

---

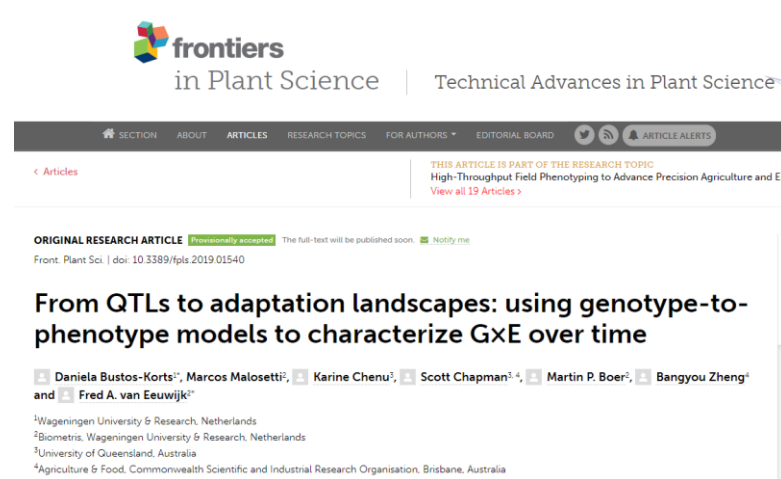
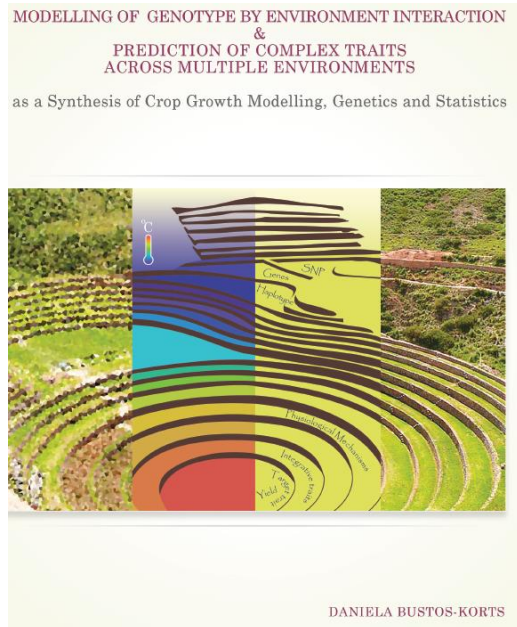
Making more of primary traits by incorporating secondary traits

6th Annual - Nordic Plant Phenotyping Network Workshop

Fred van Eeuwijk  
27 November, 2020

# Studying multi-trait dynamics by APSIM with genotype-specific parameters (Daniela Bustos Korts)

## Role of secondary phenotyping & genotype specific crop growth simulations in environmental characterization



ORIGINAL RESEARCH  
published: 27 November 2019  
doi: 10.3389/fpls.2019.01491



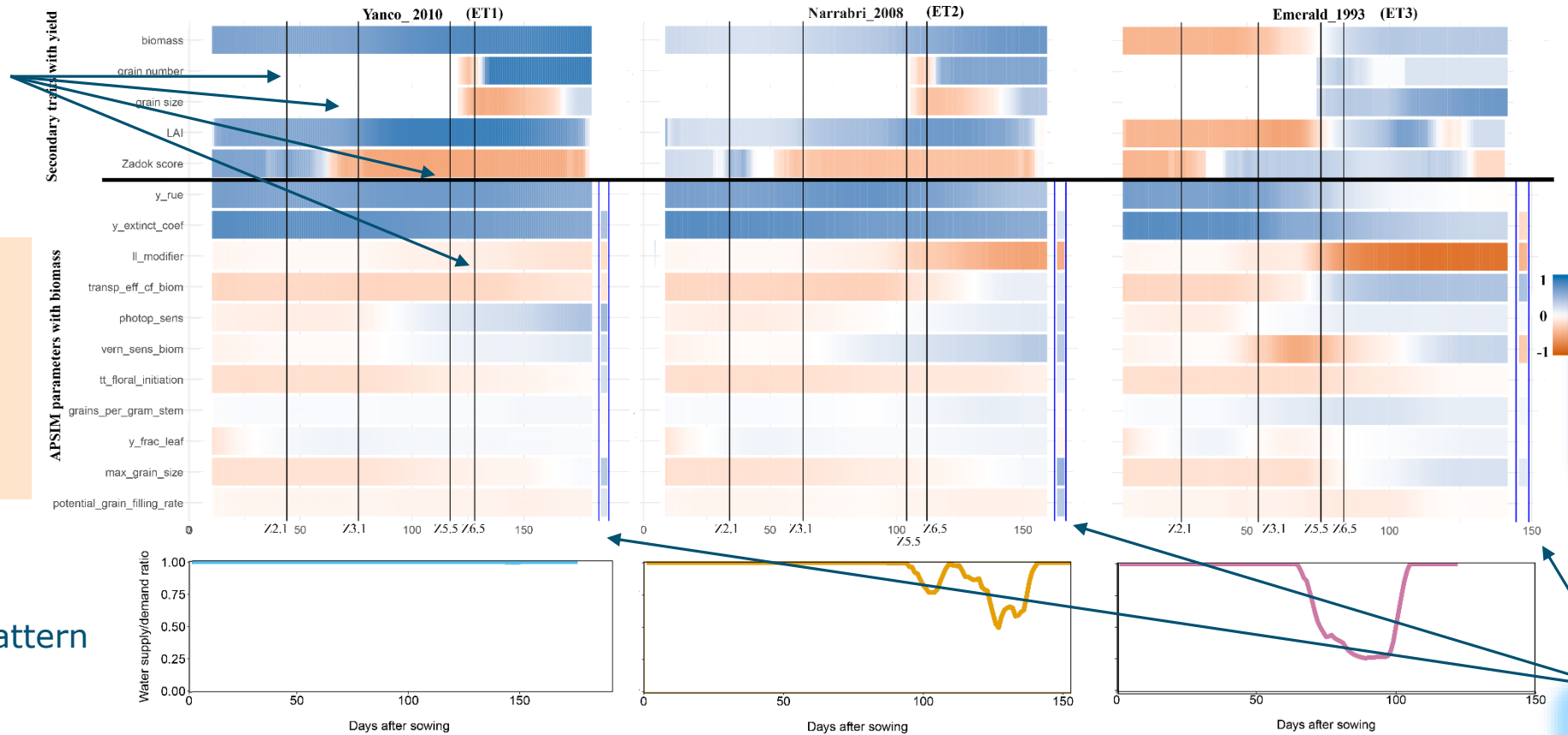
## Combining Crop Growth Modeling and Statistical Genetic Modeling to Evaluate Phenotyping Strategies

Daniela Bustos-Korts<sup>1\*</sup>, Martin P. Boer<sup>1</sup>, Marcos Malosetti<sup>1</sup>, Scott Chapman<sup>2,4</sup>, Karine Chenu<sup>3</sup>, Bangyou Zheng<sup>2</sup> and Fred A. van Eeuwijk<sup>1\*</sup>

# Correlations between target (yield), intermediate and physiological parameters **over time** (Australian wheat)

Delimitation of developmental stages

trait x trait x  
time x  
environment  
interactions



Correlations of **yield** with *intermediate traits*

Correlations of **biomass** with *physiological parameters*

Correlation of physiological parameters with **yield**

Note environment specificity of correlation patterns

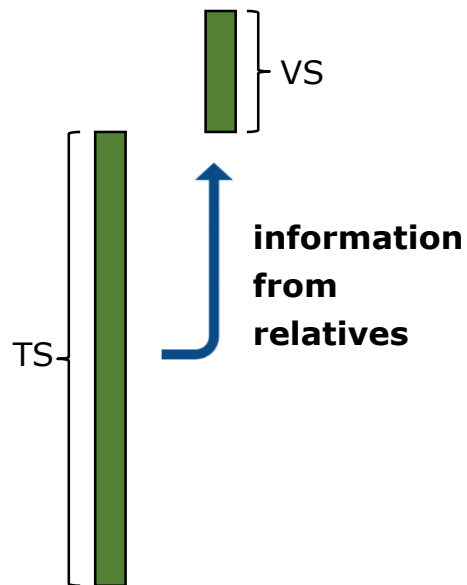
Correlation patterns between yield and yield components over time define environmental types

