Spatial modelling of plant data with the R-packages `ASReml-R` and `asremlPlus`

Chris Brien

Australian Centre for Plant Functional Genomics, University of Adelaide.

THE UNIVERSITY of ADELAIDE

appn Australian Plant Phenomics Network

OzBarley

Outline

1. The spatial models for two-dimensional plant data.
2. Functions in **`asremlPlus`** for fitting and comparing spatial models.
3. An example.
4. Summary.

1. Two types of spatial models for two-dimensional plant data

- Variance models: these model the variance.
  - `asreml` has many functions that can model the variance.
  - A very common model is one that assumes separable, first-order autocorrelation (ar1).
- Tensor-product spline models: these model a surface.
  - `asreml` provides natural cubic smoothing splines that can be used to fit a tensor-product spline (TPNCSS);
  - `TPSbits` (Welham, 2022), which has been incorporated into `asremlPlus`, can fit tensor-product linear and cubic splines (TPPSL1 & TPPSC2).
- The function `addSpatialModel` from `asremlPlus` can fit a wide range of variance and tensor-product P-spline models.

The spatial model used in `asremlPlus`

- The model, an adaptation of Cullis et al. (1997), is of the form:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}_{\mathrm{t}}\boldsymbol{\tau} + \mathbf{X}_{\mathrm{d}}\boldsymbol{\beta}_{\mathrm{d}} + \mathbf{Z}_{\mathrm{d}}\mathbf{u}_{\mathrm{d}} + \mathbf{Z}_{\ell}\mathbf{u}_{\ell} + \mathbf{e}$$

  - μ is the overall mean for the experiment;
  - $\boldsymbol{\tau}$: fixed treatment effects with indicator matrix $\mathbf{X}_{\mathrm{t}}$; (could be random)
  - $\boldsymbol{\beta}_{\mathrm{d}}$: fixed large-scale spatial (design) effects with indicator matrix $\mathbf{X}_{\mathrm{d}}$;
  - $\mathbf{u}_{\mathrm{d}}$: random large-scale spatial (design) effects with indicator matrix $\mathbf{Z}_{\mathrm{d}}$ and each set of effects assumed to be $N\left(0, \sigma_i^2\mathbf{I}_{n_i}\right)$
  - $\mathbf{u}_{\ell}$: random local spatial effects with indicator matrix $\mathbf{Z}_{\ell}$ assumed to be $N(0, \boldsymbol{\Sigma})$
    - e.g. $\boldsymbol{\Sigma}$ could be specify separable ar1 or it could involve spline terms;
  - $\mathbf{e}$: residual effects assumed to be $N(0, \sigma^2\mathbf{I}_n)$ or $N\left(0, \oplus_k\sigma_k^2\mathbf{I}_{n_k}\right)$.

Features of spatial models in `asremlPlus`

- The roles of random and residual terms are reversed from the traditional specification of spatial models:
  - Random terms $(\mathbf{Z}_\ell \mathbf{u}_\ell)$ are used to model local spatial variation.
  - Residual terms $(\mathbf{e})$ represent individual-specific variability.
- Separate models for different sections of the data can be specified.
- Each section is assumed to consist of a two-dimensional, possibly irregular, grid.
  - Generically, the dimensions are labelled rows and cols.
- Local spatial variation at other than the unit level can be fitted e.g. main units.

2. Functions in `asremlPlus` for fitting and comparing spatial models

- **`addSpatialModel`**: adds a spatial model to a fitted initial model.
- **`addSpatialModelOnIC`**: adds a spatial model to a fitted initial model when the fit is improved according to an information criterion (AIC, BIC).
  - One can fit one of (i) a variance model based on a patterned correlation matrix (corr), (ii) a TPNCSS model, or (iii) a TPPS model.
  - There are several arguments for specifying the type of P-spline model.
- **`chooseSpatialModelOnIC`**: fits four spatial models and chooses the model with the best fit according to either the AIC or BIC.
  - The models that can be compared are restricted to (i) corr, (ii) TPNCSS, (iii) TPPSC2, and (iv) TPPSL1.

