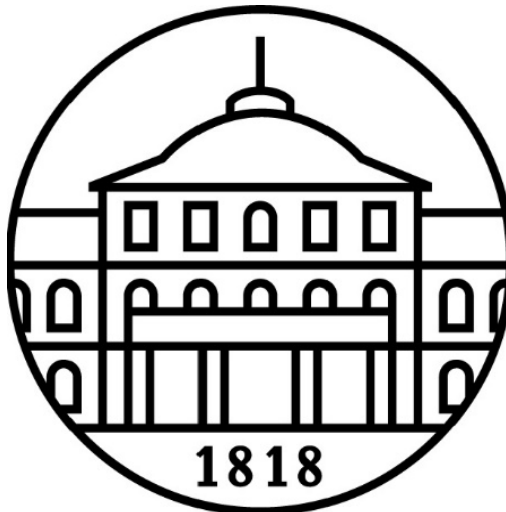


# Two-dimensional P-spline smoothing for spatial analysis of field trials

Hans-Peter Piepho  
Biostatistics Unit  
Institute of Crop Science  
University of Hohenheim  
Germany



This is joint work with:

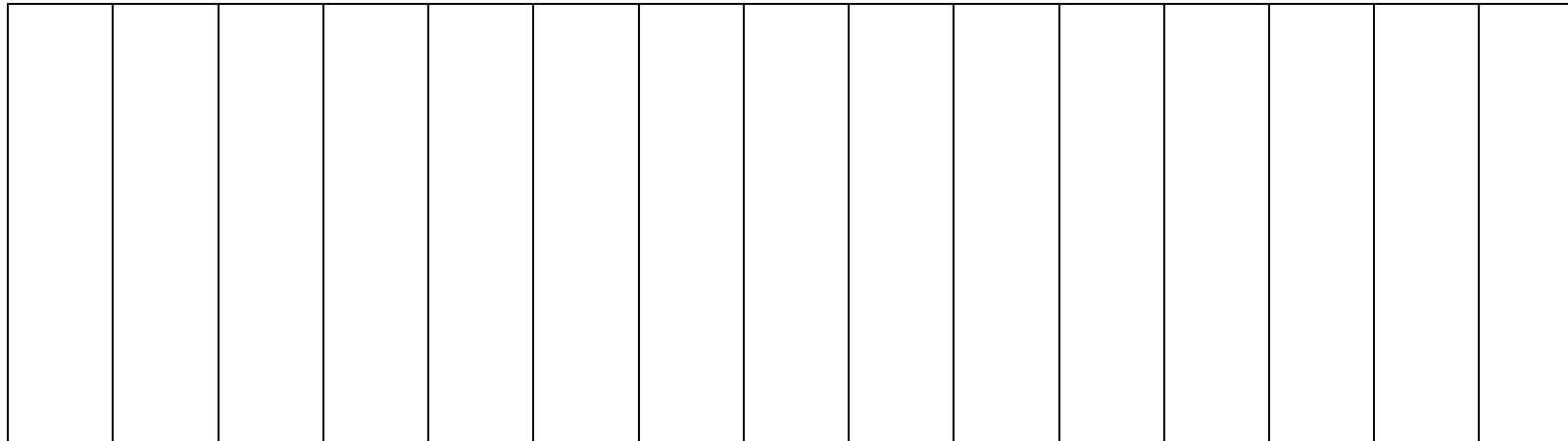
**Martin Boer**

*Wageningen University & Research (WUR), Wageningen, NL*

**Emlyn Williams**

*Australian National University (ANU), Canberra, AUS*

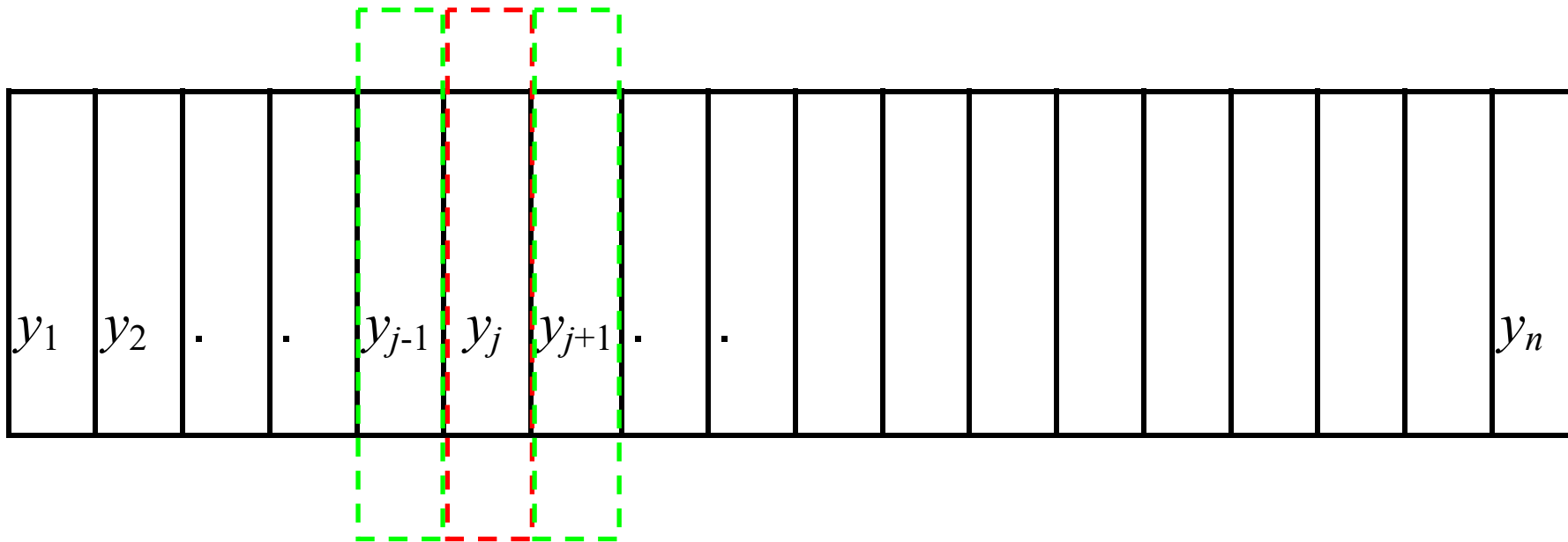
# Spatial models



Spatial correlation among neighbouring plots

**Figure:** A single row of plots. Arrow indicates direction of spatial correlation.

## Nearest neighbour analysis (NNA) (Papadakis 1937)



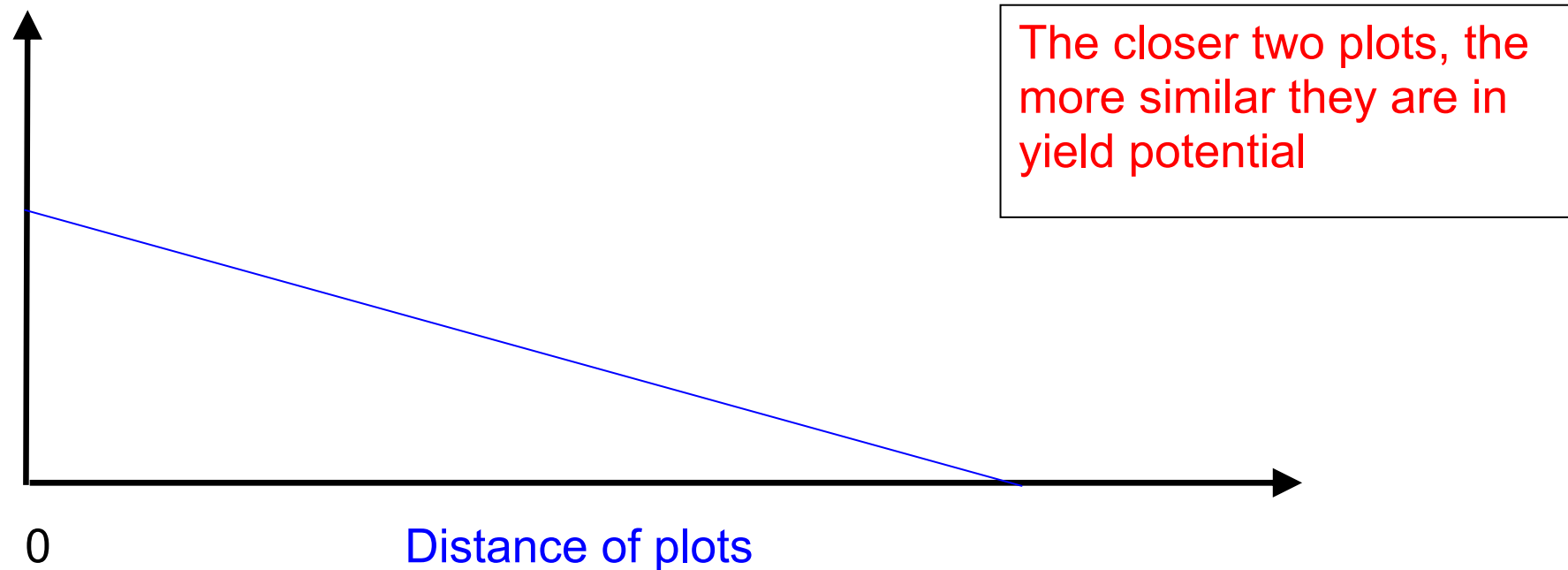
$y_j$  = yield on  $j$ -th plot

Second differences: 
$$y_j - \frac{y_{j-1} + y_{j+1}}{2}$$

## An equivalent mixed model (spatial)

Linear Variance (LV) model (Williams 1986)

Covariance ( $\Rightarrow$  similarity of plots)



$\Rightarrow$  Can implement NNA with mixed models

## Random walk models (state-space models)

$t_1, t_2$  = trend values of two adjacent plots

