \otimes 1_s = 1_k \otimes X_c = 1_k \otimes 1_s$ , corresponding to the overall intercept. When  $p = 2$  (second differences) the two unpenalized effects have an alternate representation which, apart from the overall intercept, involves a linear regression on row and column numbers with design matrices  $h_k \otimes 1_s$  and  $1_k \otimes h_s$ , respectively. We re-iterate here that while this representation is very popular and is fine for unpenalized terms, we do not advocate this representation for penalized terms as will be needed in the next section.

We note here that when the trial is randomized according to a resolvable row-column design, independent random effects for rows and columns within replicates would be routinely fitted in a randomization-based analysis. Thus, we may regard such a model as a baseline for any spatial extensions, including P-splines. If marginal smooths for rows and columns are added to a randomization-based baseline model, as suggested in this section, they compete with the independent row and column effects. Another way of looking at this is that the independent random effects for rows and columns capture any remaining non-smooth trend not represented by the smooth. Also, the associated variances can be interpreted as nugget for rows and column effects in geostatistical terminology.

## 4. Extending the smooth to cover interaction

Analogous to linear models for two-way analysis of variance (ANOVA), we may augment the smooth marginal effects  $(B_r \otimes 1_s)u_r$  and  $(1_k \otimes B_c)u_c$  with a term of the form  $(B_r \otimes B_c)u_{rc}$ , which will expand the spline space to cover interaction (Wood 2017). There are various choices for the penalty, and special attention needs to be paid to its null space because this may be quite large, e.g., when the penalty is derived from a single Kronecker product of the two marginal penalty matrices  $P_r$  and  $P_c$ , as considered in Section 4.1. In Section 4.2 we consider a penalty involving a sum of two Kronecker products that is very popular in smoothing (Lee and Durbán, 2011; Wood, 2017; Rodríguez-Álvarez et al., 2018) and is closely related to the intrinsic autoregressive (IAR) model considered in Besag and Higdon (1999).

Before going into details, we would like to stress that our philosophy here is to consider the interaction smooth and its null space as a point of departure. This view is different from other derivations that start with the marginal smooths and then from these derive the interaction smooth terms (e.g., Wood et al., 2013; Wood, 2017; Verbyla et al., 2018). We believe that it is helpful to consider the interaction term in isolation initially because this helps to understand the associated null space and make sure it is fully accounted for in the overall model. As it turns out, in Section 4.1 the null space involves terms that are confounded with the smooth marginal terms introduced in Section 3. This fact makes the derivation a bit involved, but we

think it is necessary to take this approach to be clear about the origin and fate of all terms representing the null space of smooth terms for the interaction.

#### 4.1. Penalties derived from a single Kronecker product of the two marginal penalties

In this section, we consider penalized differences  $D_{rc}u_{rc}$ , where  $D_{rc} = D_r \otimes D_c$ , and hence the penalty

$$\theta_{rc} u_{rc}^T P_{rc} u_{rc} \quad (3)$$

with  $P_{rc} = D_{rc}^T D_{rc} = D_r^T D_r \otimes D_c^T D_c = P_r \otimes P_c$ , corresponding to random effects  $u_{rc}$  with variance-covariance matrix  $\sigma_{rc}^2 P_{rc}^+$ , where  $P_{rc}^+ = P_r^+ \otimes P_c^+$ . The product-nature of the penalty already hints that its null space is substantial, so substantial in fact, that a smoothing of that null space may be useful, requiring additional penalties, as will be discussed later in this section. To determine the fixed effects needed to represent the null space of  $P_{rc}$ , note that  $(U_r \otimes U_c)(U_r^T \otimes U_c^T) = I_{m_r} \otimes I_{m_c}$  and  $U_r \otimes U_c = (U_{0r} \otimes U_{0c} : U_{0r} \otimes U_{+c} : U_{+r} \otimes U_{0c} : U_{+r} \otimes U_{+c})$ , such that using the same approach as in the one-dimensional case [see eq. (1)], the term  $(B_r \otimes B_c)u_{rc}$  can be transformed one-to-one as

$$(B_r \otimes B_c)(U_r \otimes U_c)(U_r^T \otimes U_c^T)u_{rc} = X_{00}\beta_{00} + X_{r0}\beta_{r0} + X_{0c}\beta_{0c} + Z_{rc}w_{rc} \quad (4)$$

where

$$\begin{aligned} X_{00} &= (B_r \otimes B_c)(U_{0r} \otimes U_{0c}) = X_r \otimes X_c, \\ X_{r0} &= (B_r \otimes B_c)(U_{+r} \otimes U_{0c}) = Z_r \otimes X_c, \\ X_{0c} &= (B_r \otimes B_c)(U_{0r} \otimes U_{+c}) = X_r \otimes Z_c, \text{ and} \\ Z_{rc} &= (B_r \otimes B_c)(U_{+r} \otimes U_{+c}) = Z_r \otimes Z_c \end{aligned}$$

are design matrices, and

$$\begin{aligned} \beta_{00} &= (U_{0r}^T \otimes U_{0c}^T)u_{rc}, \\ \beta_{r0} &= (U_{+r}^T \otimes U_{0c}^T)u_{rc}, \\ \beta_{0c} &= (U_{0r}^T \otimes U_{+c}^T)u_{rc}, \text{ and} \\ w_{rc} &= (U_{+r}^T \otimes U_{+c}^T)u_{rc} \end{aligned}$$

are the corresponding regression coefficients. Analogous to the derivation in Section 2, it emerges that only the term  $Z_{rc}w_{rc}$  is penalized because it is associated with the positive eigenvalues  $d_{+r}$  and  $d_{+c}$ , whereas the terms  $X_{00}\beta_{00}$ ,  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  represent the null space because they are associated with zero eigenvalues and hence need to be fitted as fixed effects to obtain a unique smooth. Note that the dimension of  $Z_{rc}$  is  $(m_r - p) \times (m_c - p)$ , whereas the null space has dimension  $m_r \times m_c - (m_r - p) \times (m_c - p)$ , which can be quite large (Wood, 2017, p. 232). This suggests that  $p$  should be chosen as small as possible.

Importantly, the unpenalized terms  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  will generally show confounding with the marginal smooths introduced in Section 3. It should also be stressed that for  $p > 1$  the dimension of the space represented by these terms is larger than that of the marginal terms in Section 3. This is why we think it is crucial in the derivation to initially consider the interaction smooth and its null space in its own right. If we insist that these fixed-effects terms be fitted to ensure uniqueness of the smooth  $Z_{rc}w_{rc}$ , then the marginal effect smooths of Section 3 are absorbed into the fixed effects and hence may be dropped. Alternatively, it may be worthwhile to replace the fixed terms  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  by smooth equivalents, which are fitted as random. In this case, however, the smooth  $Z_{rc}w_{rc}$  is no longer invariant to the choice of generalized inverse of the penalty matrix  $P_{rc}$  (Lee et al., 2021). Also, because of the confounding, the smooth marginal effects are absorbed by the smooth versions of  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  (and not vice versa!) and hence may also be dropped in this case. It follows that either way, with or without smoothing of  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$ , there is no need to explicitly add the marginal-effect smooths described in Section 3; these are accounted for by the terms in the null space of the interaction smooth for  $(B_r \otimes B_c)u_{rc}$ .

In what follows, we will consider a few important special cases paying particular attention to the null space of  $(B_r \otimes B_c)u_{rc}$  and consider how these can be turned into penalized terms. Scrutiny of these special cases then leads to our suggested general approach for tensor-spline smoothing for field trials.

#### 4.1.1 Special cases

- (i) If  $p = 1$ , then  $\sqrt{m_r}X_r = 1_k$ ,  $\sqrt{m_c}X_c = 1_s$  and  $\sqrt{m_r m_c}X_{00} = 1_k \otimes 1_s$ , coinciding with the general intercept. The design matrices  $\sqrt{m_c}X_{r0} = Z_r \otimes 1_s$  and  $\sqrt{m_r}X_{0c} = 1_k \otimes Z_c$  have dimensions  $(m_r - 1)$  and  $(m_c - 1)$ , respectively, which are usually so substantial that smoothing becomes worthwhile, though this will make the smooth non-unique. For example, for design matrix  $X_{r0}$  we may consider random effects  $w_{r0}$  with variance  $\sigma_{r0}^2 \text{diag}(d_{+r}^{-1})$ , which is equivalent to assuming design matrix  $B_r \otimes 1_s$  and random effects  $u_{r0}$  with variance-covariance matrix  $\sigma_{r0}^2 P_r^+$ . This is recognized as the smooth marginal effect for rows (Section 3). Hence, in this case both  $X_{r0}$  and  $X_{0c}$  may be absorbed into the smooth marginal effects for rows and columns, respectively. Conversely, if we insist that  $X_{r0}$  and  $X_{0c}$  are fitted as fixed, they will absorb the smooth marginal effects for rows and columns, respectively, meaning that the latter can be dropped.
- (ii) If  $p = 1$  and  $q = 1$ , and interior knots are placed at the plots, we have  $m_c = s + q - 1 = s$  and  $m_r = k + q - 1 = k$ . Moreover,  $B_c = I_s$  and  $B_r = I_k$ . Assuming that the marginal terms  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  are modelled as fixed, this model is equivalent to the LV $\otimes$ LV model of Piepho and Williams (2010) and also to the NN model based on first differences along rows and columns by Kempton et al. (1994) (also see Appendix C; Besag and Higdon, 1999; Boer et al., 2020), and all of these models can be viewed as limiting cases of the separable AR1 $\otimes$ AR1 model (Gilmour et al., 1997; Piepho and Williams, 2010). The terms  $X_{00}\beta_{00}$ ,  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  correspond to general intercept and row and column effects, respectively. The row and column effects may be many, and it may therefore be worthwhile to fit these as



random for recovery of inter-block information (Lee et al., 2021). Again, the corresponding smooth marginal effects in Section 3 may be regarded as representing these random effects already. However, when  $X_{r0}\beta_{r0}$  and  $X_{0c}\beta_{0c}$  are replaced by smooths, the smooth  $Z_{rc}w_{rc}$  is no longer unique, and also there is no longer equivalence with the  $LV\otimes LV$  model and two-dimensional first differences. Conversely, if we insist on fixed row and column effects to ensure uniqueness, the smooth marginal effects from Section 3 need to be dropped.