Selecting models using information criteria in `asremlPlus`

- Required because we are not comparing nested models.
- Can be based on the REML or the full likelihood (Verbyla, 2019).
- For correlation models, deciding whether to add a spatial model is done independently for each dimension within each section.
- For spline models, the selection of models is done independently for each section.

3. An example

- A high-throughput, phenotyping experiment was run in 2 Smarthouses at the Adelaide Plant Accelerator.
  - 215 Barley varieties had 5+ replicates for a total of 1110 pots with a single plant.
  - Varieties allocated using a latinized, semiresolved, incomplete block design:
    - optimized using the R package **`odw`** (Butler, 2021);
    - randomized using **`dae`** (Brien, 2024).
  - The grid in each Smarthouse is irregular and the number of pots differs between the Smarthouses.
  - Blocks were irregular and subBlocks were mostly 3×5.
  - The 1110 plants were each imaged for 100 days:
    - 111,000 values per trait.

Design

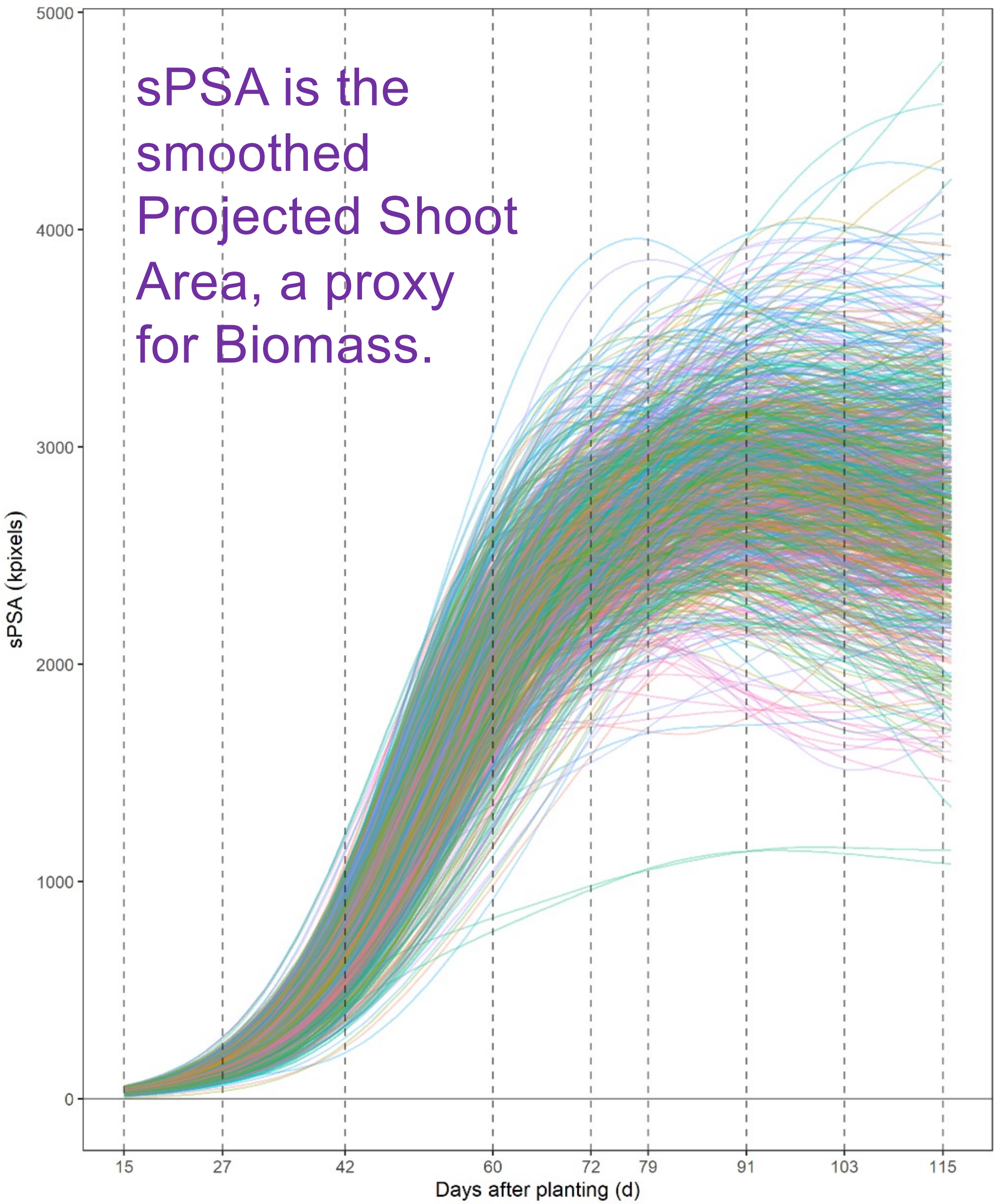
SW Smarthouse

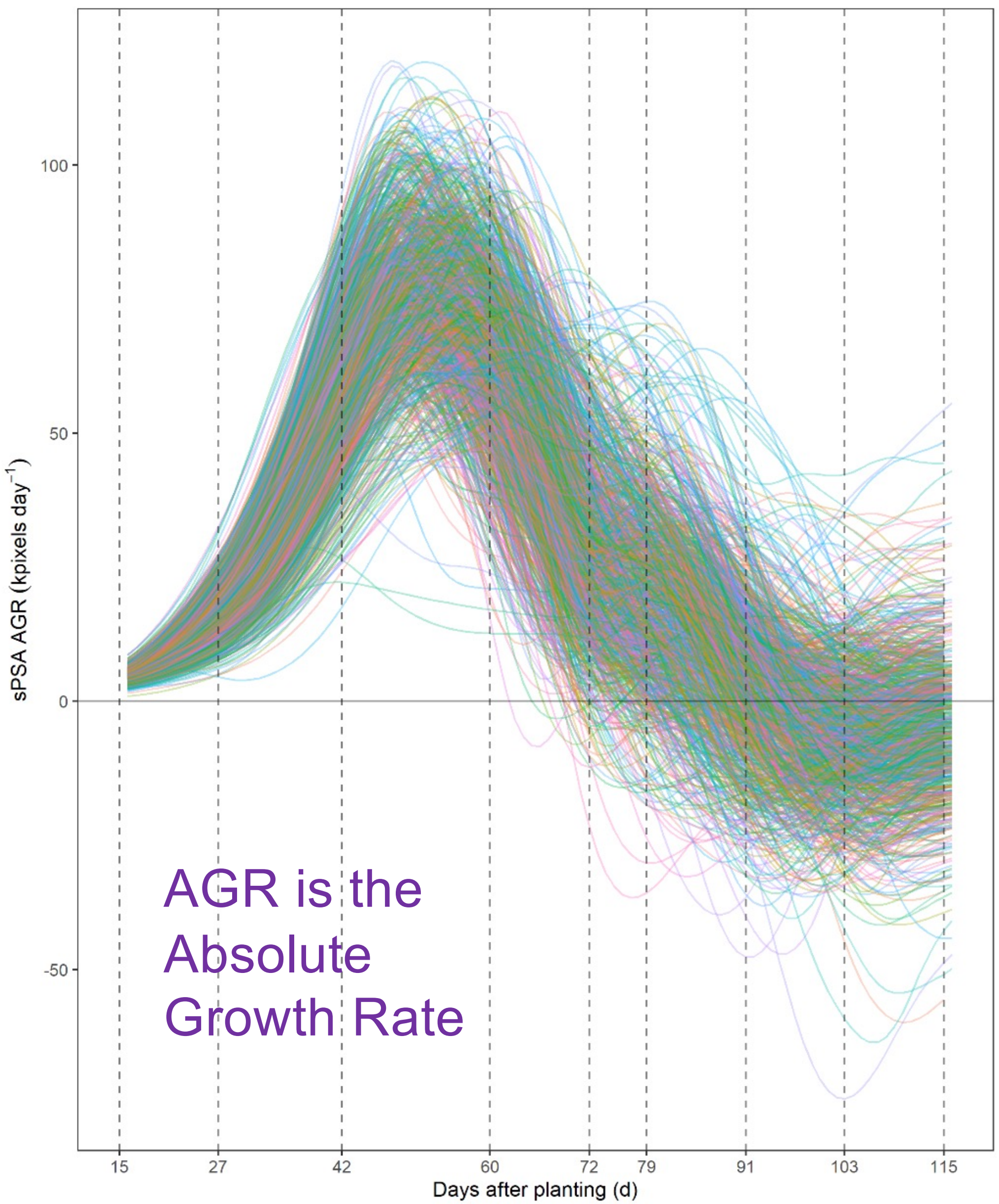
Lane \ Position	26	25	24	23	22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1
24		6	6	6	6	6	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2
23	6	6	6	6	6	6	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2
22		14	14	14	14	14	13	13	13	13	13	12	12	12	12	12	11	11	11	11	11	1	1	1	1	1
21		14	14	14	14	14	13	13	13	13	13	12	12	12	12	12	11	11	11	11	11	1	1	1	1	1
20		14	14	14	14	14	13	13	13	13	13	12	12	12	12	12	11	11	11	11	11	1	1	1	1	1
19		10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
18	10	10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
17	10	10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
16		5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
15		5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
14		5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
13		5	5	5	5	5	4	4	4	4	4		13	13	13	13	12	12	12	12	12	11	11	11	11	11
12	15	15	15	15	15	15	14	14	14	14	14	13	13	13	13	13	12	12	12	12	12	11	11	11	11	11
11	15	15	15	15	15	15	14	14	14	14	14	13	13	13	13	13	12	12	12	12	12	11	11	11	11	11
10	10	10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
9	10	10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
8	10	10	10	10	10	10	9	9	9	9	9	8	8	8	8	8	7	7	7	7	7	6	6	6	6	6
7	5	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
6	5	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1
5	5	5	5	5	5	5	4	4	4	4	4	3	3	3	3	3	2	2	2	2	2	1	1	1	1	1