**First-order random walk  $\Rightarrow$  first differences:**

$$t_1 - t_2 = a_2 \Leftrightarrow t_2 = t_1 + a_2 \quad ; \quad a_2 \sim N(0, \sigma_a^2) \quad \Rightarrow \text{equivalent to LV!}$$

**Second-order random walk  $\Rightarrow$  second differences:**

$$(t_1 - t_2) - (t_2 - t_3) = t_1 - 2t_2 + t_3 = a_2 \Leftrightarrow t_2 = (t_1 + t_3)/2 + a_2 \quad ; \quad a_2 \sim N(0, \sigma_a^2)$$

**First-order autoregressive model [AR(1)]:**

$$t_2 = \rho t_1 + a_2 \quad ; \quad a_2 \sim N(0, \sigma_a^2) \quad 0 < \rho < 1 \quad \text{(Lee et al. 2021)}$$

# P-splines

- (1) Use effects  $u_j$  to model trend values  $t_i$  for the plots
- (2) Smear out the  $u_j$  over adjacent plots
- (3) The  $u_j$  are regression coefficients for B-spline basis functions  $B = \{b_{ij}\}$

Trend value for the  $i$ -th plot:

$$t_i = b_{i1}u_1 + b_{i2}u_2 + \dots + b_{im}u_m \quad ; \quad b_{i1} + b_{i2} + \dots + b_{im} = 1$$

P = penalized  $\Rightarrow$  penalty on regression coefficients  $u_j$   
 $\Rightarrow$  first differences or second differences!