(iii) If  $p = 2$ , then  $X_{00}$  has dimension  $p^2 = 4$  and can be represented by fixed effects for regressions on the row and column numbers and their cross products (Lee and Durbán, 2011). The design matrices  $X_{r0}$  and  $X_{0c}$  have dimensions  $2 \times (m_r - 2)$  and  $(m_c - 2) \times 2$ , and smoothing is worthwhile. For example,  $X_{r0} = Z_r \otimes X_c$  could be smoothed separately for the two columns of  $X_c$ , which are in the same column space with  $1_s$ . Hence, the smooth would be confounded with the smooth marginal effect for rows. It has random effect vector  $w_{rc}$  with variance-covariance matrix  $\text{diag}(d_r^{-1}) \otimes \sigma_{r0}^2 I_2$ , which is equivalent to design matrix  $B_r \otimes X_c$  with random effect  $u_{r0}$  having variance-covariance matrix  $P_r^+ \otimes \sigma_{r0}^2 I_2$ . This smooth would represent the smooth marginal effect for rows. It is important to point out here that if  $X_c$  was replaced by  $(1_s : h_s)$ , the covariance structure would change, i.e., the model is not invariant to linear transformations with respect to  $X_c$ , as is well known for random-coefficient models (Longford, 1993; Wolfinger, 1996). Also note that our suggestion here involves fitting a single penalty for both columns of  $X_c$ , rather than two, as is commonly done (Wood et al., 2013; Verbyla et al., 2018). Fitting two separate penalties for  $X_{r0} = Z_r \otimes X_c$  amounts to fitting the variance-covariance matrix  $\text{diag}(d_r^{-1}) \otimes \text{diag}(\sigma_{r0(1)}^2, \sigma_{r0(2)}^2)$  for the random effect  $w_{r0}$ , where the variance  $\sigma_{r0(1)}^2$  is associated with the first column of  $X_c$  and the variance  $\sigma_{r0(2)}^2$  is associated with the second column of  $X_c$ . This structure may be compared to a random coefficient model with random intercepts and slopes. In such models, fitting a diagonal variance-covariance for intercepts and slopes means that the model is not invariant to linear transformations of the covariate. This is also the main reason, why we do not recommend replacing  $X_c$  by  $(1_s : h_s)$ . Invariance is achieved if we allow a covariance between intercept and slope. Such a covariance will also ensure invariance to a rotation of  $U_{0c}$  and hence of  $X_c$ . Thus, we can use the variance-covariance matrix  $\text{diag}(d_r^{-1}) \otimes \Sigma_{r0}$ , where the two variances on the diagonal of  $\Sigma_{r0}$  are  $\sigma_{r0(1)}^2$  and  $\sigma_{r0(2)}^2$  and the covariance is  $\sigma_{r0(1,2)}$ . It is also important to re-iterate that for  $p > 1$ , the null-space eigenvectors in  $U_{r0}$  and  $U_{0c}$ , and hence the matrices  $X_r$  and  $X_c$ , are only determined up to an arbitrary orthogonal rotation of the eigenvectors (see Appendix D for the special case  $p = 2$ ).

(iv) If  $p = 2$  and  $q = 1$  [a special case of (iii)], the null space can be written in a form involving fixed-effects regressions on row numbers within columns and on column numbers within rows. If these effects are modelled as random for recovery of information, we obviously have a random-coefficient model (Longford, 1993), and this is known to require a covariance among intercept and slope to ensure invariance to linear transformations of the plot coordinates (row and column numbers). We may either model the intercepts and slopes as independent between rows and between columns, or we may use a smooth across the rows and across the columns.

### 4.1.2 General case

The confounding of the smooth marginal effect introduced in Section 3 and of effects in the null space of  $P_{rc}$  requires adjustment of our notation and actually allows some simplification. The general approach emerging here for ANOVA-type smoothing is to fit fixed effects  $\beta$  with design matrix  $X_r \otimes X_c$ , smooth row marginal effects  $u_r$  with design matrix  $B_r \otimes X_c$  and variance-covariance matrix  $P_r^+ \otimes \sigma_r^2 I_p$ , smooth column marginal effects  $u_c$  with design matrix  $X_r \otimes B_c$  and variance-covariance matrix  $\sigma_c^2 I_p \otimes P_c^+$ , and smooth interaction effects  $u_{rc}$  with design matrix  $B_r \otimes B_c$  and variance-covariance matrix  $\sigma_{rc}^2 P_r^+ \otimes P_c^+$ . Note that we have re-labelled the effects and variance components here to reflect the absorption of marginal-effect smooths of Section 3 by null-space terms arising from the interaction smooth. We also note that the whole smooth only has three variance components, regardless of the order  $p$  of differencing, i.e., one for rows, one for columns and one for the interaction. For  $p > 1$  there are two further alternatives for smoothing the two marginal effects, which are described here just for the row smooth: In place of  $P_r^+ \otimes \sigma_r^2 I_p$  with homogeneous variance for the columns of  $X_c$ , we may either use the diagonal structure  $P_r^+ \otimes \Omega_r$  where  $\Omega_r = \text{diag}(\sigma_{r(1)}^2, \dots, \sigma_{r(p)}^2)$  or  $P_r^+ \otimes \Sigma_r$ , where  $\Sigma_r$  is an unstructured  $p$ -dimensional variance-covariance matrix. The former makes the smooth invariant to re-scaling of the spatial coordinates (Wood, 2017, p. 236), but not to translations or rotations (Appendix D). The latter has the advantage of full invariance regarding the representation of  $X_r$  and  $X_c$ . The lack of invariance to translations or rotations of the diagonal model is particularly relevant, since for  $p > 1$  the null-space eigenvectors spanning  $X_r$  and  $X_c$  are determined only up to an orthogonal rotation. In difference to the smooth terms, the fixed-effects term with design matrix  $X_r \otimes X_c$  is always rotation-invariant for  $p > 1$ . In summary, for any value of  $p$ , there are always four terms, i.e., the fixed-effect  $X_r \otimes X_c$ , the two marginal smooths for  $B_r \otimes X_c$  and  $X_r \otimes B_c$ , and the pure interaction smooth for  $B_r \otimes B_c$ .

What was just described as different options for  $p > 1$  can also be taken to subsume  $p = 1$  as a special case. For example, for general  $p$ , the marginal smooth for  $B_r \otimes X_c$  has variance-covariance matrix  $(B_r \otimes X_c)(P_r^+ \otimes \Sigma_c)(B_r^T \otimes X_c^T) = B_r P_r^+ B_r^T \otimes X_c \Sigma_c X_c^T$ . Then for  $p = 1$  this reduces to  $B_r P_r^+ B_r^T \otimes 1_s 1_s^T \sigma_r^2$ . This same reduction also holds for the other two options for  $p > 1$ , i.e. diagonal and homogeneous. For  $q = 1$  this reduces further to  $P_r^+ \otimes 1_s 1_s^T \sigma_r^2$ , and for knots placed at the plots this corresponds to one of the marginal term for rows of the LV $\otimes$ LV model (Piepho and Williams, 2010).

Despite the invariance of the unstructured model to linear transformations of  $X_r$  and  $X_c$ , the structure is notoriously difficult to fit, as is well known for random-coefficient models in general (Longford, 1993). This is because with poor scaling the variance-covariance matrix  $\Sigma_r$  or  $\Sigma_c$  can have correlations close to the boundary of the parameter space, and it can therefore be difficult to ensure that the matrix remains positive definite. Further problems arise when a variance converges to zero during iterations.

For  $p = 2$ , we may address these issues by considering an orthogonal rotation of the null-space eigenvectors. To illustrate this, consider the design matrix  $X = BU_0$  (the subscript  $r$  or  $c$  is dropped here for simplicity) as needed on the marginal smooth for columns. One possible set of null-space eigenvectors  $U_0$  is (Lee and Durbán, 2011)

$$U_0 = (a_m : b_m) \quad , \quad (5)$$

where  $a_m$  is an  $m$ -dimensional vector of 1's, scaled to have unit length, and  $b_m$  is the vector  $(1, 2, \dots, m)$ , centered and also scaled to have unit length. This can be orthogonally rotated to any other permissible set of eigenvectors by  $U_0 R(\varphi)$ , where

$$R(\varphi) = \begin{pmatrix} \cos(\varphi) & -\sin(\varphi) \\ \sin(\varphi) & \cos(\varphi) \end{pmatrix} \quad (6)$$

for  $\varphi \in [0, 2\pi]$ . When  $U_0$  is involved in a marginal smooth using the diagonal variance-covariance structure  $\Omega$ , we may find the optimal rotation by a grid search over  $\varphi$ . This is equivalent to fitting a reparameterized version of the unstructured model  $\Sigma$ , with  $\varphi$  acting as a third parameter over which the likelihood is profiled. Thus, our proposed procedure for fitting the unstructured model is to perform a grid search identifying a rotation angle that at least nearly maximizes the residual likelihood for the diagonal model, and then use that rotation to fit the unstructured model. The optimization for the diagonal model will ensure that the correlation between the two regression terms associated with the columns of  $X_r$  or  $X_c$  will be close to zero and safely removed from the boundary of the parameter space. For this purpose, it is not crucial that the correlation is zero exactly, which is why we only need a good guess of the  $\varphi$  that would nullify the correlation.