SE Smarthouse

Lane \ Position	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
24	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	15	15	15	15	15	16	16	16	16	16	16
23	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	15	15	15	15	15	16	16	16	16	16	16
22	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	15	15	15	15	15	16	16	16	16	16	16
21	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	11
20	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	11
19	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	11	11	11	11	11	11
18	11	11	11	11	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14						
17	11	11	11	11	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14						
16	11	11	11	11	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14						
15	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	
14	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
13	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
12	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	
11	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	
10	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	
9	11	11	11	11	11	12	12	12	12	12	13	13	13	13		4	4	4	4	4	5	5	5	5	5	
8	11	11	11	11	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	15	15	15	15	15	15
7	11	11	11	11	11	12	12	12	12	12	13	13	13	13	13	14	14	14	14	14	15	15	15	15	15	15
6	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
5	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
4	6	6	6	6	6	7	7	7	7	7	8	8	8	8	8	9	9	9	9	9	10	10	10	10	10	10
3	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5
2	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5
1	1	1	1	1	1	2	2	2	2	2	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	5

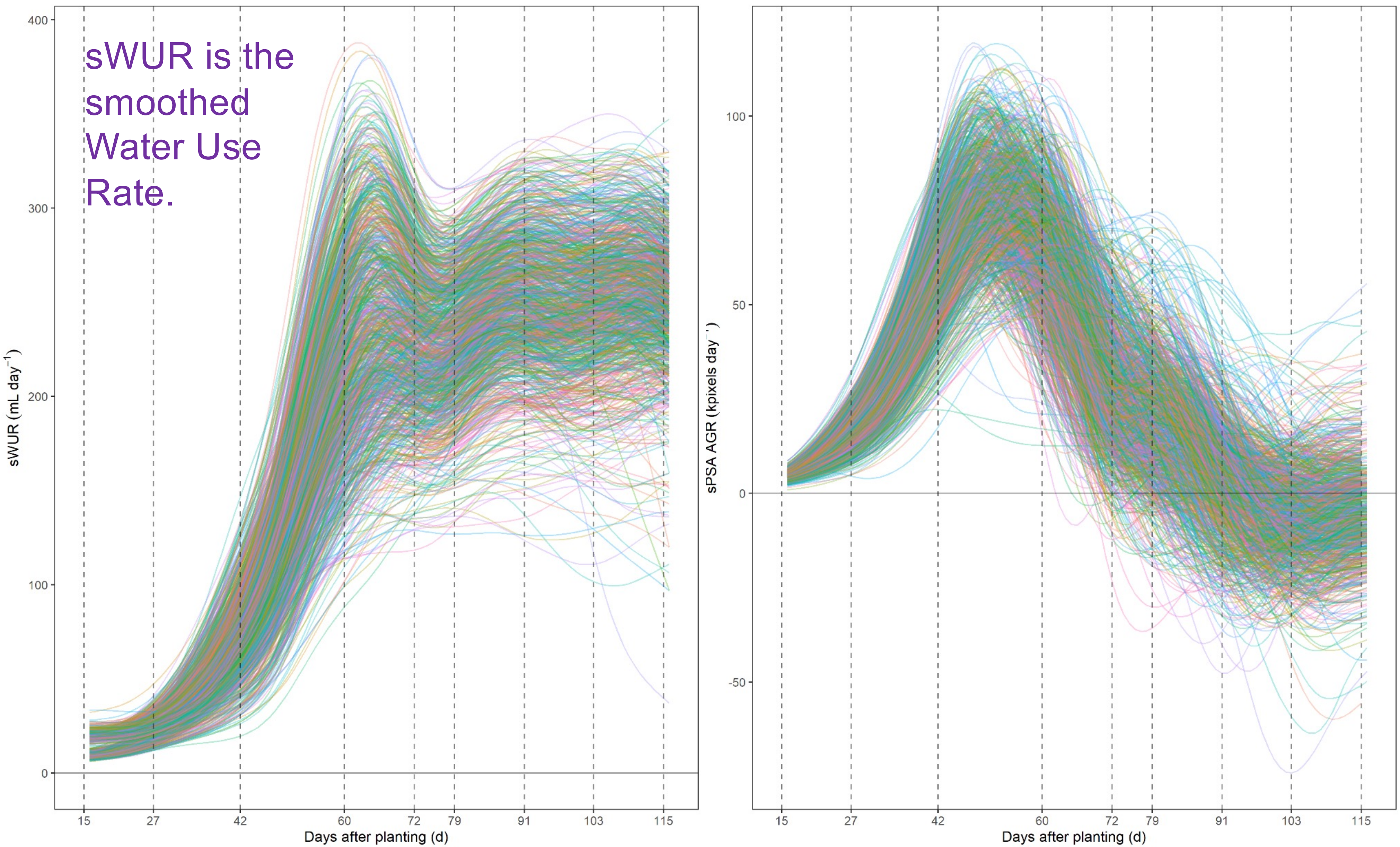
Smoothed profiles





Smoothed profiles

sWUR is the smoothed Water Use Rate.