(Rodriguez-Alvarez et al. 2018; 'SpATS' package in R)

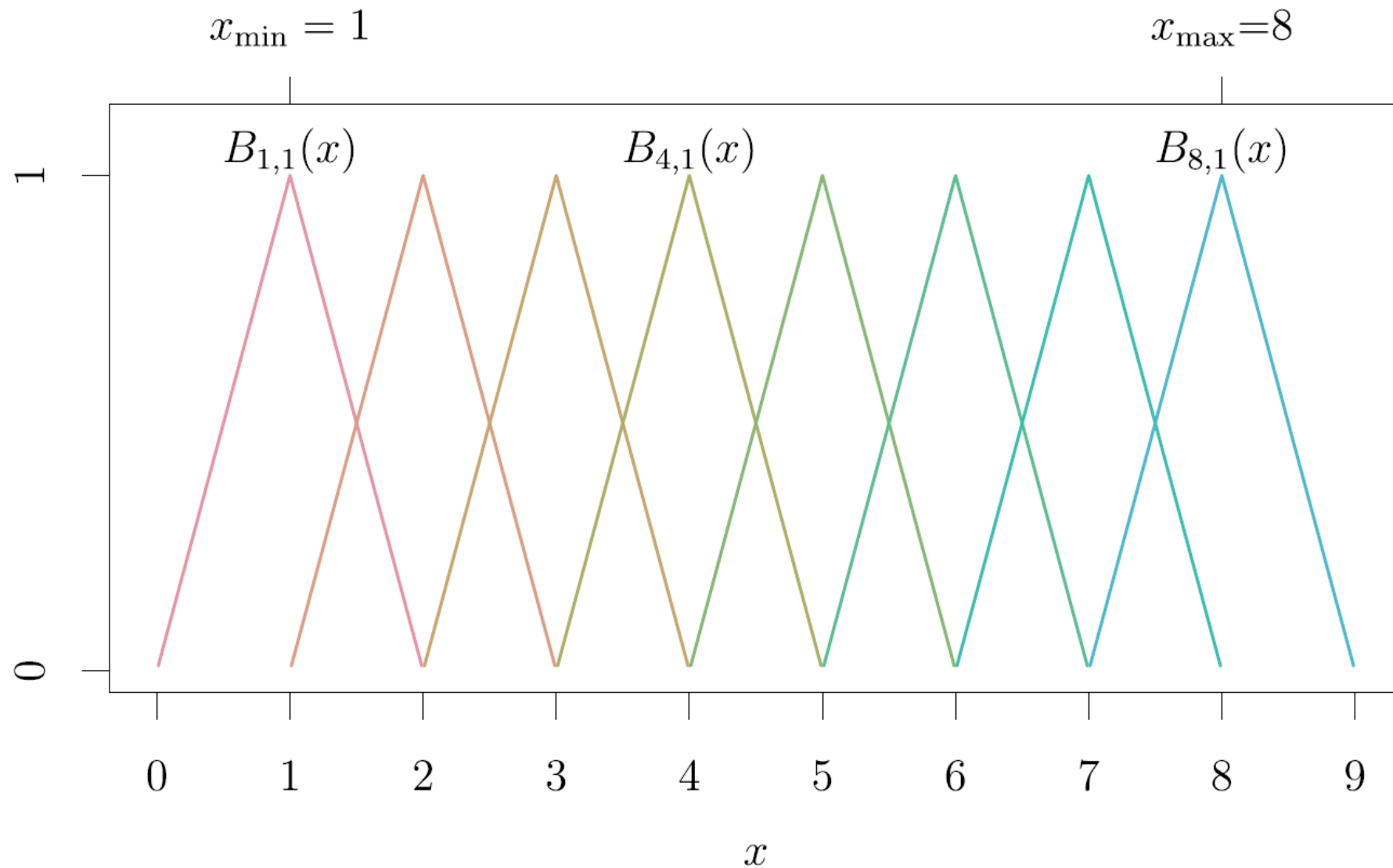


Figure 1. Example of a first-degree B-spline basis for a continuous coordinate  $x$ .