#### 4.2. Penalties derived from sum of Kronecker products

Lee and Durbán (2011) consider a penalty of the form

$$u_{rc}^T \{ \theta_{rc1} P_r \otimes I_{m_c} + \theta_{rc2} I_{m_r} \otimes P_c \} u_{rc} \quad . \quad (7)$$

Wood (2017, p. 232) points out that penalties of the form (7) can provide smoother fits than those outlined in Section 4.1. A major advantage of (7) over the interaction smooth in Section 4.1 is that its null space is  $X_r \otimes X_c$  and only has dimension  $p^2$ . For  $p = 1$ , the null space corresponds to just the intercept (Dutta and Mondal, 2015), whereas for  $p = 2$  it also comprises the regression on the row and column numbers and their products (Rodríguez-Álvarez et al., 2018). These null spaces can always be accommodated in the fixed-effects part of the mixed model at virtually no cost. A further important consequence of the low dimensionality of the null space is that there is no confounding with the marginal smooths introduced in Section 3, which may be seen as a major advantage of the IAR model. Interestingly, for  $p = 1$ ,  $q = 1$ , and knots placed at the plots, taking the conditional expectation of  $u_{i,j}$ , the fertility value for the interior plot in the  $i$ -th row and  $j$ -th column (i.e., the  $i,j$ -th element of vector  $u_{rc}$ ), given all other  $u_{i',j'}$ -values, one obtains

$$E(u_{i,j}|\dots) = \frac{1}{2}[(1-\alpha)(u_{i,j-1} + u_{i,j+1}) + \alpha(u_{i-1,j} + u_{i+1,j})] \quad (8)$$

with  $2\text{var}(u_{i,j}|\dots) = 1/(\theta_{rc1} + \theta_{rc2})$ , where  $\alpha = \theta_{rc2}/(\theta_{rc1} + \theta_{rc2})$  (modified from Dutta and Mondal, 2015). This is recognized as an NN model, where the central plot is regressed on the nearest row and column neighbours (Julian Besag, in discussion of Bartlett, 1978; Kempton and Howes, 1981). Besag and Higdon (1999) refer to this as the intrinsic autoregressive (IAR) model. Despite this appealing connection with NN models, the penalty (7), as well as variations considered for field trials (McCullagh and Clifford, 2006; Mao et al., 2020), is more difficult to translate to a standard mixed model framework because more than one parameter is involved and hence the inverse of the precision matrix is not linear in the parameters (Wood et al., 2013, p. 345). A convenient option to fit (7) using a mixed model package is by profiling the residual log-likelihood for  $\alpha$ , i.e., via a grid search over  $\alpha \in [0,1]$  (Besag and Kooperberg, 1995). It must just be kept in mind that when  $\alpha$  maximizes the residual likelihood at either  $\alpha = 0$  or  $\alpha = 1$ , the penalty in (7) changes to one with a single penalty parameter having a much larger null space, thus sacrificing its desirable properties. This is perhaps the main disadvantage of the IAR model. To obviate the problem at the boundary, a constraint may be imposed such that  $0 < \alpha < 1$ , as is done in the Separation of Anisotropic Penalties (SAP) algorithm of Rodríguez-Álvarez et al. (2015). Rodríguez-Álvarez et al. (2018) consider a simplified version of (7),

$$\theta_{rc} w_{rc}^T \{ \text{diag}(d_{+r}) \otimes I_{m_c} + I_{m_r} \otimes \text{diag}(d_{+c}) \} w_{rc} \quad (9)$$

with an associated simplification of (8), which only has a single penalty parameter, allowing this to be fitted easily with a mixed model package. The models implemented in SpATS involve adding marginal smooths for rows and columns with diagonal variance-covariance structures as described in 4.1.

## 5. Examples

### 5.1. A barley trial

Durbán et al. (2003) consider a trial with 252 barley lines laid out as a resolvable row-column design with 2 replicates, 8 rows and 34 columns per replicate [Figure 1(a)]. The replicates were adjacent so that the trial had  $k = 16$  rows and  $s = 34$  columns in total. We reanalyse this data using several special cases of our P-spline approach. In all models with spatial covariance, we allow the spatial covariance to extend across replicates. For P-splines with both  $p = 1$  and  $p = 2$ , and also for other models, we fit row and column coordinates and their product as fixed regression terms so that the deviance and Akaike information criterion based on the residual likelihood, using twice the number of variance parameters as penalty term, can be used to compare all models (Wolfinger, 1996).

- Figures 1 and 2 about here -
- Tables 2, 3, 4, and 5 about here -

Table 2 shows several models fits for  $p = 2$ ,  $i_r = k$  and  $i_c = s$ , i.e., interior knots are placed at the plots. Model M12 with  $q = 3$  and diagonal structures ( $\Omega$ ) for both marginal smooths was

used to perform a grid search determining rotation angles maximizing the deviance approximately (Figure 2). These values were then used to fit the marginal smooths with unstructured covariance ( $\Sigma$ ). We considered  $q = 3$  (cubic splines) and  $q = 1$  (ordinary second differences; Green et al., 1985). A few observations are as follows:

- (i) The unstructured model is confirmed to ensure translation invariance, as opposed to the models with diagonal or constant variance. This can be seen by comparing models M3-M6 to M6-M8, where only  $u_r$  is fitted. The diagonal models M6 and M8 fitted quite well compared to the invariant unstructured models (M3-M5). When also fitting  $u_c$ , we were not able to fit models M16-M17 with unstructured  $\Sigma_r$  and  $\Sigma_c$  because positive definiteness of these matrices could not be enforced during iterations. Fixing the covariance in  $\Sigma_c$  at zero, convergence was achieved. Note, however, that M16-M17 are equivalent to M15, which we were able to fit and which fitted best among the models in Table 2.
- (ii) M20 could not be fitted with  $q = 3$  (again fixing the covariance in  $\Sigma_c$  at zero allowed convergence), but fitted second best with  $q = 1$ , giving a fit that was almost identical to M15.
- (iii) Using the  $1_k : h_k$  as the representation of  $X_r$  in the marginal smooth gave poorer fits and also was more difficult to get to converge.
- (iv) There is only a very small difference between fits for  $q = 1$  and  $q = 3$ , suggesting that the B-splines with  $q > 1$  are not strictly needed and we can revert to simple first or second differences (Eilers, 2003) as in classical NN (Besag and Kempton, 1986; Williams, 1986; Wilkinson et al., 1983).

Table 3 shows results for first differences  $p = 1$ . These are quite competitive with second differences ( $p = 2$ ). Reducing the number of knots had hardly any effect. The best fit is obtained for  $q = 1$  and knots placed at the plots (M26) [see Figure 1(b) for the smoothed trend]. Table 4 shows the fits obtained with the IAR model with penalty (7) (Section 4.2) for a step size of 0.1 for  $\alpha$  in the grid search. Again, first differences seemed to do better than second differences. Model M41 with  $q = 1$  fitted best among all models considered for this dataset. For  $p = 2$ , we also fitted smooth marginal terms involving the full  $X_r$  and  $X_c$ , though this would not be necessary to represent the null space of the interactions smooth. We faced the same problems when trying to fit the marginal penalty for columns with unstructured  $\Sigma_c$ . When fixing the covariance to zero, convergence was easy to achieve. With the diagonal models for the marginal smooths, convergence was also unproblematic. For this model, we also report the fit for fixed  $\alpha = 0.5$  (M45 and M46), corresponding to the penalty in (9), for the case  $q = 3$ . This model is the one used in Rodríguez-Álvarez et al. (2018) and implemented in the SpATS R package, available from CRAN (<https://cran.r-project.org/package=SpATS>). The model is doing quite well in terms of AIC, but several others reported in Tables 3 and 4 fare slightly better. The best model was often obtained for  $\alpha = 1.0$ , in which case the null space is not fully represented by the model; these models would need to be extended to fully cover the null space, which we have not pursued here. We also fitted the LV $\otimes$ LV model (Piepho and Williams, 2010), which is linear in the parameters, and the AR1 $\otimes$ AR1 model, which is nonlinear (Gilmour et al., 1997) (Table 5). Both models also have good fits, with an edge in favour of the latter when a nugget was added. That model (M56) had a very similar fit to the best P-spline model in Table 3 (M26) but was inferior to the best model in Table 4 (M41).



## 5.2. A wheat trial

Stroup et al. (1994) report a trial for 56 wheat varieties laid out in  $k = 11$  rows and  $s = 22$  (Figure 3) (this is the 'Alliance' trial in their paper). The dataset displays a strong spatial trend, which is described by Stroup (2002) in these terms: 'The pattern is typical of spatial variability. In this case, relatively low yields tend to be clustered in the northwest corner. This pattern is explained by a low rise in this part of the field causing increased exposure to winter kill from wind damage, and hence depressed yield.' We here use the version of the dataset that is available in the 'agridat' package of R (Wright, 2021; <https://CRAN.R-project.org/package=agridat>), which has the yield expressed in bushels per acre. The four replicates are not arranged as a rectangular array, with some rows divided up between two replicates. We fitted the exact same models as for the barley data (Tables 2 to 5). The general picture emerging from this second example is rather similar to that from the first, so details will not be dwelt on here. It is worth mentioning that M46, which corresponds to SpATS (Rodríguez-Álvarez et al., 2018), gives a good fit in terms of AIC, landing in the mid-range of the other models reported in Tables 3 and 4. The unstructured models for the marginal smooths could only be fitted with the rotation determined by a two-dimensional grid search for the diagonal model (Figure 4). Even with this rotation, it was not easy to achieve convergence, mainly due to a variance component approaching zero. All other models were easy to fit. The results for  $p = 2$  do confirm that the diagonal model is not invariant to rotations but the unstructured model is. One main conclusion from this example is that first differences work quite well.

- Figures 3 and 4 about here -

## 6. Discussion

We considered a framework that allows the use of P-splines to model spatial trends in field trials (Rodríguez-Álvarez et al., 2018). The methodology requires making several choices, i.e., the order of differencing ( $p$ ), the degree of the B-spline basis ( $q$ ), and the number of knots ( $i_r, i_c$ ), which in field trials are always placed on a regular grid. As the two examples illustrate, there is indeed a very large number of options, and this means that model choice could easily become an overwhelming exercise. It would be advantageous for routine use if a limited set of models could be identified that works well over a broad range of settings, not requiring an 'intricate process of tailoring models to individual datasets' which always entails 'an element of subjectivity which will be difficult to eliminate' (Rob Kempton, in discussion of Verbyla et al., 1999). A natural starting point for the placement of knots are the plot locations, and this is our general recommendation. Our results for the barley and wheat data suggest that there is little gain from trying other options. We also found little difference in goodness-of-fit for different choices of the degree of the B-spline basis, with the simplest choice of  $q = 1$  (implying the simplification  $B = I$ ) working quite well, corresponding to the classical NN modelling with first or second differences for  $p = 1$  and  $p = 2$ , respectively. The examples further showed that first differences seemed to be sufficient for smoothing the spatial trend. Certainly, more empirical experience is needed to further provide guidance for routine use of the framework. We do conjecture based on our experience so far that a viable recommendation is to stick with the most parsimonious options, such as  $p = 1$ ,  $q = 1$ ,  $i_r = k$ , and  $i_c = s$ . These models are particularly easy to fit. This option is also very closely related to earlier approaches to NN analysis of field trials and so will be easy to communicate to

researchers wanting to use P-splines. For practical implementation, one may either use the specialized SpATS R package, which implements a subset of the models considered in this paper, or a general linear mixed model package such as SAS, for which sample code is provided in the supplemental material.