Some traits

- The longitudinal trend has been divided into 9 intervals, in each of which the growth dynamics are similar.
- Using these intervals, we identify 9 single-DAP sPSA traits and 8 interval-mean traits for each of sPSA AGR and sWUR.
  - These traits have a single value for each plant.
- Conducted LMM analyses that choose the best-fitting of the four spatial models (i) corr, (ii) TPNCSS, (iii) TPPSC2, and (iv) TPPSL1.

Fit the initial model

Term in $\mathbf{X}_\mathrm{d}\boldsymbol{\beta}_\mathrm{d}$

Term in $\mathbf{X}_\mathrm{t}\boldsymbol{\tau}$

Separate Smarthouse terms in $\mathbf{Z}_\mathrm{d}\mathbf{u}_\mathbf{d}$, with **`BRow`** and **`BCol`** being rows and cols of subBlocks

```
> #Fit initial model
> current.asr <- do.call(asreml,
+                        args=list(fixed = sPSA.15 ~ Smarthouse + OzBarleyID,
+                                  random = ~ at(Smarthouse, 'SW'):((SBlock/BRow)*BCol) +
+                                             at(Smarthouse, 'SE'):((SBlock/BRow)*BCol),
+                                  residual = ~
+                                    dsum(~ units | Smarthouse, levels = list(c('SW'))) +
+                                    dsum(~ units | Smarthouse, levels = list(c('SE'))),
+                                  data = indv.dat, maxiter=50))
```

Separate Smarthouse terms in $\mathbf{e}$

```
> #Create asrtests object with components asreml object, wald.tab and test.summary
> current.asrt <- as.asrtests(current.asr, NULL, NULL, IClikelihood = "full",
+                            label = "Initial design effects model")
> current.asrt <- rmboundary(current.asrt)
```

Fit and compare several spatial models

Separate spatial models for Smarthouse

Numeric centred covariates for splines

```
> #Try spatial models
> spatial.fit <- chooseSpatialModelOnIC(current.asrt, sections = "Smarthouse",
+                                        row.covar = "cLane", col.covar = "cPosn",
+                                        row.factor = "Lane", col.factor = "Position",
+                                        asreml.option = "grp",
+                                        rotateX = TRUE, ngridangles = NULL,
+                                        trySpatial = "all", maxit = 100)
```

P-spline arguments

Factors for ar1 models

```
elapsed time for bobyqa: 57.31 seconds

#### Optimal thetas: 0, 50.60790647526  with criterion 5067.536

elapsed time for bobyqa: 108.11 seconds

#### Optimal thetas: 72.2566046758365, 45.9425888740954  with criterion 5057.485
```

Choose the best fitting spatial model

```
> spatial.fit$spatial.IC
           fixedDF varDF      AIC      BIC    loglik
nonspatial     215     8 5068.155 6183.432 -2311.077
corr           215     8 5059.908 6175.186 -2306.954
TPNCSS         219     9 5044.491 6184.775 -2294.245
TPPSC2         221     9 5047.000 6197.286 -2293.500
TPPSL1         215     8 5048.682 6163.960 -2301.341
> R2.adj =  R2adj(spatial.fit$asrts$TPNCSS$asreml.obj,
+                 include.which.fixed = ~ . - OzBarleyID,
+                 orthogonalize = "eigen",
+                 include.which.random = ~ .)
> R2.adj
[1] 27.73243
attr(,"fixed")
~Smarthouse + at(Smarthouse, "SW"):cLane + cLane:at(Smarthouse,
    "SE") + at(Smarthouse, "SW"):cPosn + at(Smarthouse, "SE"):cPosn
<environment: 0x0000023935bde738>
attr(,"random")
~.
```

R2adj is an implementation of Piepho's (2023) adjusted R^2 for LMMs.