# Problem

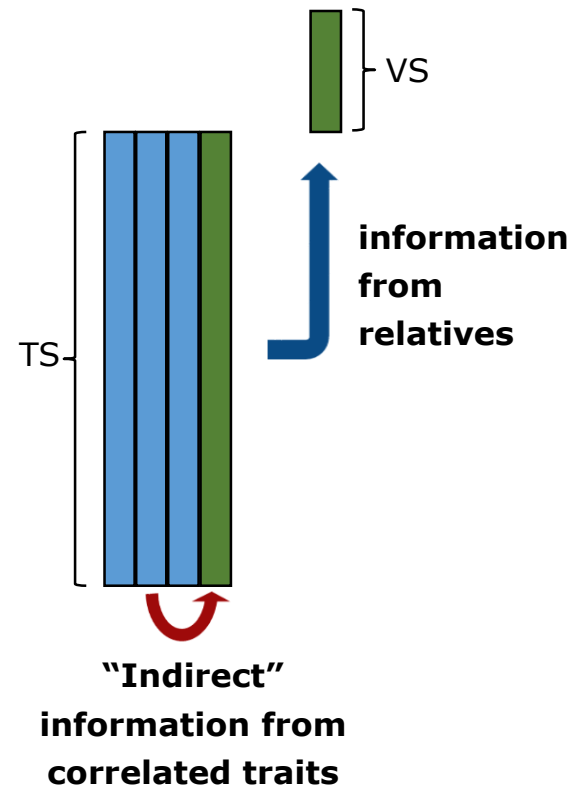
- Primary traits, like yield, can be predicted to a certain extent from molecular markers
- New phenotyping tools and devices produce large amounts of additional data on secondary phenotypes (spectra, root growth, biomass assessments)
- Can these secondary phenotyping data be useful to predict primary traits or made useful to better understand the genetic basis of primary traits (pleiotropic QTLs)?
  - Suggestion
    - Convert the secondary phenotyping data into genotypic covariables / responses and use them for prediction of primary traits / QTL mapping
  - Problems (secondary traits)
    - often come as time series
    - occur as discrete realizations of underlying continuous processes (wavelengths)
    - genetic signal requires adjustment as well as smoothing in multiple dimensions (spatial, temporal, wavelength)
    - no standard or default analysis methods available

# Prediction scenarios

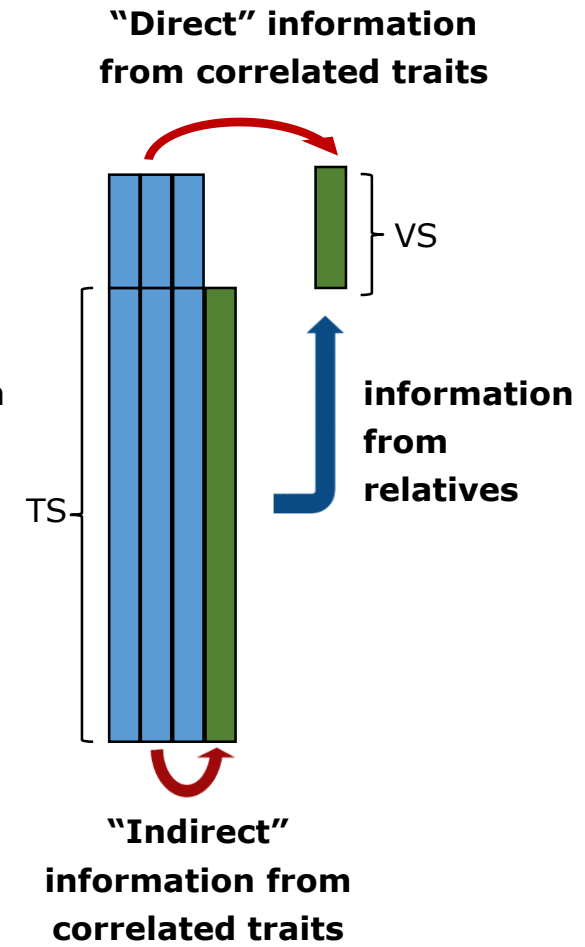
Single Trait



Multi Trait only  
in Training Set



Multi Trait in Training and  
Validation Set





## Design & analysis of phenotyping experiments

Fred van Eeuwijk & Emilie Millet

# Introduction JRA2: Design & analysis

- ▶ Develops tools for **statistical design and analysis of phenomics experiments** across platforms and scales of plant organization
- ▶ **Standardize statistical analyses and facilitate the combined analysis of data** coming from multiple installations and measurement scales (including field)
  - ▶ (i) outlier detection and data quality control
  - ▶ (ii) statistical design and analysis of single platform experiments
  - ▶ (iii) data integration between platforms and platform/field.
- ▶ Start EPPN<sup>2020</sup> clear demand for a unified set of tools and methods to analyze platform data. Diversity phenotyping techniques and increasing amount of data prevented direct application of designs, models and analysis methods originally developed for field trials
- ▶ Main objectives (strategic):
  - ▶ Level 1
    - ▶ Describe installations in a statistically intelligible way and give a motivation for design and analysis choices for the current state
  - ▶ Level 2
    - ▶ Description of the properties of the platform
    - ▶ Documented routine use of automated statistical methods and tools
    - ▶ Traceability and reversibility of decisions

## TJRA2.1 Outliers & data quality control

- ← Flagging different types of outliers
  - ← Point, time series, set of time series (plant)

## TJRA2.2 Design & analysis of single platform experiments

- ← Row-column blocking to neutralize spatial variation
- ← Choosing standard genotypes and allocating them
- ← Spatial adjustment of phenotypic traits at single and multiple time points
- ← Simultaneous analysis of multiple growth curves

## TJRA2.3 Data integration

- ← Analysis of multiple platform trials
- ← Trans platform EPPN<sup>2020</sup> experiments
- ← Combining analysis of field trials with analysis of platform trials
- ← Analysis of genotype x environment interactions
- ← Multi-trait analyses (platform & field)

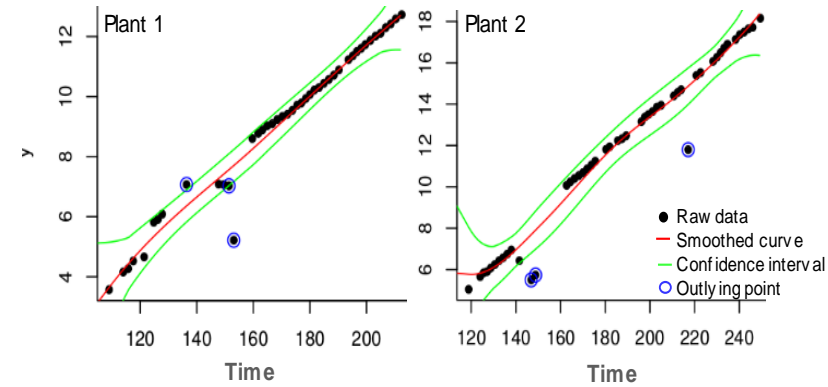


# JRA2: Design & analysis outputs

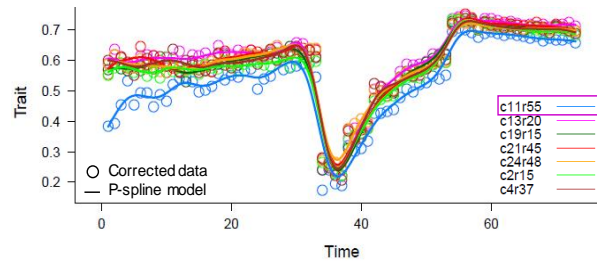
---

- ← Statistical methodology
- ← Software (R procedures)
- ← Web-applications (design generator)
- ← Instructions on how to apply the methodology
- ← Teaching material & courses
- ← Guidelines & recommendations
- ← Consultation & support

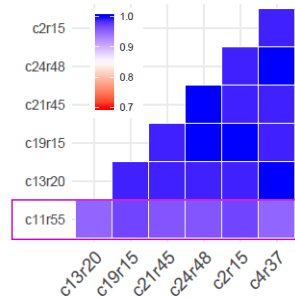
- ▶ Single points
- ▶ A time series
- ▶ A set of time series (plant)