Careful consideration needs to be given to the null space of the interaction smooth, especially for the models described in Section 4.1. As detailed there, this null space can have a rather larger number of dimensions for  $p = 2$  than for  $p = 1$  when the penalty in (3) is used. If one insists on representing this null space by fixed effects in order to make the fit invariant to the choice of g-inverse for the singular precision matrix  $P_{rc}$ , efficiency of the analysis may be adversely affected, and it is therefore worthwhile to fit these effects as random. In particular, the P-spline framework allows smoothing these terms, combining them with the marginal smooth for rows and columns as prescribed by the ANOVA-type approach to spline smoothing (Wood et al., 2013; Wood, 2017). It must be acknowledged, however, that this sacrifices the invariance to the choice of g-inverse for  $P_{rc}$ . We note in this context that the  $LV \otimes LV$  model gives the same fit as the P-spline approach for  $p = 1$ ,  $q = 1$ ,  $i_r = k$ , and  $i_c = s$ , when the null space is modelled by fixed effects for rows and columns, but not when these effects are modelled as random (Boer et al., 2020), or are smoothed (Tables 3 and 5). An advantage of the IAR model with the penalty in (7) is the small size of the null space for the interaction smooth, even for  $p = 2$ , with a simple fixed-effects regression on row and column numbers to cover that null space. Thus, there is no confounding with the marginal smooths and so these can be considered in their own right, there being no strict need to fit them. Care is needed with the interaction smooth, however, when  $\alpha$  converges to either 0 or 1 (this happened with both examples), because in that event the null space changes, increasing considerably, which may be seen as a disadvantage of the IAR model. One way out is to constrain the grid search so that  $\alpha$  stays clear of the lower and upper bound, or use specialized procedures such as the SAP algorithm outside standard mixed model packages (Rodríguez-Álvarez, 2015), and implemented in the SpATS R-package.

We have included fixed-effects regression terms for row and column numbers and their interaction in all models, even those that do not involve P-splines, and also for P-splines with first-difference penalty, where these terms are not needed to represent the null space of the smooth. This was done so we could use the REML-based likelihood for comparing all models on the same basis. These terms only cost three degrees of freedom in the fixed part of the model, and it may therefore be worth including them in their own right in order to capture global trends, even if they are not needed to represent the null space of the smooth terms.

A general advantage of the P-spline framework is that most of the variance-covariance structures are linear in the parameters, which makes them easy to fit and minimizes convergence problems which may befit spatial variance-covariance models that are nonlinear in the parameters (Piepho et al., 2015; Velazco et al., 2017). This plays out most when the random terms are all independent, i.e. the smooth terms correspond to a variance-components model with no covariances among random effects. As we have shown, when  $p > 1$ , the marginal smooth associated with the penalty (3) for the interaction may require allowing a covariance among random intercept and slope terms, a key result that seems to have been overlooked by others, and this can be difficult to fit on account of the need to ensure positive definiteness of the variance-covariance structure. Such difficulties pose an obstacle for routine use, and they speak in favour of the simpler P-spline options based on first differences. In this regard, it is interesting to observe that the  $AR1 \otimes AR1$  model with a nugget variance (M56), which fitted quite well, had relatively large correlations along rows and columns (0.8598 and

0.8952 for the barley data and 0.7488 and 0.9064 for the wheat data), in which case the covariance is well approximated by a linear variance model (Williams, 1986), corresponding to the simpler P-spline for  $p = 1$ ,  $q = 1$ ,  $i_r = k$ , and  $i_c = s$ .

On a final note we would like to stress the importance of design, which often does not receive the attention it deserves. Sometimes, the large number of modelling options for spatial analysis may raise the false impression that design does not matter, and that a sophisticated analysis takes care of everything. Nothing could be further from the truth. Certainly, given the large number of options for analysis, the design issue becomes more challenging and is far from resolved. Rather than focusing the design on a specific route of analysis, one may consider searching for a good design without a specific analysis in mind (Piepho et al., 2021), striving for robustness to model choice.

## CRediT authorship contribution statement

**Hans-Peter Piepho:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Martin P. Boer:** Conceptualization, Methodology, Formal analysis, Writing - review and editing. **Emlyn R. Williams:** Conceptualization, Methodology, Formal analysis, Writing - review and editing.

## References

- Bartlett, M.S., 1978. Nearest neighbour models in the analysis of field experiments. *Journal of the Royal Society B* 40, 147–174.
- Besag, J., Higdon, D., 1999. Bayesian analysis of agricultural field experiments (with discussion). *Journal of the Royal Society B* 61, 691–746.
- Besag, J., Kempton, R., 1986 Statistical analysis of field experiments using neighbouring plots. *Biometrics* 42, 231–251.
- Besag, J., Kooperberg, C., 1995. On conditional and intrinsic autoregressions. *Biometrika* 82, 733–746.
- Boer, M.P., Piepho, H.P., Williams, E.R., 2020. Linear variance, P-splines and neighbour differences for spatial adjustment in field trials, How are they related? *Journal of Agricultural, Biological and Environmental Statistics* 25, 676–698.
- Currie, I.D., Durban, M., 2002. Flexible smoothing with P-splines: a unified approach. *Statistical Modelling* 4, 333–349.
- Durbin, J., Koopman, S.J., 2001. Time series analysis by state space methods. Oxford University Press, Oxford.
- Durbán, M., Hackett, C.A., McNicol, W., Thomas, T.B., Currie, I.D., 2003. The practical use of semiparametric models in field trials. *Journal of Agricultural, Biological and Environmental Statistics* 8, 48–66.
- Dutta, S., Mondal, D., 2015. An h-likelihood method for spatial mixed linear models based on intrinsic auto-regressions. *Journal of the Royal Statistical Society B* 77, 699–726.
- Eilers, P.H.C., 2003 A perfect smoother. *Analytical Chemistry* 75, 3631–3636.
- Eilers, P.H.C., Marx, B.D., 1996. Flexible smoothing with B-splines and penalties. *Statistical Science* 11, 89–121.
- Eilers, P.H.C., Marx, B.D., 2003. Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Laboratory Systems* 66, 159–174.

- Gilmour, A.R., Cullis, B.R., Verbyla, A.P., 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., Jennison, C., Seheult, A., 1985. Analysis of field experiments by least squares smoothing. *Journal of the Royal Statistical Society B* 47, 299–315.
- Harvey, A.C., 1989. Forecasting, structural time series models and the Kalman filter. Cambridge University Press, Cambridge.
- Harville, D.A., 1997. Matrix algebra from a statistician's perspective. Springer, New York.
- Kempton, R.A., Howes, C.W., 1981. The use of neighbouring plot values in the analysis of variety trials. *Applied Statistics* 30, 59–70.
- Kempton, R.A., Seraphin, J.C., Sword, A.M., 1994. Statistical analysis of two-dimensional variation in variety yield trials. *Journal of Agricultural Science Cambridge* 122, 335–342.
- Lee, D.-J., Durbán, M., 2011. P-spline ANOVA-type interaction models for spatio-temporal smoothing. *Statistical Modelling* 11, 49–69.
- Lee, D.-J., Durbán, M., Eilers, P., 2013. Efficient two-dimensional smoothing with P-spline ANOVA mixed models and nested bases. *Computational Statistics and Data Analysis* 61, 22–37.
- 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, W., Piepho, H.P., Lee, Y., 2021. Resolving the ambiguity of random-effects models with singular precision matrix. *Statistica Neerlandica* <https://onlinelibrary.wiley.com/doi/abs/10.1111/stan.12244>
- Longford, N.T., 1993. Random coefficient models. Oxford University Press, Oxford.
- Mao, X., Dutta, S., Wong, R.K.W., Nettleton, D., 2020. Adjusting for spatial effects in genomic prediction. *Journal of Agricultural, Biological and Environmental Statistics* 25, 699–718.
- McCullagh, P., Clifford, D., 2006. Evidence of conformal invariance of crop yields. *Proceedings of the Royal Society A* 462, 2119–2143.
- Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.
- Papadakis, J.S., 1937. Méthode statistique pour des expériences sur champ. *Bulletin de l'Institut d'Amélioration des Plantes à Salonique* 23.
- Patterson, H.D., Thompson, R., 1971. Recovery of inter-block information when block sizes are unequal. *Biometrika* 58, 545–554.
- Piepho, H.P., Möhring, J., Pflugfelder, M., Hermann, W., Williams, E.R., 2015. Problems in the parameter estimation for power and AR1. models of spatial correlation in designed field experiments. *Communications in Biometry and Crop Science* 10, 3–16.
- Piepho, H.P., Ogutu, J.O., 2007. Simple state-space models in a mixed model framework. *The American Statistician* 61, 224–232.
- Piepho, H.P., Williams, E.R., 2010. Linear variance models for plant breeding trials. *Plant Breeding* 129, 1–8.
- Piepho, H.P., Williams, E.R., Michel, V., 2021. Generating row-column field experimental designs with good neighbour balance and even distribution of treatment replications. *Journal of Agronomy and Crop Science* 207.
- Rodríguez-Álvarez, M.X., Lee, D.-J., Kneib, T., Durbán, M., Eilers, P.H.C., 2015. Fast smoothing parameter separation in multidimensional generalized P-splines, the SAP algorithm. *Statistical Computing* 25, 941–957.
- Rodríguez-Álvarez, M.X., Boer, M.P., van Eeuwijk, F.A., Eilers, P.H.C., 2018. Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics* 23, 52–71.