The fixed and random terms

```
Wald tests for fixed effects.
                                Df  denDF    F.inc     Pr
(Intercept)                      1   34.0  23910.0 0.0000
Smarthouse                       1   34.1    296.1 0.0000
OzBarleyID                     213  743.5      5.1 0.0000
at(Smarthouse, 'SW'):cLane       1   25.0      0.2 0.6298
at(Smarthouse, 'SE'):cLane       1   21.0      2.7 0.1170
at(Smarthouse, 'SW'):cPosn       1   30.9      4.8 0.0366
at(Smarthouse, 'SE'):cPosn       1   27.3     16.2 0.0004
> summary(spatial.fit$asrts$TPNCSS$asreml.obj)$varcomp
                                              component   std.error    z.ratio bound %ch
at(Smarthouse, 'SE'):spl(cLane):cPosn       0.001316567 0.003430729  0.3837573     P   0
at(Smarthouse, 'SW'):spl(cPosn)             0.080853911 0.170515697  0.4741728     P   0
at(Smarthouse, 'SE'):spl(cPosn)             0.303036161 0.397697847  0.7619759     P   0
at(Smarthouse, 'SE'):dev(cLane)             1.860556900 1.053875627  1.7654426     P   0
at(Smarthouse, 'SW'):BCol:SBlock:BRow       0.259251881 0.654683312  0.3959959     P   0
at(Smarthouse, 'SE'):BCol:SBlock:BRow       1.135385995 1.035700006  1.0962499     P   0
at(Smarthouse, 'SE'):spl(cPosn):spl(cLane)  0.286450480 0.338224956  0.8469230     P   0
Smarthouse_SE!R                            30.525386920 2.228570202 13.6972965     P   0
Smarthouse_SW!R                            25.092817128 2.039614855 12.3027233     P   0
```

- The only large-scale spatial terms (from $\mathbf{Z}_d\mathbf{u}_d$) remaining in the model.
  - The rest have gone to zero.
  - These two represent sub-block variability within a Smarthouse.

Fixed Lines				Random Lines			
Trait	AIC change	Fitted model	R^2 adj. (%)	Trait	AIC change	Fitted model	R^2 adj. (%)
sPSA.15	-23.7	TPNCSS	27.7	sPSA.15	-13.8	TPPSC2	27.2
sPSA.27	-48.0	TPPSC2	27.2	sPSA.27	-38.5	TPNCSS	28.1
sPSA.42	-92.9	TPPSC2	37.1	sPSA.42	-67.9	TPPSC2	37.9
sPSA.60	-102.1	TPPSC2	30.0	sPSA.60	-81.9	TPPSC2	29.6
sPSA.72	-72.7	TPPSC2	15.0	sPSA.72	-58.8	TPPSC2	17.4
sPSA.79	-57.3	TPPSC2	16.0	sPSA.79	-41.7	TPPSC2	16.5
sPSA.91	-32.4	TPPSC2	8.7	sPSA.91	-12.7	AR1	7.5
sPSA.103	-26.4	TPPSC2	5.7	sPSA.103	-11.0	TPPSC2	5.9
sPSA.115	-29.5	TPPSL1	9.7	sPSA.115	-21.3	TPPSC2	10.7
sPSA.AGR.15to27	-62.4	TPPSC2	29.0	sPSA.AGR.15to27	-47.5	TPPSC2	30.0
sPSA.AGR.27to42	-103.9	TPPSC2	38.3	sPSA.AGR.27to42	-76.6	TPPSC2	39.1
sPSA.AGR.42to60	-54.2	TPPSC2	17.5	sPSA.AGR.42to60	-44.6	TPPSC2	17.5
sPSA.AGR.60to72	-45.0	TPPSC2	24.3	sPSA.AGR.60to72	-33.4	TPPSC2	23.8
sPSA.AGR.72to79	-12.1	TPNCSS	20.3	sPSA.AGR.72to79	-3.7	TPNCSS	20.3
sPSA.AGR.79to91	-13.7	TPPSC2	7.3	sPSA.AGR.79to91	-6.8	TPPSC2	5.5
sPSA.AGR.91to103	-76.8	TPPSC2	36.2	sPSA.AGR.91to103	-51.6	TPPSC2	17.3
sPSA.AGR.103to115	-21.3	TPPSC2	19.5	sPSA.AGR.103to115	-12.6	TPPSC2	21.9
sWUR.15to27	-153.1	TPPSC2	92.5	sWUR.15to27	-126.3	TPPSC2	92.1
sWUR.27to42	-95.0	TPPSC2	42.4	sWUR.27to42	-63.1	TPPSC2	41.6
sWUR.42to60	-120.9	TPPSC2	49.5	sWUR.42to60	-86.3	TPPSC2	49.4
sWUR.60to72	-210.3	TPPSC2	74.9	sWUR.60to72	-159.6	TPPSC2	75.3
sWUR.72to79	-305.6	TPPSC2	82.5	sWUR.72to79	-241.3	TPPSC2	84.4
sWUR.79to91	-258.9	TPPSC2	80.8	sWUR.79to91	-215.5	TPPSC2	83.0
sWUR.91to103	-189.5	TPPSC2	72.3	sWUR.91to103	-147.5	TPPSC2	72.7
sWUR.103to115	-124.0	TPPSC2	70.2	sWUR.103to115	-96.2	TPPSC2	69.6