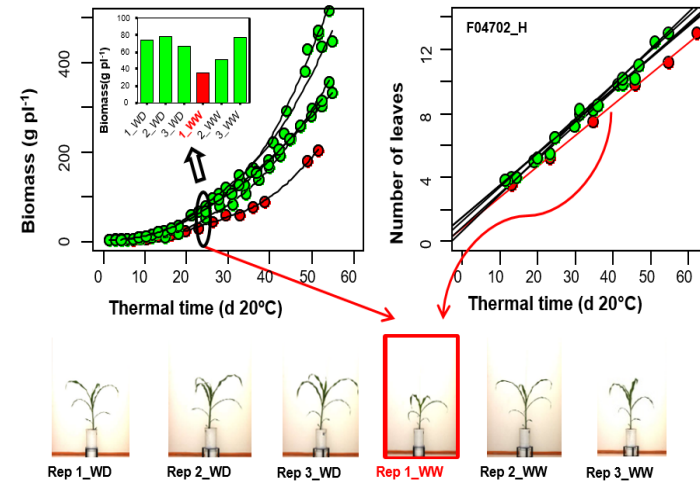
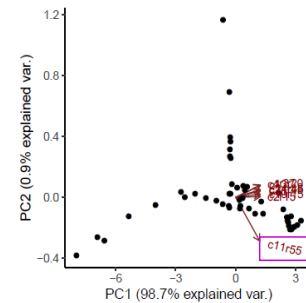
A – Corrected data and p-spline model



B – Pairwise correlation coefficient

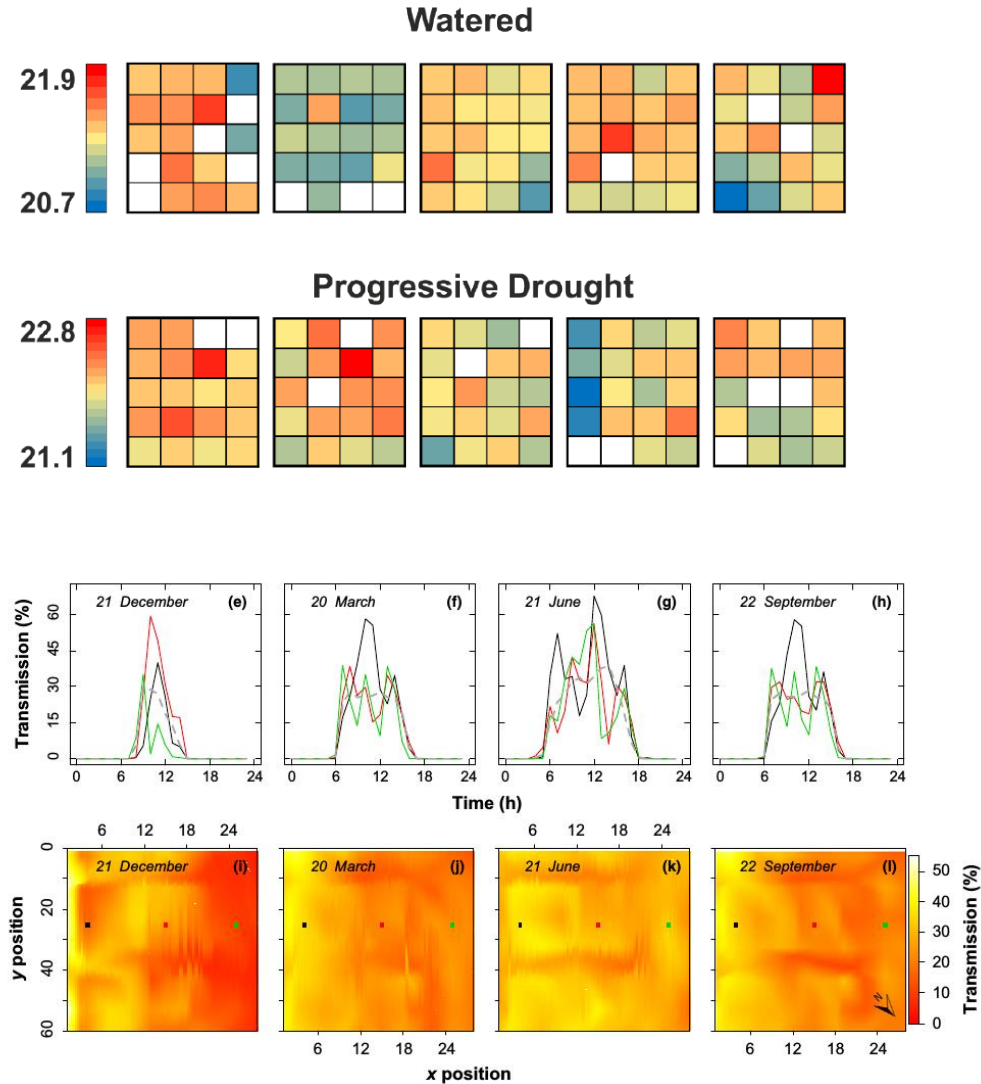


C- Principal Component Analysis



## Observed trends

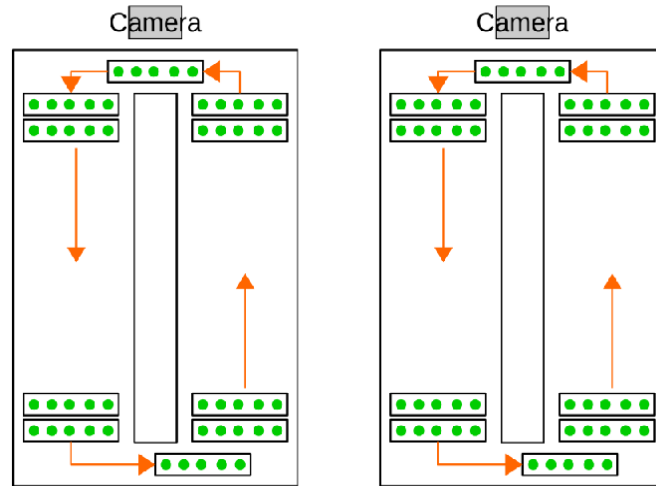
To counteract spatial trends on platforms, we can use row-column designs and perform spatial analyses



# Rhizotubes Dijon Resolvable Row-Column Design

21	29	2	26	4	22	16	6	1	30	14	12	3	5	10
33	17	25	7	5	13	14	8	30	23	16	24	2	1	11
16	12	30	19	9	21	23	7	20	32	33	15	17	31	29
20	7	26	1	25	29	9	12	11	17	27	28	24	4	6
9	13	6	27	24	15	29	19	18	31	22	2	7	14	26
32	31	19	14	18	8	25	21	26	5	4	6	33	15	20
14	24	27	28	23	12	15	3	5	25	10	20	18	13	16
3	10	8	30	2	20	24	4	13	11	9	26	32	12	19
15	28	11	32	16	10	2	27	33	7	18	1	8	22	21
5	1	23	6	33	17	32	22	31	21	13	3	9	28	27
4	22	18	31	11	3	17	28	10	29	8	19	23	25	30
22	20	15	5	17	11	27	30	28	12	2	23	14	32	4
8	32	29	16	12	14	3	33	4	22	25	13	19	27	7
23	5	33	29	10	2	19	15	12	24	6	31	1	11	25
24	30	4	3	32	31	1	10	17	14	15	7	26	18	33
11	19	1	4	27	9	21	13	7	3	20	18	5	29	23
31	18	3	24	6	1	8	2	14	27	19	16	12	21	17
10	27	16	8	26	33	20	31	29	9	5	4	28	24	2
17	25	14	15	19	7	11	5	9	26	28	21	22	16	13
6	26	12	20	13	18	22	23	16	8	17	11	10	30	31
7	2	28	22	21	25	18	24	6	33	30	32	20	9	3
13	21	9	23	30	28	26	25	32	10	1	29	6	8	15