- Ruppert, D., Wand, M.P., Carroll, R.J., 2003. Semiparametric regression. Cambridge University Press, Cambridge.
- Stroup, W.W., 2002. Power analysis based on spatial effects mixed models, A tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological and Environmental Statistics* 7, 491–511.
- Stroup, W.W., Baenziger, P.S., Mulitze, D.K., 1994. Removing spatial variation from wheat yield trials, a comparison of methods. *Crop Science* 34, 62–66.
- Tsimikas, J.V., Ledolter, J., 1997. Mixed model representation of state-space models, New smoothing results and their application to REML estimation. *Statistica Sinica* 7, 973–991.
- Velazco, J.G., Rodríguez-Álvarez, M.X., Boer, M.P., Jordan, D.R., Eilers, P.H.C., Malosetti, M., van Eeuwijk, F.A., 2017. Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theoretical and Applied Genetics* 130, 1375–1392.
- Verbyla, A.P., Cullis, B.R., Kenward, M.G., Welham, S.J., 1999. The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Applied Statistics* 48, 269–311.
- Verbyla, A.P., De Faveri, J., Wilkie, J.D., Lewis, T., 2018. Tensor cubic smoothing splines in designed experiments requiring residual modelling. *Journal of Agricultural, Biological and Environmental Statistics* 23, 478–508.
- Wand, M.P., Omerod, J.T., 2008. On semiparametric regression with O'Sullivan penalized splines. *Australian and New Zealand Journal of Statistics* 50, 179–198.
- Wilkinson, G.N., Eckert, S.R., Hancock, T.W., Mayo, O., 1983. Nearest neighbour (NN) analysis of field experiments (with discussion). *Journal of the Royal Statistical Society B* 45, 151–211.
- Williams, E.R., 1986. A neighbour model for field experiments. *Biometrika* 73, 279–287.
- Wolfinger, R.D., 1996. Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological and Environmental Statistics* 1, 205–230.
- Wood, S.N., 2017. Generalized additive models, An introduction with R. Second edition. Chapman & Hall/CRC Press, Boca Raton.
- Wood, S.N., Scheipl, F., Faraway, J.J., 2013. Straightforward intermediate rank tensor product smoothing in mixed models. *Statistical Computing* 23, 341–360.
- Wright, K., 2021. agridat: Agricultural Datasets. R package version 1.18.

## Appendices

### A. Row-wise Kronecker products

Eilers and Marx (2003) proposed the use of row-wise Kronecker products, also known as box products or tensor products of B-spline bases for non-additive smoothing. A salient feature of field trials, however, is that plots are usually placed on a regular grid, for which the tensor product coincides with the usual Kronecker product (Lee and Durbán, 2011). Thus, if  $\otimes_r$  denotes the row-wise Kronecker product and  $\otimes$  denotes the ordinary Kronecker product, then for a B-spline basis matrix  $B_r$  with  $k$  rows and arbitrary number of columns corresponding to regularly spaced knots and a B-spline basis matrix  $B_c$  with  $s$  rows and an arbitrary number of columns (knots), we have

$$B_{rc} = (B_r \otimes 1_s) \otimes_r (1_k \otimes B_c) = B_r \otimes B_c. \quad (\text{A1})$$



This important fact can hardly be over-emphasized for those wanting to use P-splines for analysis of spatial data on a regular grid in general and on a regular grid of field trial plots in particular, but drawing on a literature that seeks to embrace irregularly gridded data as the dominant application case, e.g. in longitudinal studies. In the case of field trials this means that a row-wise Kronecker product can be written as an ordinary Kronecker product of matrices associated with rows and columns (Verbyla et al., 2018), which greatly simplifies the notation and also clarifies the relationship with other spatial modelling approaches for field trials, such as the LV $\otimes$ LV model (Williams and Piepho, 2010) and the AR1 $\otimes$ AR1 model (Gilmour et al., 1997). We make use of the important equation (A1) as a design matrix throughout this paper to model spatial interactions between rows and columns. Readers less familiar with P-splines should be aware that the simplified notation arising from (A1) may look somewhat unfamiliar compared to results for P-splines applicable to irregular spatial grids (Lee and Durbán, 2011; Wood, 2017, p. 232), even though the regular case is just a special version of the irregular one.

## B. Relations with state-space models

The use of differences among neighbouring plots (Besag and Kempton, 1986) has close ties with state-space models (Harvey, 1989; Durbin and Koopman, 2001), which in turn can also be represented as mixed models (Tsimikas and Ledolter, 1997; Piepho and Ogutu, 2007). Specifically, we may assume that the response  $y_i$  on the  $i$ -th plot is related to a latent level  $\mu_i$  according to  $y_i = \mu_i + e_i$ , where  $e_i \sim N(0, \sigma_e^2)$ . If we further assume that the state on the neighbouring  $(i+1)$ -th plot is related to that on the  $i$ -th plot by  $\mu_{i+1} = \mu_i + u_i$ , where  $u_i \sim N(0, \sigma_u^2)$ , then  $\mu_i$  follows a *first-order random walk* model, which is a simple form of state-space model. For first differences we then have  $\mu_{i+1} - \mu_i = u_i \sim N(0, \sigma_u^2)$ , demonstrating the connection between the use of first differences and state-space models as used mostly for time series. By comparison, the AR1 model is known as damped model in state-space modelling and written as  $\mu_{i+1} = \rho\mu_i + u_i$  with  $u_i \sim N(0, \sigma_u^2)$ , where  $0 < \rho < 1$  is the autocorrelation (Harvey, 1989, p. 46).

The first-order random walk model can be extended to what is known in state-space modelling as *local linear trend* model. The starting point is a simple linear regression of the form  $\mu_i = \beta_0 + \beta_1 i$ . One idea to make this deterministic trend stochastic is to let both  $\beta_0$  and  $\beta_1$  vary randomly, but this would lead to discontinuities at the plots (Harvey, 1989, p. 37). A better option is to work directly with the current level,  $\mu_i$  in place of the intercept  $\beta_0$ . Noting that the deterministic trend can be rewritten as  $\mu_{i+1} = \mu_i + \beta_1$  with  $\mu_0 = \beta_0$ , we may consider updating both the current level and the slope term by a random walk as follows:

$$\mu_{i+1} = \mu_i + b_i + u_{0i}, \quad u_{0i} \sim N(0, \sigma_{u0}^2) \quad (B1)$$

$$b_{i+1} = b_i + u_{1i}, \quad u_{1i} \sim N(0, \sigma_{u1}^2) \quad (B2)$$

where  $b_i$  is the random slope at the  $i$ -th plot with  $b_0 = \beta_1$ . Thus, on the  $i$ -th plot we have linear trend  $b_i$ , and this is updated by a random increment  $u_{1i}$ , i.e. by a random walk, as we move to the  $(i+1)$ -th plot. The intercept at the  $i$ -th plot is  $a_i = \mu_i - b_i i$ , and hence  $a_{i+1} = \mu_i - (1-i)b_i + u_{0i}$ . Essentially, we have randomly varying regression lines, pieced

together at the plots. Due to the update  $u_{0i}$ , however, there are discontinuities at the plots (Figure 1). In the limiting case when  $\sigma_{u0}^2 = 0$ , the trend is a polygon with breaks at the plots but no discontinuities. Interestingly, the model then is equivalent to  $\mu_{i+1} - 2\mu_i + \mu_{i-1} = u_{1i}$ , which is a *second-order random walk* and involves second differences (Durbin and Koopman, 2001, p. 39). To see this, we may subtract the state equation  $0 = -\mu_i + \mu_{i-1} + b_{i-1}$  from  $\mu_{i+1} = \mu_i + b_i$ , yielding  $\mu_{i+1} = 2\mu_i - \mu_{i-1} + b_i - b_{i-1} = 2\mu_i - \mu_{i-1} + u_{1,i-1}$ . By comparison, the first-order random walk model for  $\mu_i$  can be depicted as horizontal regression lines with jumps at the plots (Figure B1). It is worth observing that both the local linear trend and its limiting case with  $\sigma_{u0}^2 = 0$ , which is equivalent to use of second differences, have the fixed trend component  $\beta_0 + \beta_1 i$ , corresponding to the null space of P-splines with second differences penalty. Thus, the state-space view provides a very natural explanation why a fixed intercept as well as a fixed slope are needed with second differences.

- Figure B1 about here -

### C. Two-dimensional first differences

Kempton et al. (1994) formulate a correlation structure that is separable between rows and columns. Their assumption is that first differences along rows and along variances have variances  $\phi_r = \psi_r \sigma^2$  and  $\phi_c = \psi_c \sigma^2$ , respectively, and that plot differences calculated across both rows and columns, i.e.,  $v_{i,j} - v_{i+1,j} - v_{j,i+1} + v_{i+1,j+1}$ , where  $v_{i,j} - v_{i+1,j} - v_{j,i+1} + v_{i+1,j+1}$  is the plot value in the  $i$ -th row and  $j$ -th column, have variance  $\phi_{rc} = (2 + \psi_r)(2 + \psi_c)\sigma^2$ . In what follows, we will relate this model to the LV $\otimes$ LV model (Piepho and Williams, 2010) and P-splines with  $p = 1$ ,  $q = 1$ ,  $i_r = k$ , and  $i_c = s$ . We may collect plot values for spatial trend into a random vector  $v_{rc}$ . Following Kempton et al. (1994), the first differences across rows can then be assumed to have variance

$$\text{var}[(D_r \otimes I_s)v_{rc}] = I_{k-p} \otimes A_s \phi_r \quad (C1)$$

for some  $s \times s$  correlation matrix  $A_s$ . For first differences across columns, we may postulate accordingly that

$$\text{var}[(I_k \otimes D_c)v_{rc}] = A_k \otimes I_{s-p} \phi_c \quad (C2)$$

for some  $k \times k$  correlation matrix  $A_k$ , and for differences across both rows and columns

$$\text{var}[(D_r \otimes D_c)v_{rc}] = I_{k-p} \otimes I_{s-p} \phi_{rc} \quad (C3)$$