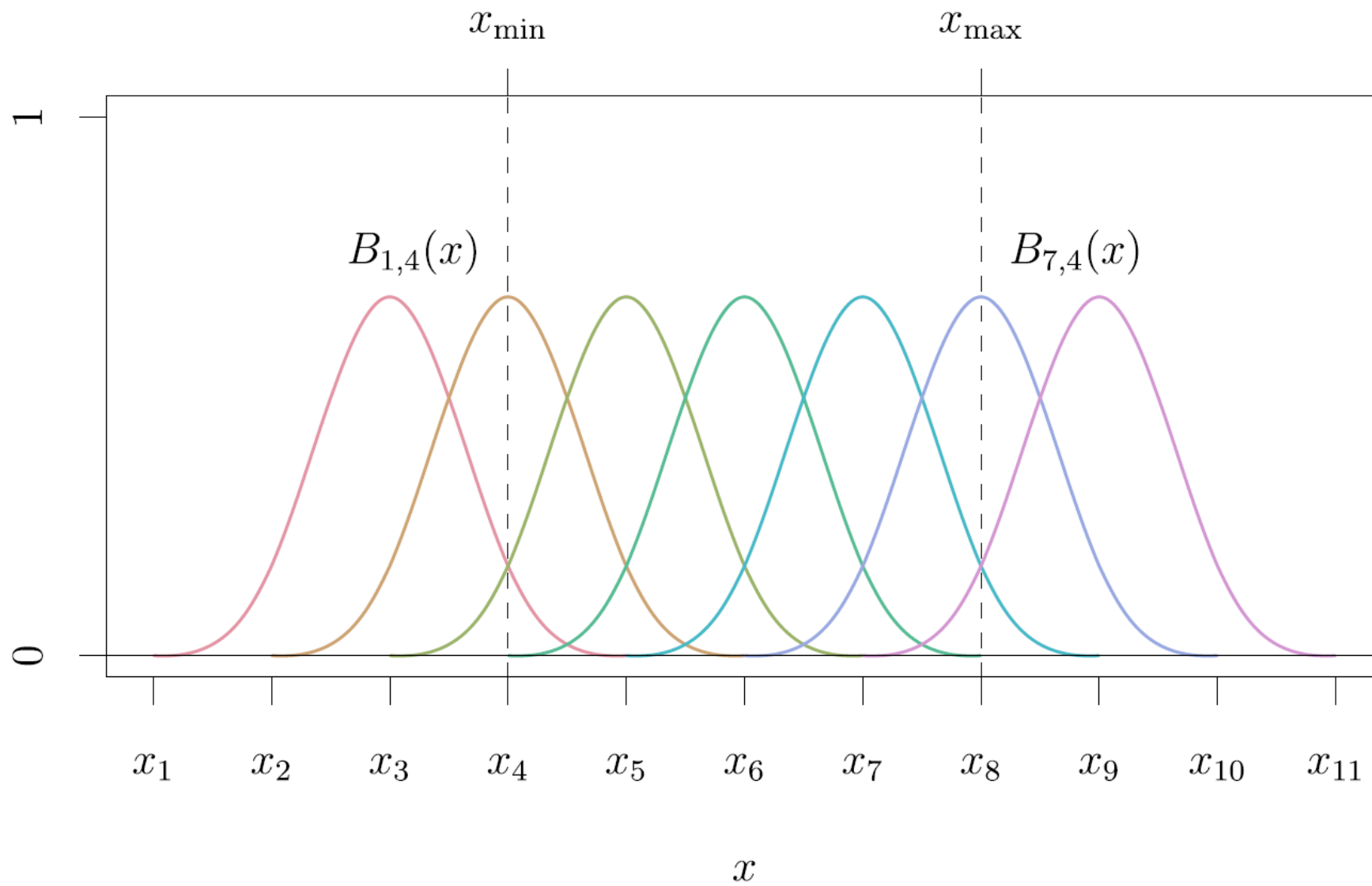
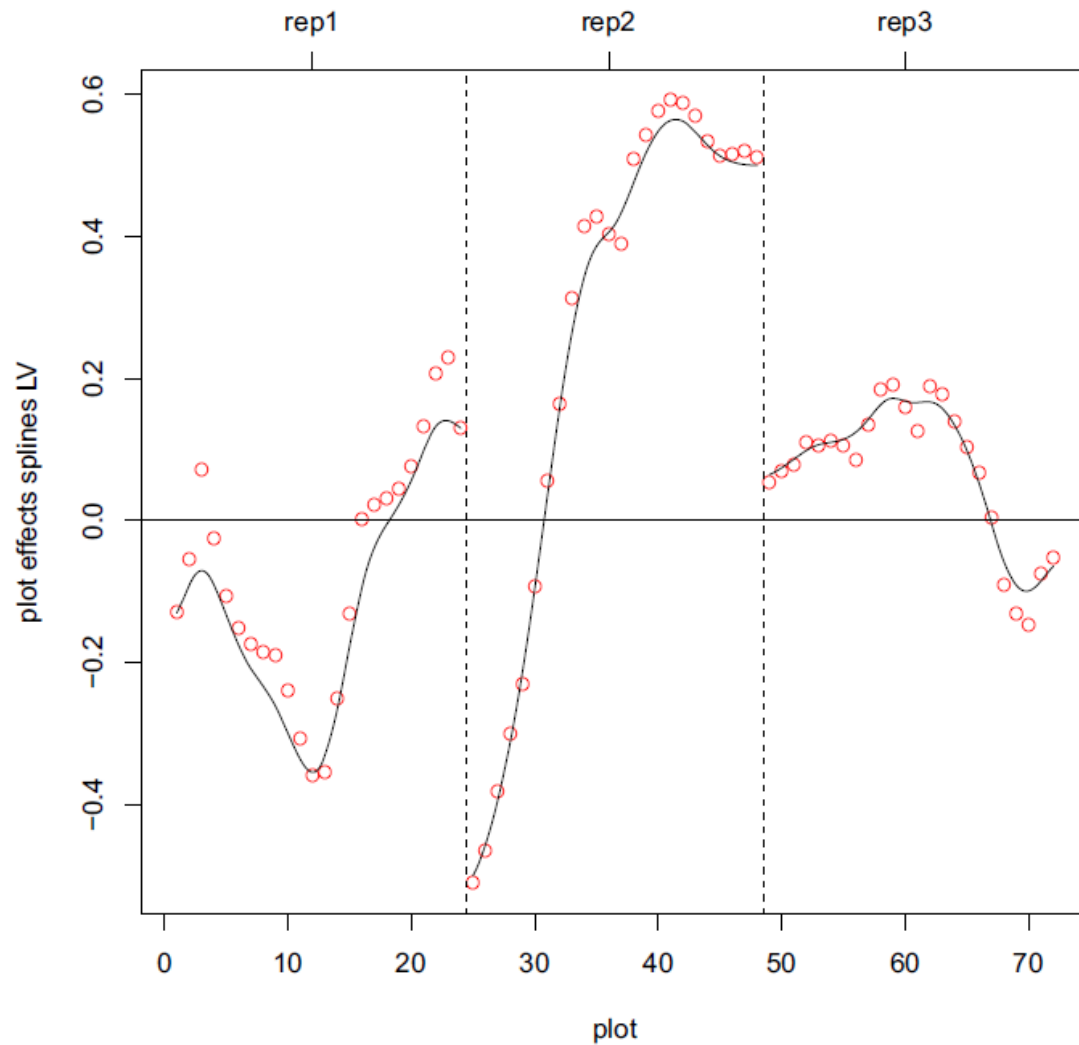


Figure 2. Cubical B-spline basis for a continuous coordinate  $x$ .