# Rhizotron UCL Augmented Row Column Design



- Guidelines, software & support for choosing suitable statistical designs and corresponding statistical analyses
- Experimental design with row and column block structures to correct for spatial trends
- Choice and allocation of standard genotypes
- Unbiased estimates of treatment differences with good precision at acceptable costs

119	140	118	3	68
109	50	10	101	135
127	144	22	65	14
5	56	27	89	48
106	27	135	60	147
96	129	24	80	113
114	54	54	58	48
123	88	91	143	18
116	29	40	83	33
135	51	131	59	63
100	42	112	66	105
12	43	104	47	147
64	39	75	46	78
125	98	24	143	141
85	147	94	90	145
21	82	67	7	102
76	130	17	9	16
70	148	55	26	97
134	5	136	108	108
28	117	41	89	72
124	110	25	103	139
124	126	143	120	77
79	137	31	38	13
35	53	36	37	115
57	34	81	74	121
57	73	44	69	86
143	81	132	23	45
92	133	71	20	148
15	93	145	107	30
146	87	122	84	128
8	147	142	22	111
52	11	2	148	19
138	95	62	147	4
67	91	64	135	135
29	55	43	126	36
102	97	138	57	96
86	79	141	131	124
147	20	77	93	88
148	44	128	61	89
122	49	48	143	143
59	109	22	100	98
26	90	34	50	81
56	83	3	148	38
143	85	42	116	31
65	2	134	113	103
74	143	119	146	76
132	15	139	142	117
95	114	13	32	148
29	17	148	24	144
136	23	101	12	144
108	99	11	141	106
30	18	130	54	87
104	16	143	75	69
147	10	63	115	92
148	35	8	82	82
107	14	47	140	123
66	45	27	5	46
105	71	28	40	60
111	112	121	143	52
133	21	19	72	27
147	84	137	73	70
9	143	33	94	118
25	120	60	62	53
39	145	6	68	24
51	58	37	125	41
77	78	100	148	99
54	69	85	143	42
44	111	113	51	94
98	118	147	2	57
68	28	73	106	12
50	89	35	147	56
24	122	20	97	136
60	146	82	13	109
53	65	30	137	143
68	63	26	52	43
117	62	104	104	29
47	33	32	133	1
11	147	81	76	116
27	148	49	79	22
84	131	123	145	134
75	108	74	147	71
128	15	80	6	141
16	138	59	36	93
40	67	4	148	90
115	124	72	130	66
48	103	39	10	17
18	142	143	87	64
126	121	70	139	114
58	135	83	9	143
31	92	23	110	61
38	107	5	55	8
41	119	129	88	101
34	148	14	105	120
7	46	86	148	140
91	125	96	19	147

96	197	412	58	90	20	282	45	190	32	29	383	78	372	415	416	185	414	206	241	403	943	371	9	437		
409	62	270	347	92	29	25	77	404	140	136	417	105	201	418	283	265	247	2	231	406	9	242	237	338		
184	413	16	243	256	396	257	40	401	179	309	422	108	406	39	350	283	435	314	392	290	33	223	117	390		
43	230	112	411	302	418	172	345	15	63	164	191	403	412	358	130	353	420	204	46	224	329	149	402	362		
417	234	343	226	239	44	402	295	407	36	122	200	100	405	157	111	108	323	5	404	110	281	219	382	119	374	
398	406	414	158	276	390	380	286	95	152	413	306	265	317	187	232	215	332	401	4	279	47	250	412	12		
35	333	322	254	240	377	82	285	204	423	5	409	221	69	418	36	181	180	228	97	66	58	416	403	387		
102	292	261	418	251	261	435	236	410	17	65	373	161	64	186	307	411	142	340	406	369	66	214	135	43		
407	47	21	141	341	83	473	294	150	87	366	354	370	153	410	394	244	120	417	55	107	72	1	409	11		
96	13	61	178	383	211	101	405	227	379	258	259	284	400	305	225	54	170	146	416	303	381	411	59	419		
109	402	321	203	26	129	389	148	213	411	138	406	386	84	248	287	88	403	418	48	275	301	77	86	245		
217	384	423	193	165	112	271	196	416	122	194	1	5	364	238	249	31	315	210	162	216	346	404	15	1068	104	311
98	404	144	300	386	361	409	408	176	273	267	359	19	405	56	73	57	402	399	352	229	73	260	117	415	116	
7	367	213	156	71	337	156	316	414	412	263	420	75	212	190	255	16	437	357	74	126	391	123	413	418	116	
331	274	219	27	403	406	209	70	99	330	403	334	351	252	23	335	413	41	310	419	313	182	200	128	245	207	
312	325	375	43	416	209	114	318	417	360	368	444	51	166	411	4	342	291	10	134	80	401	324	205	91		
113	419	192	416	280	68	24	175	403	355	208	31	220	103	262	412	344	298	408	289	375	25	42	345	218		
320	205	355	52	401	14	416	289	189	163	404	204	365	116	418	406	393	40	103	186	33	296	84	76	736		
183	298	263	405	408	76	67	323	277	390	151	50	165	402	39	60	36	137	167	338	378	340	414	158	417		
223	127	385	151	272	22	598	121	273	397	145	171	335	407	39	65	8	118	253	418	413	240	132	124	174		
277	17	300	422	284	416	412	369	409	373	419</																

# Spatial analysis by 2D spline surface

Thèse Appl Genet  
DOI 10.1007/978-94-007-2894-4



ORIGINAL ARTICLE

## Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model

Julio G. Velasco<sup>1,2</sup> · María Xosé Rodríguez-Álvarez<sup>1,4</sup> · Martin P. Boer<sup>1</sup> · David R. Jordan<sup>3</sup> · Paul H. C. Eilers<sup>5</sup> · Marcos Makosetti<sup>1</sup> · Fred A. van Eeuwijk<sup>1</sup>

Spatial Statistics 23 (2018) 52–71



Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: [www.elsevier.com/locate/spasta](http://www.elsevier.com/locate/spasta)



## Correcting for spatial heterogeneity in plant breeding experiments with P-splines

María Xosé Rodríguez-Álvarez<sup>a,b,\*</sup>, Martin P. Boer<sup>c</sup>, Fred A. van Eeuwijk<sup>c</sup>, Paul H.C. Eilers<sup>d</sup>

<sup>a</sup> BCAM — Basque Center for Applied Mathematics, Alameda de Mazarredo, 14. E-48009 Bilbao, Basque Country, Spain

<sup>b</sup> IKERBASQUE, Basque Foundation for Science, Bilbao, Spain

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

<sup>d</sup> Erasmus University Medical Centre, Rotterdam, The Netherlands



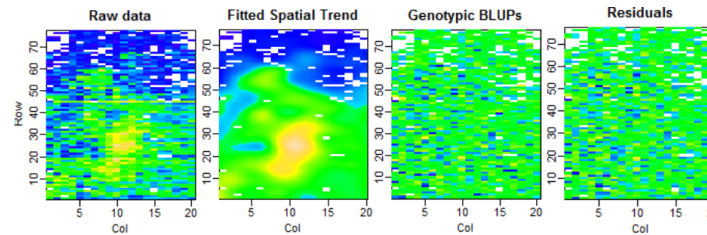
$$y = f(r, c) + G + e$$

genetic effects

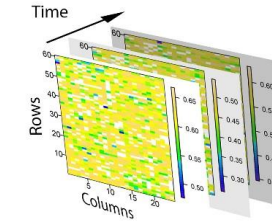
ANOVA-type decomposition of the smoothed surface:

$$\hat{f}(r, c) \equiv f_r(r) + f_c(c) + f_{r|c}(r, c)$$

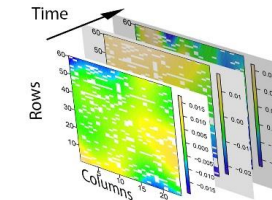
$$y = f(r, c) + G + e$$



## A Raw phenotypic data

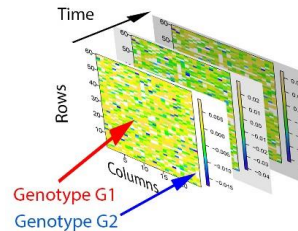


## B Fitted spatial trend



Correction fro spatial effects:  
- Blocks  
- Spatially-dependent error  
- AR1⊗AR1 (Gilmour et al., 1997)  
- 2D-splines  
(Rodríguez-Álvarez et al., 2018,  
Velasco et al., 2017)