In line with our P-spline approach, the variance-covariance matrix for  $v_{rc}$  may be assumed to take the form

$$\text{var}(v_{rc}) = F_r \otimes J_s \sigma_r^2 + J_k \otimes F_c \sigma_c^2 + F_r \otimes F_c \sigma_{rc}^2 \quad (C4)$$

for suitable variance-covariance matrices  $F_r$  and  $F_c$ . This can also be expressed in terms an ANOVA-type decomposition

$$v_{rc} = u_r \otimes 1_s + 1_k \otimes u_c + u_{rc} \quad (C5)$$

with  $\text{var}(u_r) = F_r \sigma_r^2$ ,  $\text{var}(u_c) = F_c \sigma_c^2$  and  $\text{var}(u_{rc}) = F_r \otimes F_c \sigma_{rc}^2$ . We now need to find  $F_r$  and  $F_c$  such that  $D_r F_r D_r^T = I_{k-p}$  and  $D_c F_c D_c^T = I_{s-p}$ . For  $F_r$ , e.g., we can meet these assumptions using  $F_r = P_r^+$  or  $F_r = U_{0r} U_{0r}^T + P_r^+$ , corresponding to our P-spline approach, but this does not yield the desired form of  $A_s$ . One (but not the only) choice that will yield the desired form of  $A_s$  is  $2F_r = M_k = J_k(k-1) - L_k$ , where  $L_k$  is a  $k \times k$  matrix with  $(i_1, i_2)$ -th element equal to  $|i_1 - i_2|$ , corresponding to the LV $\otimes$ LV model (Piepho and Williams, 2010). To see this, we first use (C4) to find

$$\begin{aligned} \text{var}[(D_r \otimes I_s)v_{rc}] &= (D_r \otimes I_s)(F_r \otimes J_s \sigma_r^2 + J_k \otimes F_c \sigma_c^2 + F_r \otimes F_c \sigma_{rc}^2)(D_r^T \otimes I_s) = \\ &D_r F_r D_r^T \otimes J_s \sigma_r^2 + D_r F_r D_r^T \otimes F_c \sigma_{rc}^2 = I_{k-p} \otimes (J_s \sigma_r^2 + F_c \sigma_{rc}^2) \end{aligned} \quad (C6)$$

Comparing coefficients with (C1), we find  $A_s \phi_r = J_s \sigma_r^2 + F_c \sigma_{rc}^2$ . Then using  $2F_r = M_k$ , we find that the diagonal elements of  $A_s \phi_r$  equal  $\sigma_r^2 + \frac{1}{2}(k-1)\sigma_{rc}^2$ , and hence

$$\phi_r = \sigma_r^2 + \frac{1}{2}(k-1)\sigma_{rc}^2. \text{ Similarly, } \phi_c = \sigma_c^2 + \frac{1}{2}(s-1)\sigma_{rc}^2. \text{ Moreover,}$$

$$\text{var}[(D_r \otimes D_c)v_{rc}] = D_r F_r D_r^T \otimes D_c F_c D_c^T \sigma_{rc}^2, \text{ from which } \phi_{rc} = \frac{1}{4}(k-1)(s-1)\sigma_{rc}^2.$$

#### ***D. Orthogonal rotations of the null-space eigenvectors for $p = 2$***

It is interesting to observe the connection of a diagonal variance-covariance model with rotation parameter with an unstructured model. Let  $\Omega$  be a diagonal matrix

$$\Omega = \begin{pmatrix} \omega_1 & 0 \\ 0 & \omega_2 \end{pmatrix}. \quad (D1)$$

Then using  $U_0 = (a_m : b_m)R(\varphi)$  [see eqs. (5) and (6)] we have

$$U_0 \Omega U_0^T = (a_m : b_m) \Sigma (a_m : b_m)^T \quad (D2)$$

where

$$\Sigma = R(\varphi) \Omega R^T(\varphi) = \begin{pmatrix} (\omega_1 - \omega_2) \cos^2(\varphi) + \omega_2 & \frac{1}{2} \sin(2\varphi)(\omega_1 - \omega_2) \\ \frac{1}{2} \sin(2\varphi)(\omega_1 - \omega_2) & -(\omega_1 - \omega_2) \cos^2(\varphi) + \omega_1 \end{pmatrix} \quad (D3)$$

If  $\omega_1 = \omega_2 = \omega$ , we have  $\Sigma = \Omega = \omega I_2$ , showing that if we chose the identity matrix, the eigenvectors can be chosen arbitrarily, i.e. the model does not depend on the rotation parameter  $\varphi$ .

**Table 1:** Notation for matrices and vectors used for P-splines

Matrix <sup>§</sup>	Description	Dimensions <sup>§</sup>
$h_r$	Row coordinates of plots	$k \times 1$
$B_r$	$q$ -th degree B-spline bases for $h_r$	$k \times m_r$
$D_r$	Differencing matrix	$(m_r - p) \times m_r$
$P_r$	Penalty matrix; $P_r = D_r^T D_r$	$m_r \times m_r$
$U_r$	All eigenvectors of $P_r$	$m_r \times m_r$
$U_{+r}$	Eigenvectors of $P_r$ with positive eigenvalues	$m_r \times (m_r - p)$
$U_{0r}$	Eigenvectors of $P_r$ with zero eigenvalues	$m_r \times p$
$d_r$	Vector of all eigenvalues of $P_r$	$m_r \times 1$
$d_{+r}$	Vector of positive eigenvalues of $P_r$	$(m_r - p) \times 1$
$i_r$	Number of interior knots	$1 \times 1$
$m_r$	Number of B-spline bases <sup>%</sup>	$1 \times 1$

§ The subscript  $r$  refers to rows. For columns, replace subscript  $r$  with  $c$  (for columns).

§  $k$  = number of rows;  $p$  = order of differences. For columns, replace  $k$  with  $s$  (the number of columns) and  $m_r$  with  $m_c$  (the number of B-spline bases used for columns).

%  $m_r = i_r + q - 1$ , where  $q$  is the degree of the B-spline bases;  $m_r = k + q - 1$  when the interior knots are placed at the plots. For columns, replace  $m_r$  with  $m_c$  (the number of B-spline bases used for columns).



**Table 2:** Analysis of barley data of Durbán et al. (2003) and wheat data of Stroup et al. (1994) using P-spline approach with  $p = 2$  (second differences),  $i_r = k$ ,  $i_c = s$ , and the penalty in (3) for  $B_{rc}$ . All models have fixed effects for replicates, genotypes, row numbers  $h_r$ , column numbers  $h_c$  and their product  $h_r h_c$ .

Model	$\text{var}(u_r)^s$	$X_c^s$	$\text{var}(u_c)^s$	$X_r^s$	$\text{var}(u_{rc})$	Barley data				Wheat data			
						$q = 1$		$q = 3$		$q = 1$		$q = 3$	
						Deviance	AIC	Deviance	AIC	Deviance	AIC	Deviance	AIC
M1	1	$1_s$	1	$1_k$	-	295.22	301.22	295.70	301.70	1076.03	1082.03	1075.93	1081.93
M2	1	$1_s$	1	$1_k$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	292.20	300.20	293.09	301.09	1075.78	1083.78	1071.88	1079.88
M3	$\Sigma$	$\ddot{X}_c$	-	-	-	378.69	386.69	378.56	386.56	1075.45	1083.45	1075.51	1083.51
M4	$\Sigma$	$1_s : h_s$	-	-	-	378.69	386.69	378.56	386.56	%	%	%	%
M5	$\Sigma$	$a_s : b_s$	-	-	-	378.69	386.69	378.56	386.56	%	%	%	%
M6	$\Omega$	$\widehat{X}_c$	-	-	-	378.72	384.72	378.62	384.62	1078.49	1084.49	1075.15	1079.15
M7	$\Omega$	$1_s : h_s$	-	-	-	381.95	387.95	383.41	389.41	1082.32	1088.32	1081.92	1087.92
M8	$\Omega$	$a_s : b_s$	-	-	-	378.87	384.87	378.72	384.72	1081.76	1087.76	1081.08	1087.08
M9	$I_2$	$\widehat{X}_c$	$I_2$	$\widehat{X}_r$	-	296.46	302.46	296.56	302.56	1058.02	1064.02	1060.05	1066.05
M10	$I_2$	$1_s : h_s$	$I_2$	$1_k : h_k$	-	317.58	323.58	316.77	322.77	1087.29	1093.29	1082.64	1088.64
M11	$I_2$	$a_s : b_s$	$I_2$	$a_k : b_k$	-	296.45	302.45	297.13	303.13	1058.02	1064.02	1059.84	1065.84
M12	$\Omega$	$\widehat{X}_c$	$\Omega$	$\widehat{X}_r$	-	296.41	306.41	295.23	305.23	1049.91	1059.91	1057.51	1067.51
M13	$\Omega$	$1_s : h_s$	$\Omega$	$1_k : h_k$	-	291.15	301.15	291.52	301.52	1057.10	1067.10	1057.34	1067.34
M14	$\Omega$	$a_s : b_s$	$\Omega$	$a_k : b_k$	-	287.88	297.88	288.27	298.27	1057.45	1067.45	1059.24	1069.24
M15	$\Sigma$	$\ddot{X}_c$	$\Sigma$	$\ddot{X}_r$	-	283.65	297.65	284.12	298.12	1043.37	1057.37	1044.22	1058.22
M16	$\Sigma$	$1_s : h_s$	$\Sigma$	$1_k : h_k$	-	%	%	%	%	%	%	%	%
M17	$\Sigma$	$a_s : b_s$	$\Sigma$	$a_k : b_k$	-	%	%	%	%	%	%	%	%
M18	$I_2$	$\widehat{X}_c$	$I_2$	$\widehat{X}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	293.23	301.23	293.60	301.60	1058.02	1066.02	1060.05	1068.05

M19	$\Omega$	$\widehat{X}_c$	$\Omega$	$\widehat{X}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	293.16	305.16	292.28	304.28	1049.91	1061.91	1057.49	1067.49
M20	$\Sigma$	$\ddot{X}_c$	$\Sigma$	$\ddot{X}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	280.12	296.12	280.27	296.27	1042.80	1058.80	1044.22	1058.22

\$ 1:  $P_r^+ \otimes \sigma_r^2$  or  $\sigma_c^2 \otimes P_c^+$ ;  $I_2$ :  $P_r^+ \otimes \sigma_r^2 I_2$  or  $\sigma_c^2 I_2 \otimes P_c^+$ ;  $\Omega$ :  $P_r^+ \otimes \Omega_r$  or  $\Omega_c \otimes P_c^+$ ;  $\Sigma$ :  $P_r^+ \otimes \Sigma_r$  or  $\Sigma_c \otimes P_c^+$

§ Representation of  $X_r$  or  $X_c$  in the marginal smooth;  $\widehat{X}$ : obtained using  $U_0$  of spectral decomposition of  $P$  as obtained by software (IML procedure of SAS);  $\ddot{X}$ : obtained using near optimal orthogonal rotation (Figures 2 and 4)

% Did not converge due to poor scaling of  $X_r$  and/or  $X_c$

**Table 3:** Analysis of barley data of Durbán et al. (2003) and wheat data of Stroup et al. (1994) using P-spline approach with  $p = 1$  (first differences) and the penalty in (3) for  $B_{rc}$ . All models have fixed effects for replicates, genotypes, row numbers  $h_r$ , column numbers  $h_c$  and their product  $h_r h_c$ . The marginal smooths use  $\text{var}(u_r) = P_r^+ \otimes \sigma_r^2$ ,  $X_c = 1_s$ ,  $\text{var}(u_c) = \sigma_c^2 \otimes P_c^+$  and  $X_r = 1_k$  for all models.

