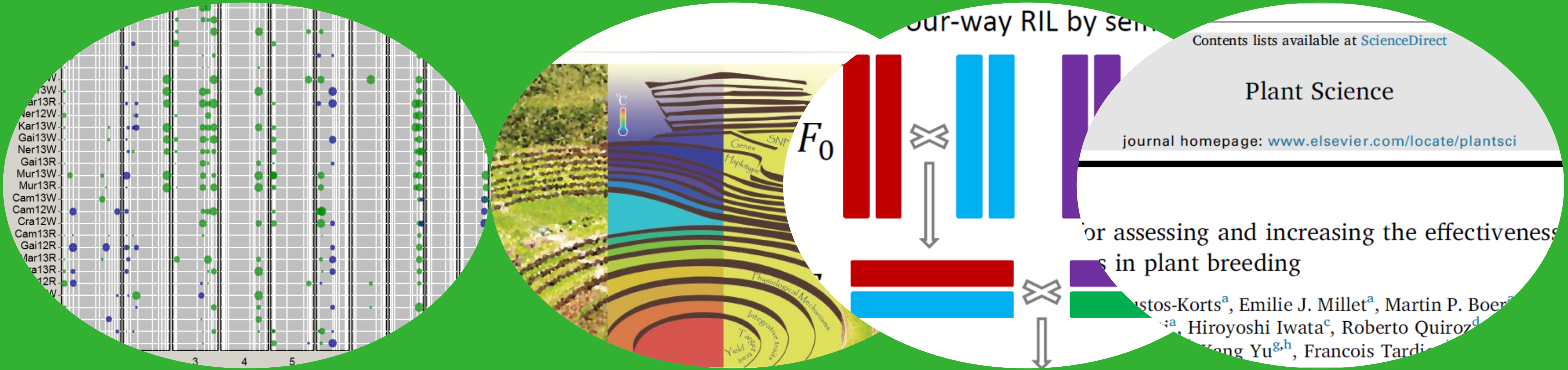
Statistical genetics in plants:
Finding an equilibrium between academic and commercial interests

Biostatistical Challenges in R&D: joint BMS-ANed and PSDM meeting

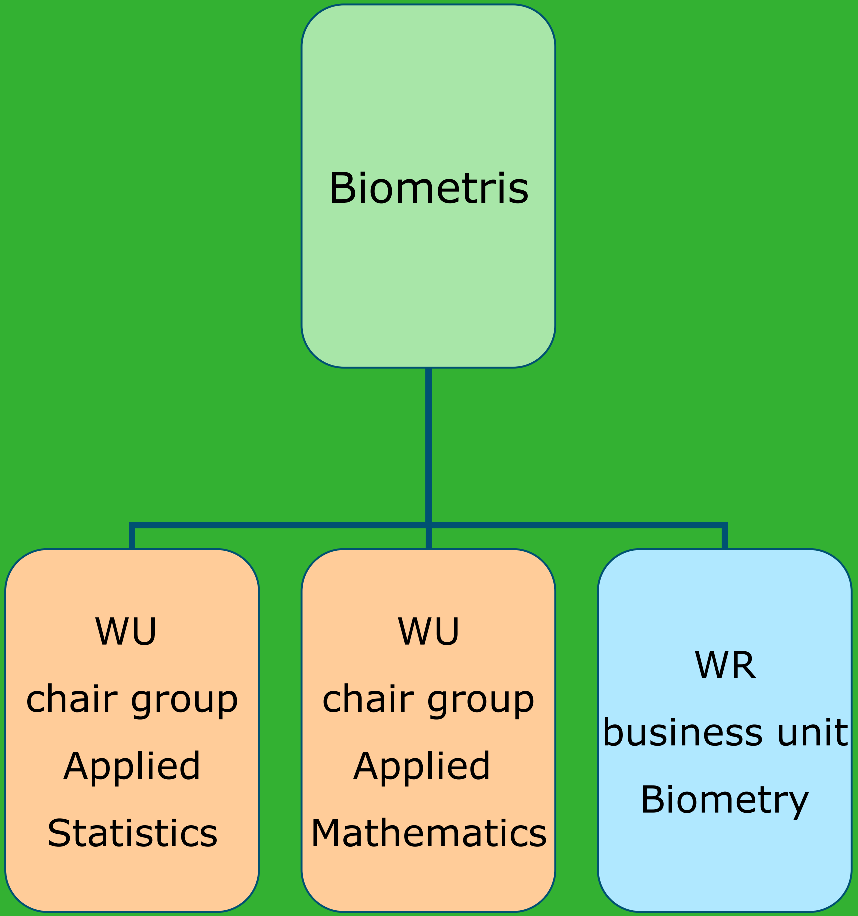
Fred van Eeuwijk, 23 November, Wageningen



WAGENINGEN UR
For quality of life

Organisation

Biometris: Structure



Biometris

WU chair group Applied Statistics

WU chair group Applied Mathematics

WR business unit Biometry

Quantitative methods brought to life

- WR permanent: +/- 15 fte
- WU permanent: +/- 30 fte
- DLO/WU temporary: 10 fte
- PhD: 25 & post doc: 5