## C Genotypic values (BLUPs)





## Modelling steps

Choosing the experimental design



Size of  
input data

Model  
dependence  
of output

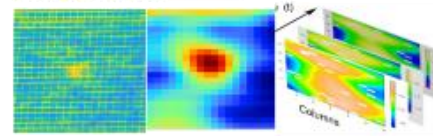
Level 1

Feature extraction



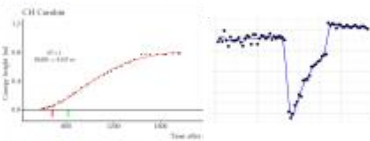
Level 2

Correcting for design factors  
and spatial trends



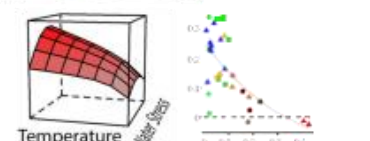
Level 3

Dynamical modelling



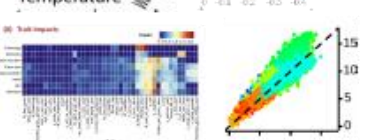
Level 4

Modelling dependence on  
environmental gradients



Level 5

Target trait prediction



ELSEVIER

Contents lists available at ScienceDirect

Plant Science

journal homepage: [www.elsevier.com/locate/plantsci](http://www.elsevier.com/locate/plantsci)



Modelling strategies for assessing and increasing the effectiveness of new phenotyping techniques in plant breeding

Fred A. van Eeuwijk<sup>a,\*</sup>, Daniela Bustos-Korts<sup>a</sup>, Emilie J. Millet<sup>a</sup>, Martin P. Boer<sup>a</sup>, Willem Kruijer<sup>a</sup>, Addie Thompson<sup>b,1</sup>, Marcos Malosetti<sup>a</sup>, Hiroyoshi Iwata<sup>c</sup>, Roberto Quiroz<sup>d</sup>, Christian Kuppe<sup>e</sup>, Onno Muller<sup>e</sup>, Konstantinos N. Blazakis<sup>f</sup>, Kang Yu<sup>g,h</sup>, Francois Tardieu<sup>i</sup>, Scott C. Chapman<sup>j,k</sup>



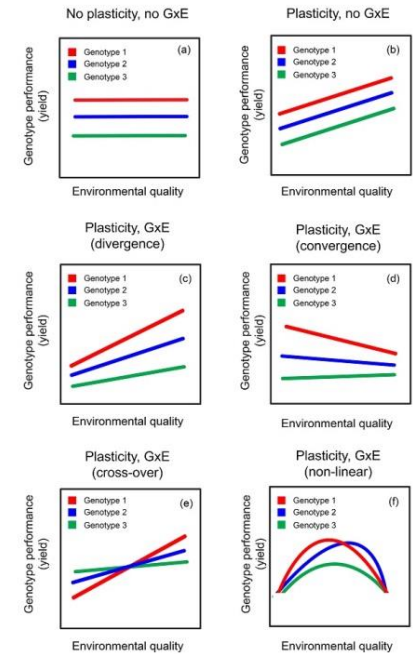
# Baseline genotype to phenotype model includes GxE

Published August 30, 2016

RESEARCH

## What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype $\times$ Environment Interactions?

Fred A. van Eeuwijk,\* Daniela V. Bustos-Korts, and Marcos Malosetti



- Phenotype =
  - Genotype +
  - Environment +
  - Genotype by Environment Interaction +
  - Error
- Two-way ANOVA, fixed genotypes and environments, GxE fixed (lack of fit) term
  - $y_{ij} = \mu + g_i + e_j + \textcolor{red}{g}e_{ij} + \epsilon_{ij}; \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
  - Subscript  $i$  for genotype,  $j$  for environment
- Mixed model formulation, random genotypes, GxE as heterogeneity of genetic variances and correlations
  - $y_{ij} = \mu_{ij} + \textcolor{red}{G}_{ij} + \epsilon_{ij}; \text{VCOV}(y_{ij}) = \Sigma_{gge} + R_\epsilon$

# GxE and separability

- $y_{ij} = \mu_{ij} + \underline{G}_{ij} + \underline{\epsilon}_{ij}$ 
  - Multi-environment (multi-trait) model
- Mean  $\mu_{ij}$ 
  - $\mu_{ij} = \mu(\boldsymbol{\beta}_i^g, \boldsymbol{\beta}_j^e, \mathbf{x}_i, \mathbf{z}_j) = \sum_m a_{mi} b_{mj}$ 
    - $\boldsymbol{\beta}_i^g$ : genotypic sensitivities
    - $\mathbf{x}_i$ : genotypic covariables / molecular markers / groupings
    - $\boldsymbol{\beta}_j^e$ : environment related quality / QTL
    - $\mathbf{z}_j$ : environmental covariable / index / classification / management
- Variance-Covariance matrix  $\underline{G}_{ij}$ :  $VCOV(\underline{G}_{ij})$ ;
  - $VCOV(\underline{G}_{ij}) = \sum_v \mathbf{A}_v^g \otimes \mathbf{B}_v^e$ 
    - $VCOV(\underline{G}_{ij}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^g \otimes \boldsymbol{\Sigma}^e$
  - Structuring  $\underline{G}_{ij}$  on genotypic dimension
    - Markers (GRM), pedigree, omics, phenotyping, grouping
  - Structuring  $\underline{G}_{ij}$  on environmental dimension
    - Weather, soil, management, indices, classifications

# Product models for GxE (separable)

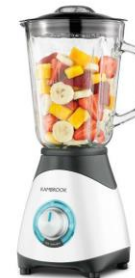
- Finlay-Wilkinson / AMMI / GGE / SREG; implicit genotypic and environmental characterization
  - $\gamma_i \delta_j + \varepsilon_{ij}$
- **Explicit genotypic characterization**: QTL mapping (markers) & **secondary phenotyping** (platform / field)
  - $x_i \beta_j^e + \varepsilon_{ij}$
  - $y_i^* \beta_j^e + \varepsilon_{ij}$
  - $(y_i^*)^t \beta_j^e + \varepsilon_{ij}$
  - $(y(x)_i^*)^t \beta_j^e + \varepsilon_{ij}$
- Explicit environmental characterization (factorial regression)
  - $\beta_i^g z_j + \varepsilon_{ij}$
  - $\beta_i^g s(z_j) + \varepsilon_{ij}$
- Modelling using genotypic and environmental covariates
  - $\beta(x_i) z_j + \varepsilon_{ij} = \kappa x_i z_j + \beta_i^* z_j + \varepsilon_{ij}$
  - $\beta(y_i^*) z_j + \varepsilon_{ij} = \kappa y_i^* z_j + \beta_i^* z_j + \varepsilon_{ij}$



# Kernel methods

## ■ Generalized GE-BLUP

- $y_{ij} = \mu_{ij} + G_{1ij} + G_{2ij} + \epsilon_{ij}$
- $VCOV(G_{1i}) = \Sigma^{g_1} \otimes \Sigma^e$  with  $\Sigma^{g_1} = K_1 \sigma_{G_1}^2$  and  $K_1$  a function of markers (genotypes)
- $VCOV(G_{2i}) = \Sigma^{g_2} \otimes \Sigma^e$  with  $\Sigma^{g_2} = K_2 \sigma_{G_2}^2$  and  $K_2$  a function of omics / secondary phenotyping, defined on genotypes
- $\Sigma^e$  is a structuring matrix defined on the environments



# Introducing a time component in GxE models

## 2-way P-spline ANOVA and 3-level hierarchical P-spline model

### ■ Martin Boer: P-spline 2-way ANOVA

- $y_{ij}(t) = \mu(t) + f_j^E(t) + f_i^G(t) + f_{ij}^{G \cdot E}(t) + \varepsilon_{ij}(t)$
- $y_{ij}(t) = \{\mu + \text{lin}(t) + \text{spl}(t)\} + \{g + e + g \cdot e\} \cdot \{\text{lin}(t) + \text{spl}(t)\} + \varepsilon_{ij}(t)$ 
  - Sparse mixed model equations, special choice of basis functions
  - Wood. *Generalized additive models: an introduction with R*. CRC press, 2017.