Model	$q$	$i_r$	$i_c$	$\text{var}(u_{rc})$	Barley data		Wheat data	
					Deviance	AIC	Deviance	AIC
M21	3	$k$	$s$	-	293.37	299.37	1072.47	1078.47
M22	3	$k$	$s$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	279.28	287.28	1046.40	1054.40
M23	2	$k$	$s$	-	293.56	299.56	1071.91	1077.91
M24	2	$k$	$s$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	279.18	287.18	1046.06	1054.06
M25	1	$k$	$s$	-	295.78	301.78	1075.14	1081.14
M26	1	$k$	$s$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	278.45	286.45	1047.12	1055.12
M27	3	10	20	-	292.92	298.92	1073.38	1079.38
M28	3	10	20	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	281.18	289.18	1056.41	1064.41
M29	2	10	20	-	291.61	297.61	1073.27	1079.27
M30	2	10	20	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	279.74	287.74	1057.09	1065.09
M31	1	10	20	-	296.75	302.75	1075.43	1081.43
M32	1	10	20	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	281.31	289.31	1052.41	1060.41

**Table 4:** Analysis of barley data of Durbán et al. (2003) and wheat data of Stroup et al. (1994) using the IAR model with the penalty in (7) for  $u_{rc}$  with  $i_r = k$  and  $i_c = s$  with and without marginal smooths added. All models have fixed effects for replicates, genotypes, row numbers  $h_r$ , column numbers  $h_c$  and their product  $h_r h_c$ . Fits obtained by a grid search over  $\alpha = 0, 1(0.1)$ .

Model	$q$	$p$	Marginal smooth <sup>§</sup>	Barley data			Wheat data		
				$\alpha$	Deviance	AIC	$\alpha$	Deviance	AIC
M33	3	1	-	0.5	284.53	290.53	0.9	1052.33	1058.33
M34	2	1	-	0.5	283.94	289.94	0.9	1053.18	1059.18
M35	1	1	-	0.6	281.52	287.52	0.8	1055.06	1061.06
M36	3	2	-	0.6	295.90	301.90	0.9	1063.35	1069.35
M37	2	2	-	0.6	296.58	302.58	0.9	1064.66	1070.66
M38	1	2	-	0.6	297.00	303.00	0.9	1064.94	1070.94
M39	3	1	1	0.9	274.84	284.84	0.9	1050.84	1058.84
M40	2	1	1	0.9	274.85	284.85	1.0	1050.61	1060.61
M41	1	1	1	0.9	273.81	283.81	0.9	1053.08	1063.08
M42	3	2	1	1.0	284.63	294.63	0.9	1059.55	1069.55
M43	2	2	1	1.0	284.79	294.79	0.9	1060.54	1070.54
M44	1	2	1	1.0	284.94	294.94	0.9	1061.39	1071.39
M45 <sup>%</sup>	3	2	$\Omega$	0.5	283.58	295.58	0.5	1063.57	1077.57
M46 <sup>§</sup>	3	2	$\Omega$	0.5	290.92	302.92	0.5	1052.70	1066.70
M47	3	2	$\Omega$	1.0	285.24	297.24	1.0	1049.40	1063.40
M48	2	2	$\Omega$	1.0	287.06	299.06	1.0	1044.35	1058.35
M49	1	2	$\Omega$	1.0	287.20	299.20	1.0	1046.99	1060.99
M50	3	2	$\Sigma$	1.0	273.67	289.67	1.0	1044.28	1062.28
M51	2	2	$\Sigma$	1.0	274.77	290.77	1.0	1041.51	1059.51
M52	1	2	$\Sigma$	1.0	273.51	289.51	1.0	1038.81	1056.81

§ 1: for  $p = 1$  just the intercepts for rows and columns.  $\Omega$ : for  $p = 2$  using  $P_r^+ \otimes \Omega_r$  and  $\Omega_c \otimes P_c^+$  with  $\hat{X}_c$  and  $\hat{X}_r$ , respectively.  $\Sigma$ : using  $P_r^+ \otimes \Sigma_r$  and  $\Sigma_c \otimes P_c^+$  with  $\ddot{X}_c$  and  $\ddot{X}_r$ , respectively;  $\hat{X}$ : obtained using  $U_0$  of spectral decomposition of  $P$  as obtained by software (IML procedure of SAS);  $\ddot{X}$ : obtained using near optimal orthogonal rotation (Figures 2 and 4)

% Fit for fixed  $\alpha = 0.5$ ; using  $\hat{X}_c$  and  $\hat{X}_r$  for  $X_c$  and  $X_r$ , respectively.

§ Fit for fixed  $\alpha = 0.5$ ; using  $a_s : b_s$  and  $a_k : b_k$  for  $X_c$  and  $X_r$ , respectively (Rodríguez-Álvarez et al. 2018).

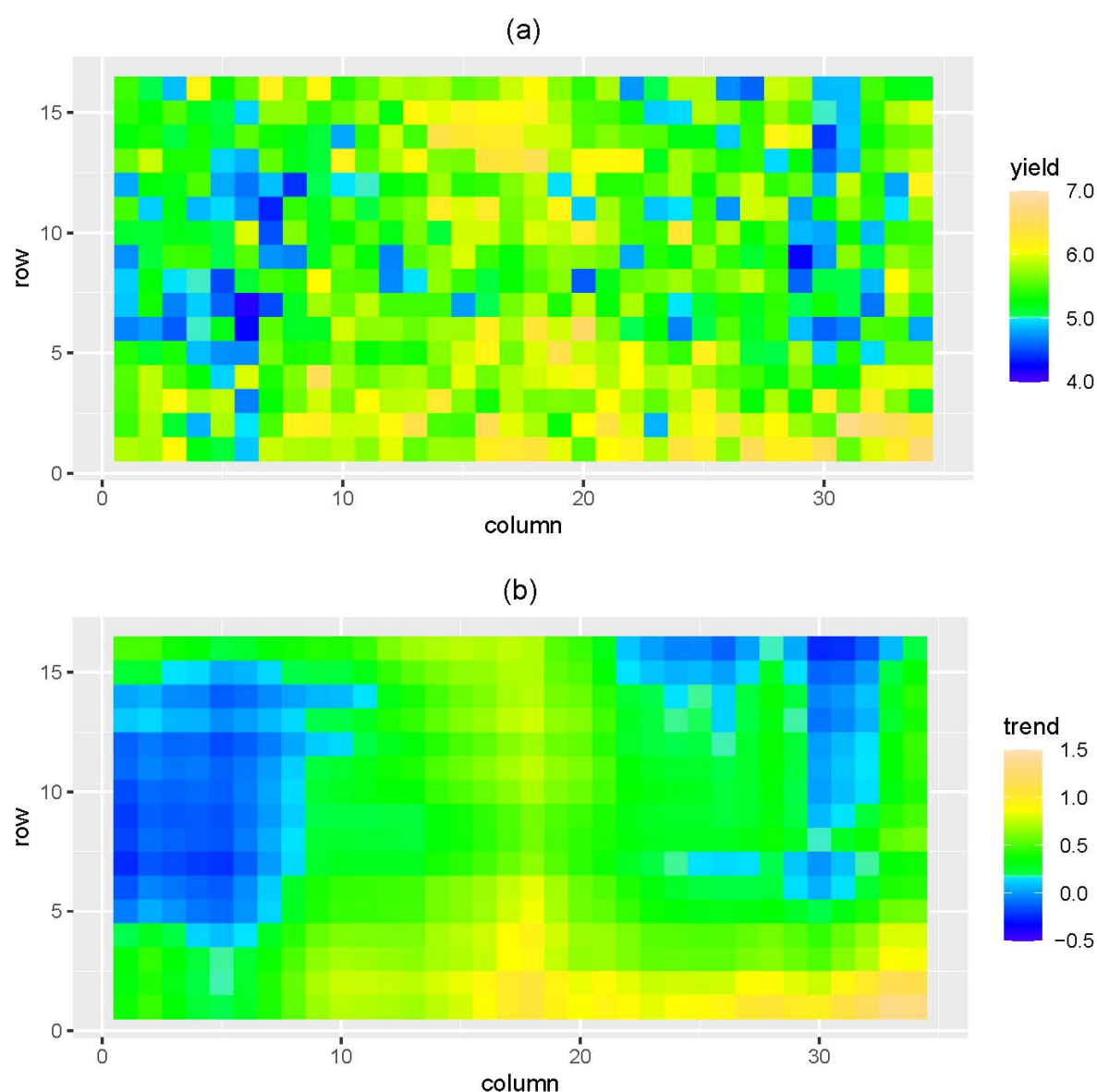
**Table 5:** Analysis of barley data of Durbán et al. (2003) and wheat data of Stroup et al. (1994) using other common models. All models have fixed effects for replicates, genotypes, row numbers  $h_r$ , column numbers  $h_c$  and their product  $h_r h_c$ .