4. Summary

- **`R`** package **`asremlPlus`** in conjunction with **`ASReml-R`**:
  - robustly fits a wide range of variance models for local spatial variation;
  - Fits tensor-product natural cubic smoothing splines (TPNCSS);
  - Fits tensor-product P-splines (TPPS) that can vary in their degree and the order of the differencing for the penalty, as well as incorporating the Piepho et al. (2022) modifications.
- Separate models for sections whose two-dimensional grids can differ.
- Local spatial variation at other than the unit level can be fitted.
- Can use AIC or BIC with reml or full likelihood in choosing models.
- Checks for, and can exclude, unconverged models or models with fixed correlations and removes bound terms, if possible.
- Available on CRAN or my **`R`** repository (http://chris.brien.name/rpackages).
- Compatible with **`ASReml-R`** version 4.2.

References

- Brien, C. J. (2024) `asremlPlus`*: augments ASReml-R in fitting mixed models and packages generally in exploring predictions. R* package version 4.4.29. http://chris.brien.name/rpackages or http://cran.at.r-project.org/package=asremlPlus.
- Brien, C. J. (2024). dae: functions useful in the design and ANOVA of experiments. R package version 3.2.21. http://chris.brien.name/rpackages or https://cran.at.r-project.org/package=dae/
- Butler, D. G. (2021). odw: Generate optimal experimental designs (Version 2.1.2): . https://mmade.org/
- Butler, D. G., Cullis, B. R., Gilmour, A. R., Gogel, B. J., & Thompson, R. (2023). ASReml-R reference manual. (Version 4.2). Hemel Hempstead: VSN International Ltd. URL http://asreml.org
- 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*(3), 269-293.
- Piepho, H.-P. (2023). An adjusted coefficient of determination (R2) for generalized linear mixed models in one go. *Biometrical Journal, 65*(7), 2200290. https://doi.org/10.1002/bimj.202200290
- Piepho, H.-P., Boer, M. P., & Williams, E. R. (2022). Two-dimensional P-spline smoothing for spatial analysis of plant breeding trials. *Biometrical Journal, 64*(5), 835-857.
- Verbyla, A. P. (2019). A note on model selection using information criteria for general linear models estimated using REML. *Australian & New Zealand Journal of Statistics, 61*(1), 39—50.
- 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*(4), 478-508.
- Welham, S. J. (2022) `TPSbits`: *creates structures to enable fitting and examination of 2D tensor-product splines using ASReml-R.* Version 1.0.0 https://mmade.org/tpsbits/

Thank you for your attention!