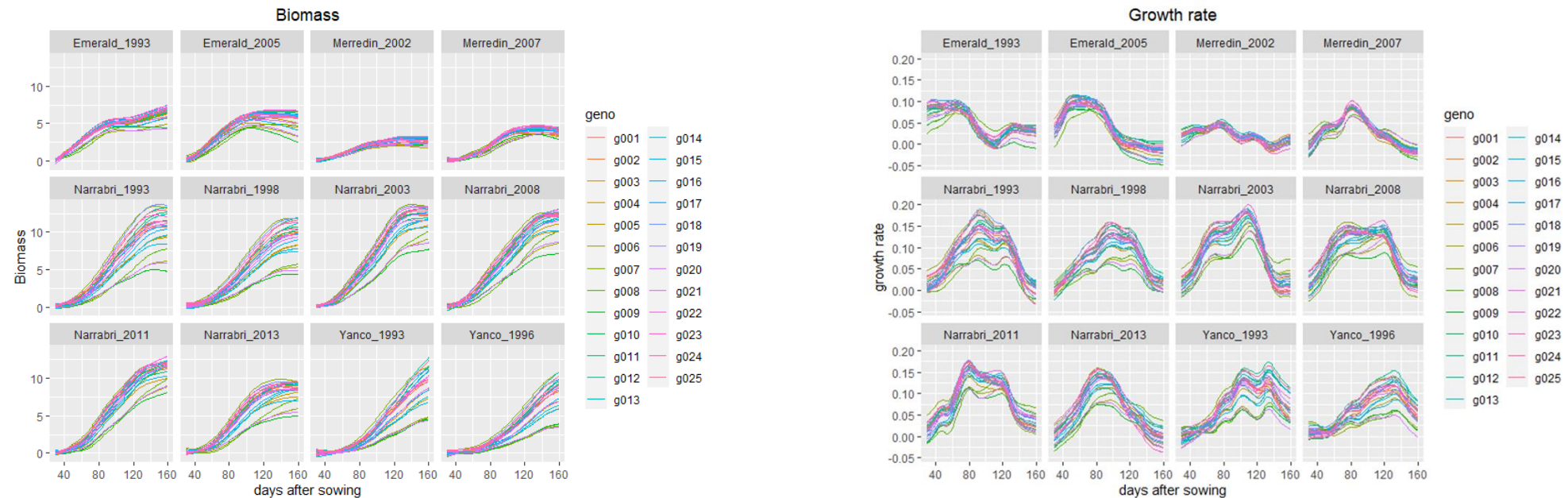
### ■ Diana M. Perez & Maria Xose Alvarez-Rodriguez

- $y_{ijk}(t) = \mu(t) + f_j^E(t) + f_{ij}^{E \cdot G}(t) + f_{ijk}^{E \cdot G \cdot P}(t) + \varepsilon_{ijk}(t)$
- Choice of number of basis splines critical
  - Brumback and Rice. *JASA*, 1998; Durban, Harezlak, Wand and Carroll. *Stat Med*, 2005.

# Dynamical predictions of biomass (G+E+G.E) and first derivatives

Data: APSIM wheat simulation – Daniela Bustos Korts

Analysis: Martin Boer

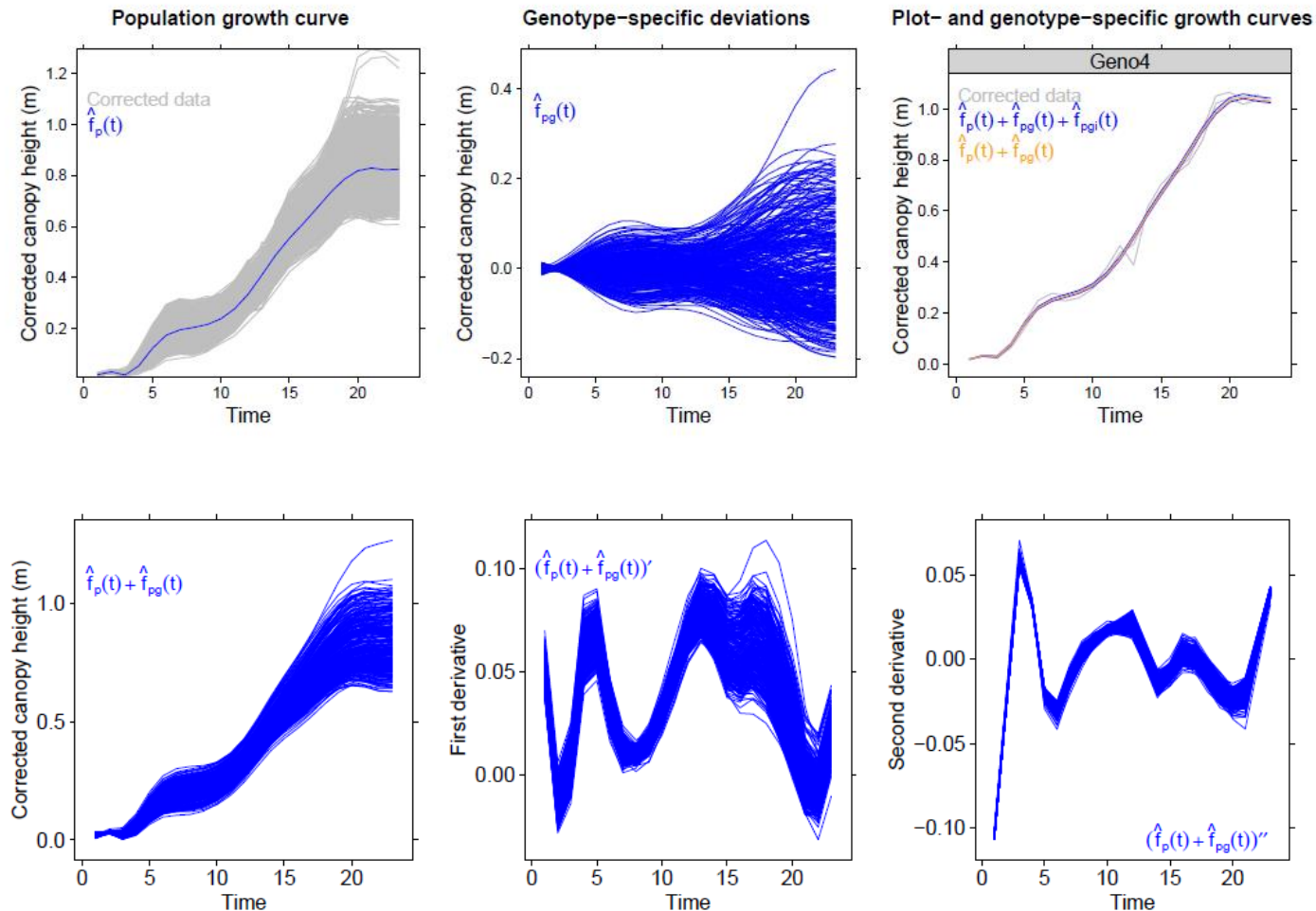


$$y_{ij}(t) = \mu(t) + f_G(t) + f_E(t) + f_{G \times E}(t) + \varepsilon_{ij}(t)$$

# Illustration 3 level hierarchical splines and derivatives

Analysis: Diana M. Perez & Cote Alvarez Rodriguez (BCAM – Bilbao)

Data: wheat - group Andreas Hund (ETH Zurich)

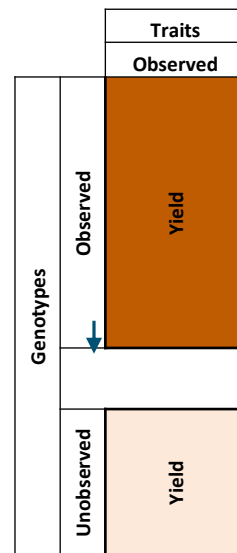


- Curves
  - population (prediction)
  - genotype (deviation)
  - plant / plot (prediction)
- Genotype prediction
  - first derivative
  - second derivative