- Trend LV model
- Trend 3<sup>rd</sup> degree P-spline

(Boer et al., 2020)

Figure 4. Comparison of plot effects  $u$  for the oats data. The red points are the estimates for the LV model and equivalent to first-order P-splines. The black curve is based on P-splines, using a third degree B-splines with 12 segments, so half the number of plots per replicates.

## Examples for 5 plots

Cubical = third-degree, 7 knots at plots (including "outer" plots!):

$$B = \begin{pmatrix} 0.167 & 0.666 & 0.167 & 0 & 0 & 0 & 0 \\ 0 & 0.167 & 0.666 & 0.167 & 0 & 0 & 0 \\ 0 & 0 & 0.167 & 0.666 & 0.167 & 0 & 0 \\ 0 & 0 & 0 & 0.167 & 0.666 & 0.167 & 0 \\ 0 & 0 & 0 & 0 & 0.167 & 0.666 & 0.167 \end{pmatrix}$$

⇒ need  $u_1, \dots, u_7$

$$\Rightarrow t_1 = 0.167 \times u_1 + 0.666 \times u_2 + 0.167 \times u_3 + 0 \times u_4 + 0 \times u_5 + 0 \times u_6 + 0 \times u_7$$

$$\Rightarrow t_2 = 0 \times u_1 + 0.167 \times u_2 + 0.666 \times u_3 + 0.167 \times u_4 + 0 \times u_5 + 0 \times u_6 + 0 \times u_7$$

etc.

## Examples for 5 plots

Cubical = third-degree, just 5 knots:

$$B = \begin{pmatrix} 0.167 & 0.666 & 0.167 & 0 & 0 \\ 0.021 & 0.479 & 0.479 & 0.021 & 0 \\ 0 & 0.167 & 0.666 & 0.167 & 0 \\ 0 & 0.021 & 0.479 & 0.479 & 0.021 \\ 0 & 0 & 0.167 & 0.666 & 0.167 \end{pmatrix}$$

⇒ need  $u_1, \dots, u_5$

$$\Rightarrow t_1 = 0.167 \times u_1 + 0.666 \times u_2 + 0.167 \times u_3 + 0 \times u_4 + 0 \times u_5$$

$$\Rightarrow t_2 = 0.021 \times u_1 + 0.479 \times u_2 + 0.479 \times u_3 + 0.021 \times u_4 + 0 \times u_5$$

etc.

## Examples for 5 plots

First-degree, just 3 knots:

$$B = \begin{pmatrix} 1 & 0 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0 & 0.5 & 0.5 \\ 0 & 0 & 1 \end{pmatrix}$$

⇒ need  $u_1, \dots, u_3$

$$\Rightarrow t_1 = 1 \times u_1 + 0 \times u_2 + 0 \times u_3$$

$$\Rightarrow t_2 = 0.5 \times u_1 + 0.5 \times u_2 + 0 \times u_3$$

etc.

## Examples for 5 plots

First-degree, 5 knots at plots:

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

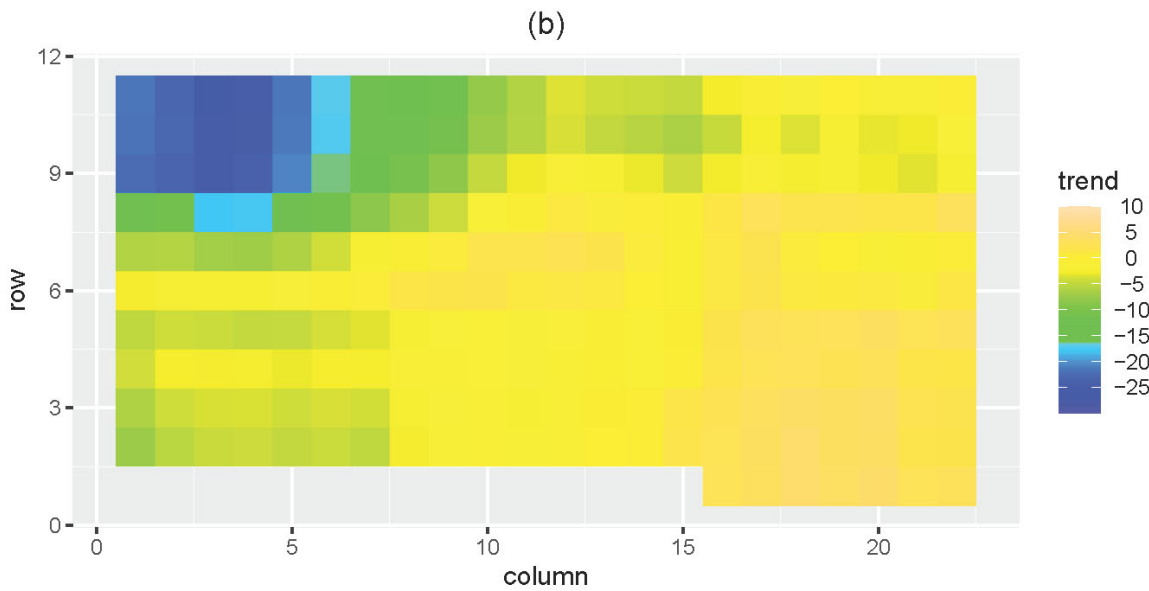
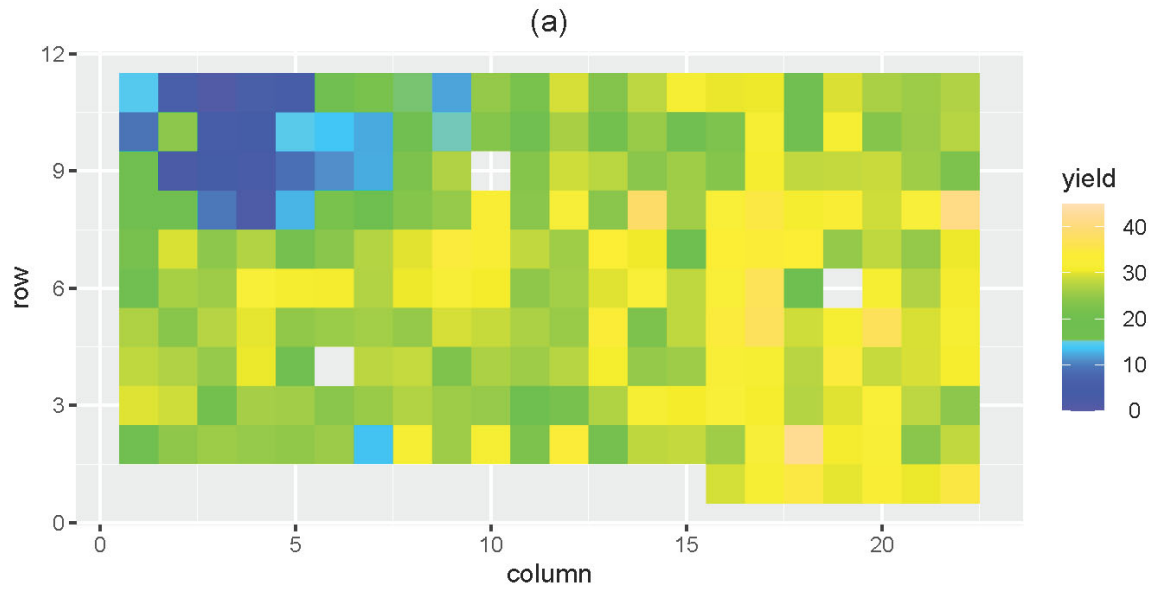
⇒ need  $u_1, \dots, u_5$