Model	Description	Barley data		Wheat data	
		Deviance	<i>AIC</i>	Deviance	<i>AIC</i>
M53	Baseline <sup>§</sup>	410.19	412.19	1101.53	1103.53
M54	Baseline + row & column <sup>§</sup>	352.40	358.40	1083.52	1089.52
M55	AR1 $\otimes$ AR1	299.60	305.60	1067.32	1073.32
M56	AR1 $\otimes$ AR1 + nugget	278.20	286.20	1050.34	1058.34
M57	LV $\otimes$ LV	283.71	291.71	1051.33	1059.33

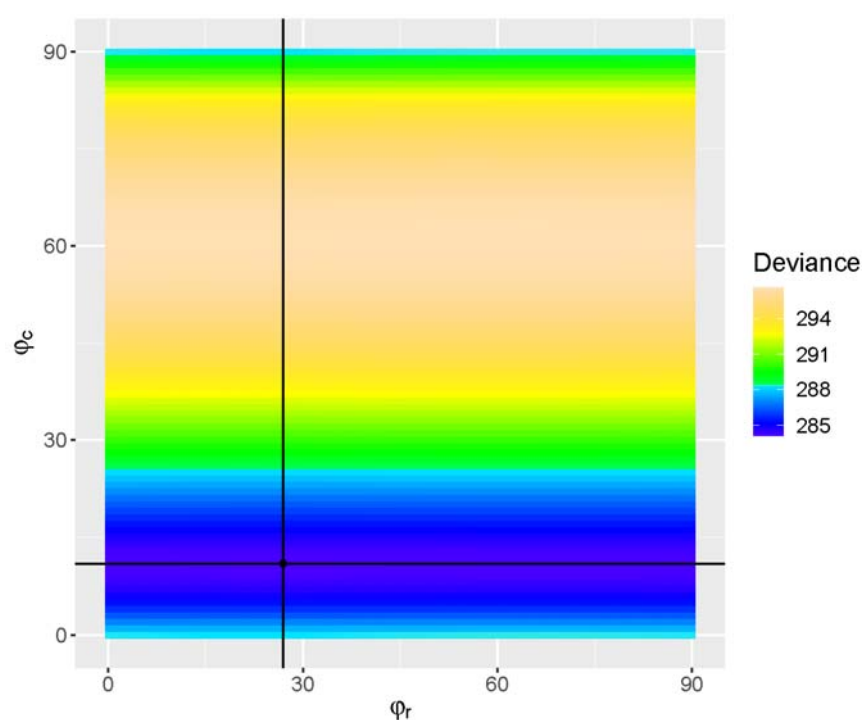
§ Model with fixed effects for genotype, replicate, linear regression on row and column numbers as well as their product, and i.i.d. residual error.

§ Baseline, adding random effects for rows and columns nested within replicates

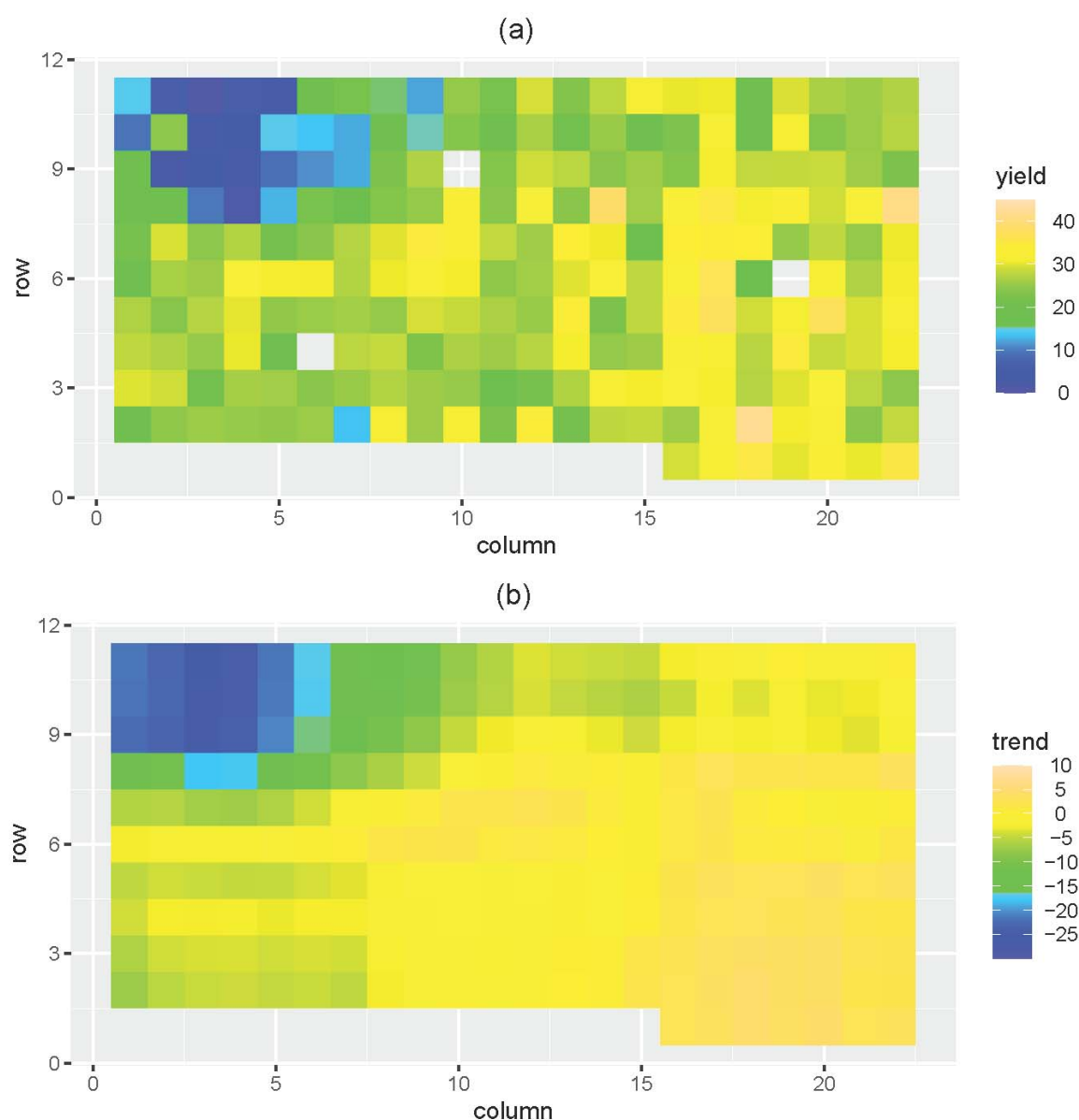




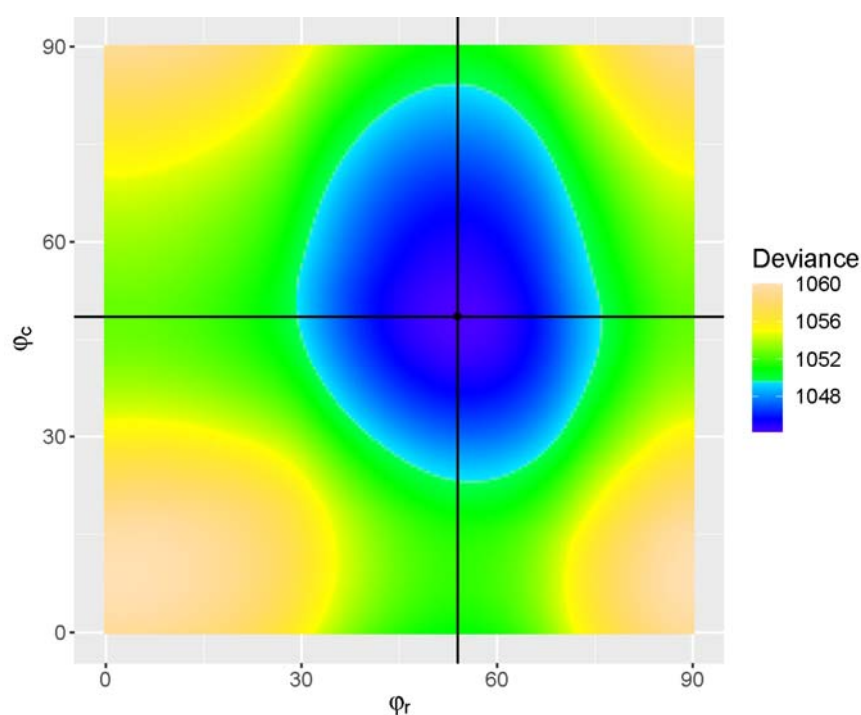
**Figure 1:** Heatmap of barley data of Durbán et al. (2003). (a) Raw data; (b) Smooth trend fitted by model M26 (Table 2).



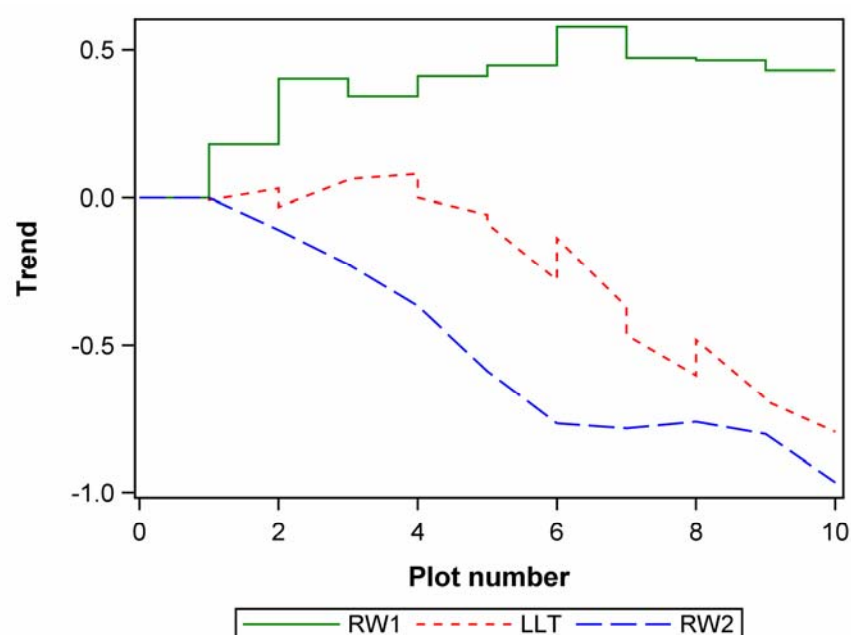
**Figure 2:** Deviance profiled for  $\varphi_r$  and  $\varphi_c$  for barley data using model M12 with  $q = 3$  and diagonal structures ( $\Omega$ ) for both marginal smooths (Table 2). Minimum at  $\varphi_r = 27^\circ$  and  $\varphi_c = 11^\circ$ .



**Figure 3:** Heatmap of wheat data of Stroup et al. (1994). (a) Raw data; (b) Smooth trend fitted by model M26 (Table 2).



**Figure 4:** Deviance profiled for  $\varphi_r$  and  $\varphi_c$  for wheat data using model M12 (Table 2). Minimum at  $\varphi_r = 52^\circ$  and  $\varphi_c = 48^\circ$ .



**Figure B1:** Simulated trend for three state-space models. RW1 = first-order random walk; RW2 = second-order random walk; LLT = local linear trend;  $\mu_0 = b_0 = 0$ ;  $\sigma_{u0}^2 = \sigma_{u1}^2 = 0.1$ .