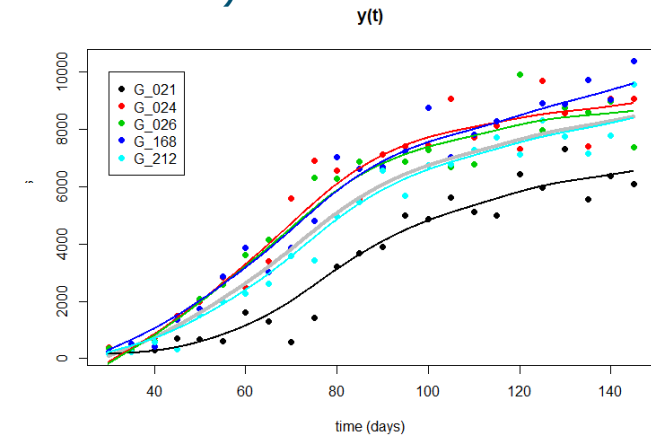
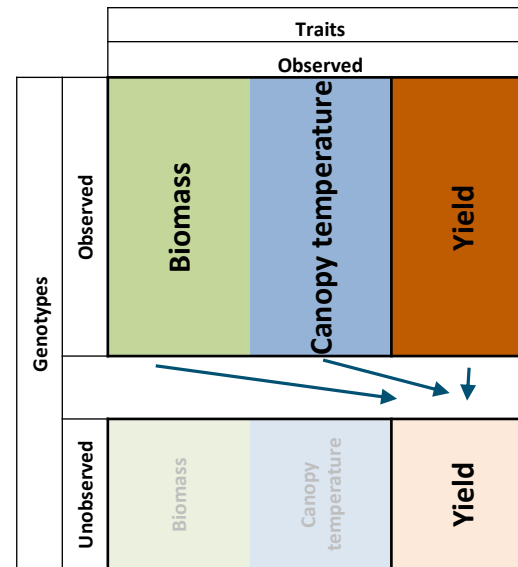
# Data integration

Use dynamical parameters or full functional information of secondary phenotypes in a better prediction of target traits like yield (including GxE) (Daniela Bustos-Korts)

## Single-trait



## Multi-trait



$$\underline{y}(t)_{ij} = \mu(t)_j + \underline{G}(t)_{ij} + \underline{\epsilon}(t)_{ij}$$



Technical Advances in Plant Science®

ORIGINAL RESEARCH ARTICLE | [Preprint only accepted](#) | The full-text will be published soon. [Notify me](#)  
Front. Plant Sci. | doi: 10.3389/fpls.2019.01491

## Combining crop growth modelling and statistical genetic modelling to evaluate phenotyping strategies

[Daniela Bustos-Korts](#)<sup>1\*</sup>, [Martin P. Boer](#)<sup>1</sup>, [Marcos Malosetti](#)<sup>2</sup>, [Scott Chapman](#)<sup>2,3</sup>, [Karine Chenu](#)<sup>1</sup>, [Banyou Zheng](#)<sup>1</sup> and [Fred A. Van Eeuwijk](#)<sup>1\*</sup>

<sup>1</sup>Biometris, Wageningen University & Research, Netherlands

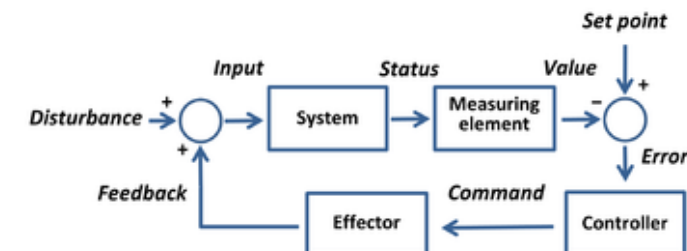
<sup>2</sup>University of Queensland, Australia

<sup>3</sup>Agriculture & Food, Commonwealth Scientific and Industrial Research Organisation, Brisbane, Australia



# Modelling principles: phenology and multiple traits

- Multiple kernel models
  - integrate genomic information with omics, phenotyping and environmental information
  - usually little attention for feature selection
- Feature selection in the form of characterization of genotypes and environments always pays off (mean and VCOVs)
  - adaptation & maturity groups (phenology)
  - eco zones, mega environments, environmental scenarios
- Yield = integrated output of processes triggered by environmental and internal stimuli following interactions between upstream traits starting from a DNA basis
  - Predicting GxE requires multi-trait modelling
  - Phenotyping platforms give access to many traits
- Prediction models for GxE need to include phenology/ timing of developmental events implicitly or explicitly
  - phenotyping platforms / devices provide temporal information
  - feature selection needs to respect phenology, scalability and granularity



Review

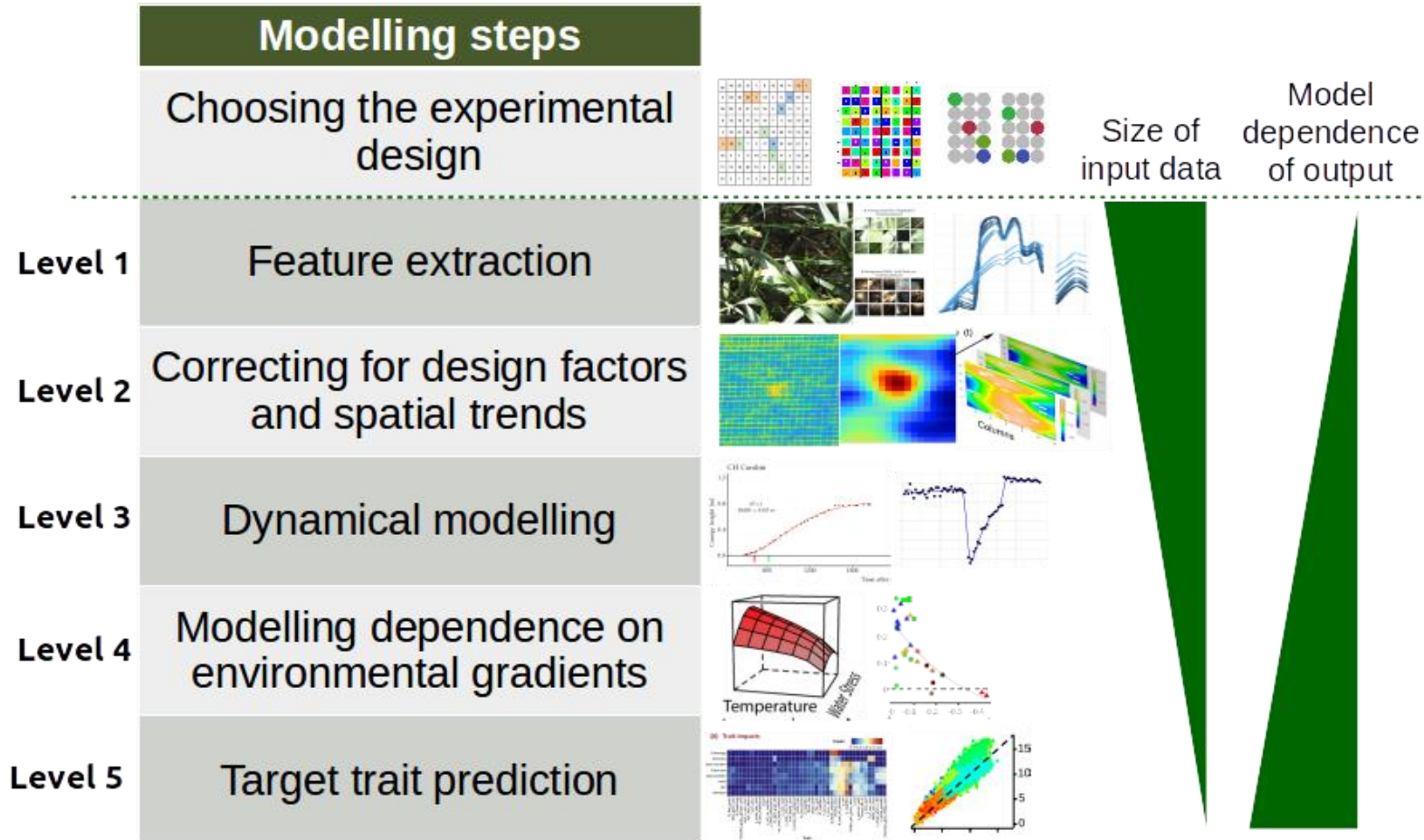
TRENDS in Plant Science Vol.11 No.12

Full text provided by www.sciencedirect.com  
ScienceDirect

## Models for navigating biological complexity in breeding improved crop plants

Graeme Hammer<sup>1</sup>, Mark Cooper<sup>2</sup>, François Tardieu<sup>3</sup>, Stephen Welch<sup>4</sup>, Bruce Walsh<sup>5</sup>, Fred van Eeuwijk<sup>6</sup>, Scott Chapman<sup>7</sup> and Dean Podlich<sup>2</sup>