$$\Rightarrow t_1 = 1 \times u_1 + 0 \times u_2 + 0 \times u_3 + 0 \times u_4 + 0 \times u_5 = u_1$$

$$\Rightarrow t_2 = 0 \times u_1 + 1 \times u_2 + 0 \times u_3 + 0 \times u_4 + 0 \times u_5 = u_2$$

etc.

⇒ This is LV when first differences are used to penalize  $u_i$ !



**Figure** Heatmap of wheat data of Stroup et al. (1994).

(a) Raw data

(b) Smooth trend (P-spline)

(Piepho et al., 2022)

## Intermediate summary on P-splines

- Modelling smooth trend
- Just another spatial covariance model
- All covariance structures linear in the variance components
- LV and random walk are special cases when using first differences
- Many knobs (degree of B-spline basis, number of knots, difference penalty)
- SpATS uses second differences, but often first differences sufficient

(Boer et al. 2020; Piepho et al. 2022)



## Fitting P-splines as mixed models: one column of plots

$$t = Bu$$

$t$  = vector of trend values for  $k$  plots

$B = \{b_{ij}\} = k \times m$  matrix of  $m$  B-spline bases of  $q$ -th degree

$u$  =  $m$ -vector of coefficients

### Penalty for coefficients $u$ :

$$\theta u^T D^T D u$$

$\theta$  = penalty parameter

$D = (m - p) \times m$  matrix of  $p$ -th differences

## The penalty

$$\theta u^T P u$$

$$P = D^T D = \text{penalty matrix}$$

⇒ equivalent to quadratic form for random effect in likelihood for random effects  $u$

$$\sigma^{-2} u^T P u$$

where

$$\sigma^2 = \text{variance}$$

$$P = \text{precision matrix}$$

## Next question

What is the variance-covariance matrix for  $u$ ?

⇒ needed for mixed model package

If  $P$  were positive-definite, we could just use  $\text{var}(u) = P^{-1}\sigma^2$

But  $P$  is singular!

⇒ use spectral decomposition of  $P$

## Spectral decomposition of $P$

$$P = U \operatorname{diag}(d) U^T$$

$U$  = eigenvectors of  $P = (U_+ | U_0)$

$d$  = eigenvalues of  $P$

where

$U_+$  =  $(m - p)$  eigenvectors corresponding to  $d_+$

$U_0$  =  $p$  eigenvectors corresponding to zero eigenvalues

$d_+$  = sub-vector of  $(m - p)$  positive eigenvalues in  $d$

## Decomposing $Bu$

$$Bu = BUU^T u = BU_0 U_0^T u + BU_+ U_+^T u = X\beta + Zw$$

where

$$X = BU_0 \quad \text{and} \quad \beta = U_0^T u \quad \text{with precision zero!} \Rightarrow \text{fixed effect!}$$

$$Z = BU_+ \quad \text{and} \quad w = U_+^T u \quad \text{with precision } \theta \text{ } \textit{diag}(d_+)$$

## The penalty

Using

$$P = U \operatorname{diag}(d) U^T = U_+ \operatorname{diag}(d_+) U_+^T$$

and

$$w = U_+^T u$$

we find

$$\theta u^T P u = \theta u^T U_+ \operatorname{diag}(d_+) U_+^T u = \theta w^T \operatorname{diag}(d_+) w$$

$$\Rightarrow \text{can assume } \operatorname{var}(w) = \theta^{-1} \operatorname{diag}(d_+^{-1}) = \sigma^2 \operatorname{diag}(d_+^{-1})$$

## The variance-covariance matrix for $u$

The penalty again:

$$\theta u^T P u = \theta u^T U_+ \text{diag}(d_+) U_+^T u = \theta w^T \text{diag}(d_+) w$$

Observing that

$$P^+ = U_+ \text{diag}(d_+^{-1}) U_+^T$$

is the Moore-Penrose inverse of  $P$ , we can also just fit the random effect  $Bu$

assuming that

$$\text{var}(u) = \sigma^2 P^+$$

**But:** Must also fit fixed effects  $X\beta = BU_0^T u$

(Lee et al. 2021)

## What are the fixed effects? (What's in the null space?)

$p = 1$ :