# Wrap up



## Statistical approaches / tools

- Design
- Data quality control
- Single experiment analysis
  - Longitudinal data analysis
- Data integration
  - Primary trait prediction
  - Multi-experiment analysis
  - Multi-trait hierarchical models



Contents lists available at ScienceDirect

Plant Science

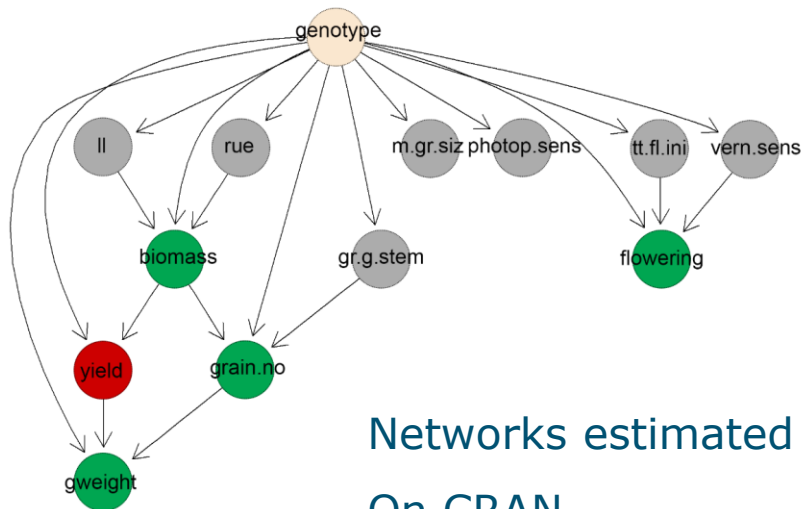
journal homepage: [www.elsevier.com/locate/plantsci](http://www.elsevier.com/locate/plantsci)



# Yield as part of a trait network with a genetic root

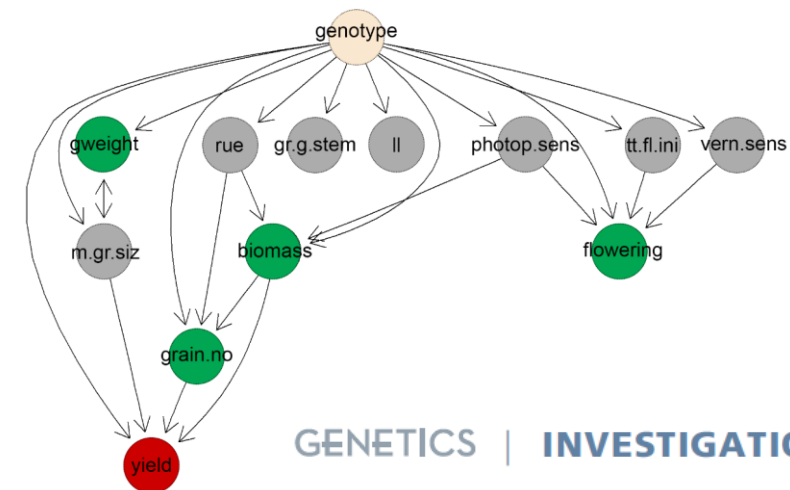
- Networks help visualizing genetic and phenotypic trait relationships (biomass plays a central role for yield)
- Topologies / adaptive mechanisms differ between environmental types with an important role for phenology
- Trait organisation shows '**modules**' that are **environment type** dependent, such modules can guide feature selection in prediction models (including kernel models)
- Elucidation of direct and indirect genetic effects on target traits

## Drought



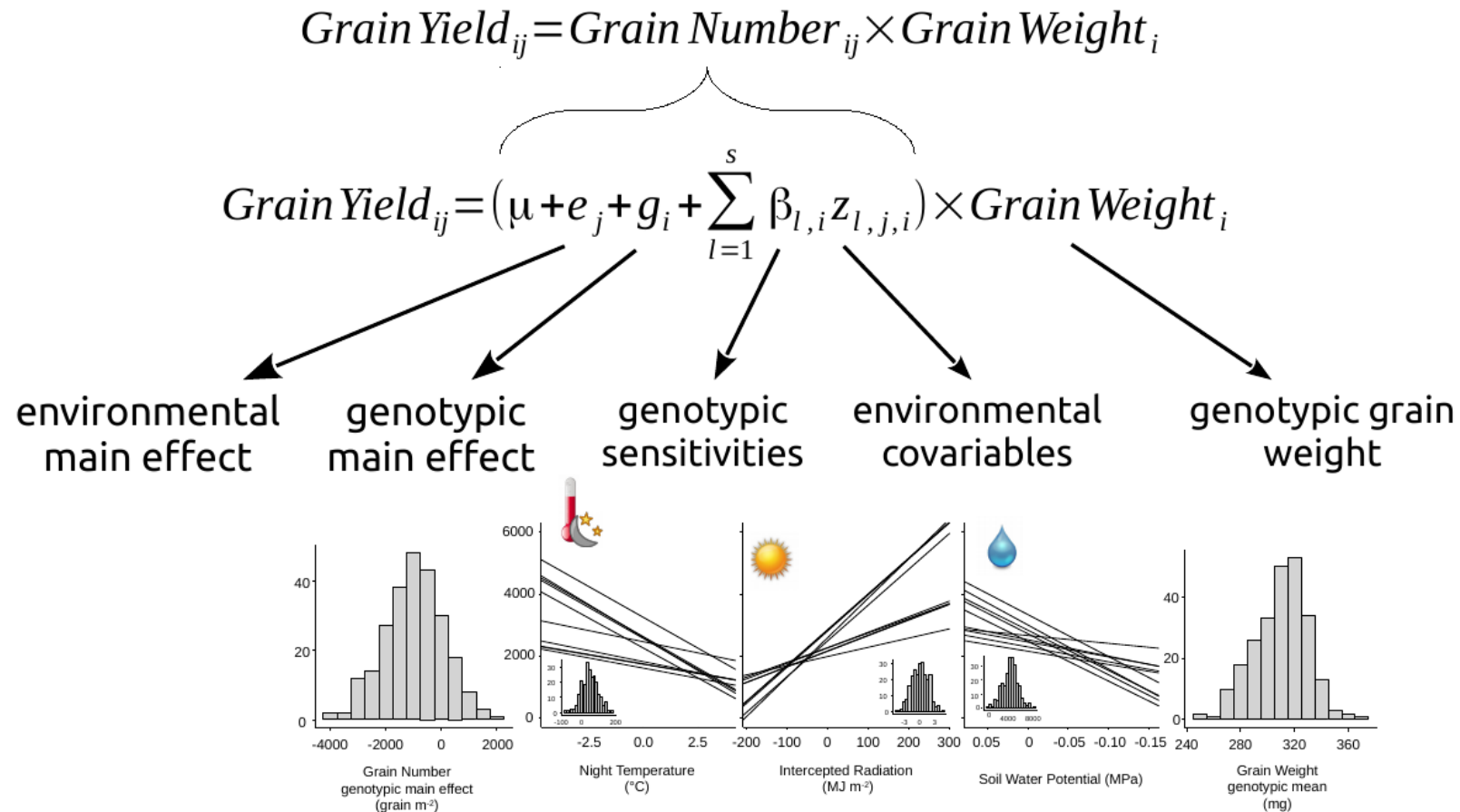
Networks estimated with **pcgen** algorithm  
On CRAN

## Non drought



Reconstruction of Networks with Direct and Indirect  
Genetic Effects

# Data integration



⇒ Can be predicted using marker profile and/or environmental indices