$X = 1$  (just an intercept)

$p = 2$ :

$X = (1 \quad h)$

$h$  = vector of plot numbers

⇒ linear regression on plot numbers



## Extending the model to two dimensions

Plots of a field trial on a regular grid with  $k$  rows and  $s$  columns

Model for spatial trend:

$$t = B_{rc} u_{rc}$$

where

$$B_{rc} = B_r \otimes B_c$$

$B_r$  = matrix of  $m_r$  B-spline bases for  $k$  rows

$B_c$  = matrix of  $m_c$  B-spline bases for  $s$  columns

## A separable two-dimensional penalty

Consider differences  $D_{rc}u_{rc}$  where

$$D_{rc} = D_r \otimes D_c$$

$D_r$  = matrix of  $p$ -th differences for  $k$  rows

$D_c$  = matrix of  $p$ -th differences for  $s$  columns

$$\Rightarrow \text{penalty } \theta_{rc} u_{rc}^T P_{rc} u_{rc}$$

where

$$P_{rc} = P_r \otimes P_c = D_r^T D_r \otimes D_c^T D_c$$

## What are the fixed effects?

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

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

Fixed effects (null space) have dimension

$$m_r m_c - (m_r - p)(m_c - p)$$

$\Rightarrow$  need extra smoothing terms for row and column main effects

## An important detail when $p = 2$

Have design matrices

$$X_r = \begin{pmatrix} 1 & h_r \end{pmatrix} \text{ and } X_c = \begin{pmatrix} 1 & h_c \end{pmatrix}$$

When smoothing these using P-splines, have a [random coefficient regression](#)

⇒ need to allow for a covariance between intercept and slope

⇒ this is not nice but necessary to ensure invariance

⇒ to our knowledge this fact has been ignored in literature on P-splines

⇒ this complication is a good reason to favor  $p = 1$

## Special cases

Several older approaches turn out to be special cases when knots are at the plots and first differences are used ( $p = 1$ ).

One particular one occurs when first-degree B-spline bases are used ( $q = 1$ ). In this case the model is closely related to the

LV  $\otimes$  LV

model (Piepho and Williams 2010; Boer et al. 2020).

This model, in turn, is a limiting case of the very popular

AR(1)  $\otimes$  AR(1)

model (Gilmour et al. 1997).

## Penalties derived from sum of Kronecker products

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

$$u_{rc}^T \left\{ \theta_{rc1} P_r \otimes I_{m_c} + \theta_{rc2} I_{m_r} \otimes P_c \right\} u_{rc}$$

Wood (2017, p. 232): This can provide smoother fits than interaction smooth

A major advantage: null space is  $X_r \otimes X_c$  and only has dimension  $p^2$

$p = 1$ : null space is just the intercept (Dutta & Mondal 2015)

$p = 2$ : null space also comprises regression on row and column numbers and their products (Rodríguez-Álvarez et al., 2018)  $\Rightarrow$  SpATS

## Interesting special case $p = 1$ & $q = 1$

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:

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

$$2 \text{ var}(u_{i,j} | \dots) = 1 / (\theta_{rc1} + \theta_{rc2}) \quad , \text{ where}$$

$$\alpha = \theta_{rc2} / (\theta_{rc1} + \theta_{rc2}) \quad (\text{modified from Dutta and Mondal, 2015})$$

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.

More difficult to translate to a standard mixed model framework because inverse of the precision matrix is not linear in the parameters (**Wood et al., 2013, p. 345**).

Used here: profiling the residual log-likelihood for  $\alpha$ , i.e., via a grid search over  $\alpha \in [0,1]$  (**Besag and Kooperberg, 1995**).



Rodríguez-Álvarez et al. (2018) consider a simplified version of the IAR penalty:

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

The models implemented in [SpATS](#) involve adding marginal smooths for rows and columns with diagonal variance-covariance structures

## Two Examples

Barley data or Durban et al. (2003)

Wheat data of Stroup et al. (1994)

⇒ Fitted randomization-based baseline model and added spatial components

TABLE 5 Analysis of barley data of Durban 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 the product of row and column numbers

Model	Description	Barley data		Wheat data	
		Deviance	AIC	Deviance	AIC
M53	Baseline <sup>a</sup>	410.19	412.19	1101.53	1103.53
M54	Baseline + row & column <sup>b</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

<sup>a</sup>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.

<sup>b</sup>Baseline, adding random effects for rows and columns nested within replicates.

## The separable penalty ( $p = 1$ )

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 the product of row and column numbers. 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

M26 is doing very well  $\Rightarrow$  mathematically close to  $LV \otimes LV$

$\Rightarrow$  **LVIS** = **L**inear **V**ariance with **I**nteraction **S**plines

# The separable penalty ( $p = 2$ )

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 the product of row and column numbers

Model	$\text{var}(\mathbf{u}_r)^a$	$\mathbf{X}_c^b$	$\text{var}(\mathbf{u}_c)^a$	$\mathbf{X}_r^b$	$\text{var}(\mathbf{u}_{rc})$	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	$\Sigma$	$\bar{\bar{X}}_c$	–	–	–	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	$\Omega$	$\bar{X}_c$	–	–	–	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	$I_2$	$\bar{X}_c$	$I_2$	$\bar{X}_r$	–	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	$\Omega$	$\bar{\bar{X}}_c$	$\Omega$	$\bar{X}_r$	–	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	$\Sigma$	$\bar{\bar{X}}_c$	$\Sigma$	$\bar{\bar{X}}_r$	–	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	$I_2$	$\bar{X}_c$	$I_2$	$\bar{X}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	293.23	301.23	293.60	301.60	1058.02	1066.02	1060.05	1068.05
M19	$\Omega$	$\bar{X}_c$	$\Omega$	$\bar{X}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	293.16	305.16	292.28	304.28	1049.91	1061.91	1057.49	1067.49
M20	$\Sigma$	$\bar{\bar{X}}_c$	$\Sigma$	$\bar{\bar{X}}_r$	$\sigma_{rc}^2 P_r^+ \otimes P_c^+$	280.12	296.12	280.27	296.27	1042.80	1058.80	1044.22	1058.22

<sup>a</sup>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  $\Sigma_c \otimes P_c^+$ .

<sup>b</sup>Representation of  $X_r$  or  $X_c$  in the marginal smooth;  $\bar{X}$  obtained using  $U_0$  of spectral decomposition of  $P$  as obtained by software (IML procedure of SAS);  $\bar{\bar{X}}$  obtained using near optimal orthogonal rotation (Figures 2 and 4).

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

⇒ Second differences not doing so well

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 the product of row and column numbers. Fits obtained by a grid search over  $\alpha = 0, 1(0.1)$

Model	q	p	Marginal smooth <sup>a</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>b</sup>	3	2	$\Omega$	0.5	283.58	295.58	0.5	1063.57	1077.57
M46 <sup>c</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

<sup>a</sup>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  $\widehat{X}_c$  and  $\widehat{X}_r$ , respectively.  $\Sigma$ : using  $P_r^+ \otimes \Sigma_r$  and  $\Sigma_c \otimes P_c^+$  with  $\widehat{X}_c$  and  $\widehat{X}_r$ , respectively;  $\widehat{X}$  obtained using  $U_0$  of spectral decomposition of  $P$  as obtained by software (IML procedure of SAS);  $\widetilde{X}$ : obtained using near optimal orthogonal rotation (Figures 2 and 4).

<sup>b</sup>Fit for fixed  $\alpha = 0.5$ ; using  $\widehat{X}_c$  and  $\widehat{X}_r$  for  $X_c$  and  $X_r$ , respectively.

<sup>c</sup>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).

## Brief summary

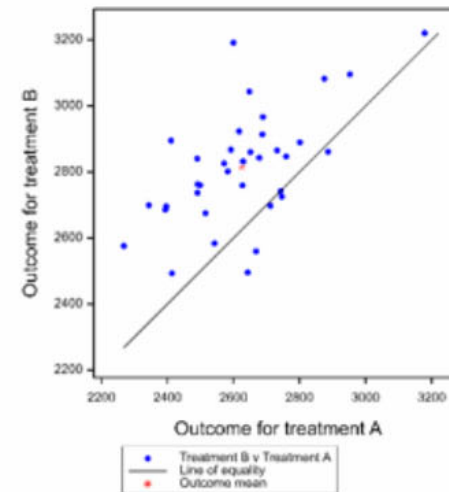
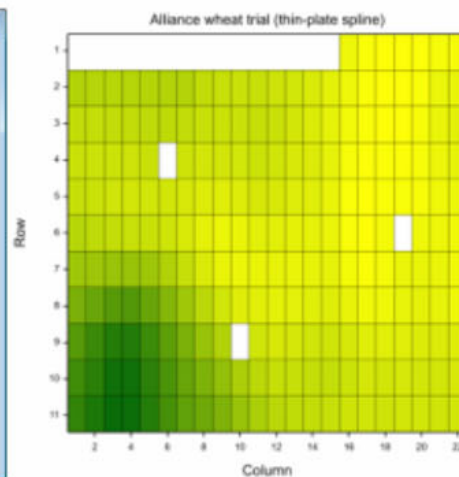
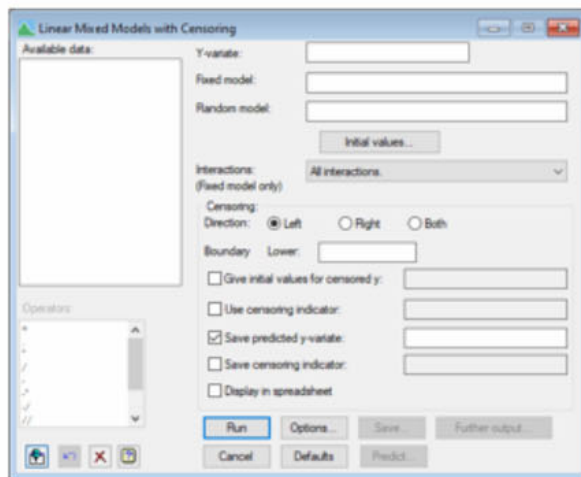
- P-splines are a very versatile option for spatial modelling in field trials
- Several choices to make (many knobs to turn)
- First differences simpler than second differences; work quite well
- Okay to place knots at plots and use first-degree B-spline bases
- Close ties with older approaches [NN-analysis, random walk, LV, AR(1)]
- Our philosophy: always start with a randomization-based model as baseline; any spatial modelling is just an add-on
- Important disclaimer: A sophisticated spatial analysis is no substitute for good experimental design

previously answered question:

## What does Genstat offer?

Genstat provides extensive data analysis facilities, delivering reliable and insightful results no matter how complex your data may be. Our latest update provides even more functionality, including:

- The ability to analyse censored data using ANOVA and regression
- 2-dimensional spline models (i.e., LVIS and SpATS) for modelling spatial correlation
- Tools to design and analyse n-of-1 medical trials



(accessed on 8 August 2024)

## References

Boer, M., 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.

Lee, W., Piepho, H.P., Lee, Y. (2021): Resolving the ambiguity of random-effects models with singular precision matrix. *Statistica Neerlandica* **75**, 482-499.

Piepho, H.P., Boer, M., Williams, E.R. (2022): Two-dimensional P-spline smoothing for spatial analysis of field trials. *Biometrical Journal* **64**, 835-857.

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., Richter, C., Williams, E.R. (2008): Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biometrical Journal* **50**, 164-189.

Piepho, H.P., Williams, E.R. (2010): Linear variance models for plant breeding trials. *Plant Breeding* **129**, 1-8.