Working area and mission

Quantitative methods brought to life

- Mission
  - Provide a solid quantitative framework to support and facilitate education and research in life and environmental sciences as relevant to WUR:
    - Plant / Animal / Food / Environment / Social science
- Main areas of activity
  - Education (JM, FvE)
  - Statistical genetics and genomics (FvE, MB)
  - Analysis of omics data (RW, FvE)
  - Systems biology (JM, PdR)
  - Food & Health (HV, HB)
  - Consultation (PG, SB)

Objective

- Centre of expertise for development and application of quantitative methods for life and environmental sciences (academic and commercial)
  - Design and analysis of experimental and observational studies
  - Modeling of static and dynamical behavior of life and environmental systems
- Centre of education for quantitative methods, knowledge transfer at various levels (BSc, MSc, PhD, post doc)
  - WUR (applied maths and stats courses)
  - Netherlands (MSc Statistical Science)
  - IAMZ (Mediterranean Institute for Agronomy Zaragoza)
  - Agro-food/ life science industry (courses on demand for professionals)
  - Developing world (BMGF, GCP, NUFFIC, FAO)

Realization

- Translation of stats and maths results to practical contexts
  - close collaborations
    - WUR groups (BU Bioscience, Plant Breeding, ASG,AFSG, TIFN)
    - National consortia and institutes (CBSG, NMC, NCSB, RIVM, LUMC)
    - Foreign research organizations (BIOSS, INRA, Max Planck, VIB)
    - Coordination of national (STW) international projects (EU, GCP)
    - Companies (Corteva/ DuPont-Pioneer, KWS, Nunhems, SES v/d Have, Syngenta, Rijk Zwaan, Enza, Danone)
- Education
  - create enthusiasm for quantitative methods (MSc, PhD, post doc)
  - collaboration with other research groups through joint supervision of BSc and MSc students
  - participation in MSc Statistical Science with Leiden University

Structure of projects / collaborations

- Company / institute
    - Multidisciplinary team with substance matter specialists and quantitative experts / information specialists
- Biometris
    - Key person for day to day work (PhD, post doc, staff) that is embedded in an advisory/ participatory group of 2-4 people

Statistical genetics and genomics

WAGENINGEN UR
For quality of life

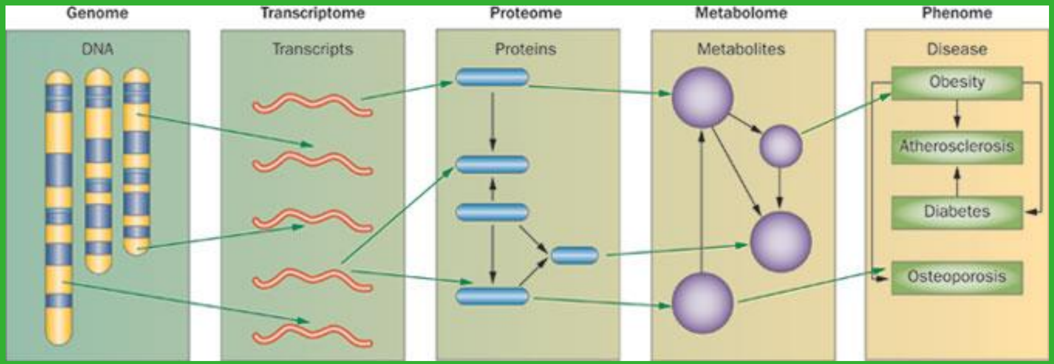
Starting point

- Plant breeding: improve yield, quality, sustainability for new cultivars by
  - Phenotypic selection
  - Genotype-to-phenotype (G2P) models with as input
    - Other phenotype(s)
    - DNA markers
    - Omics
    - Environmental descriptiojns
- Dynamic phenotypes as collected by new phenotyping techniques require an integration of
  - Static non-hierarchical (multi-trait) statistical genetic models with dynamic multi-trait hierarchical crop growth models

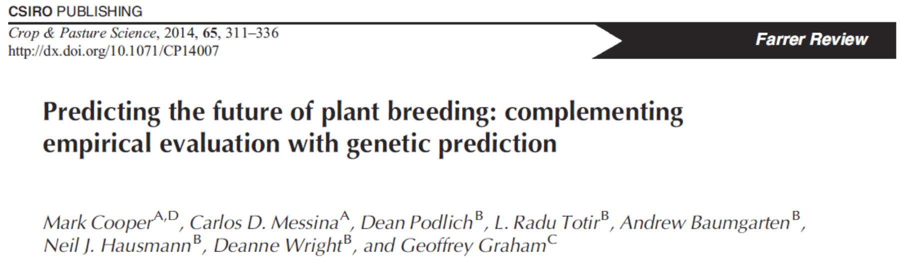
Genotype to Phenotype (G2P) Models

Models to predict phenotypes from genotypes (markers)

Then select best genotype



System Genetics



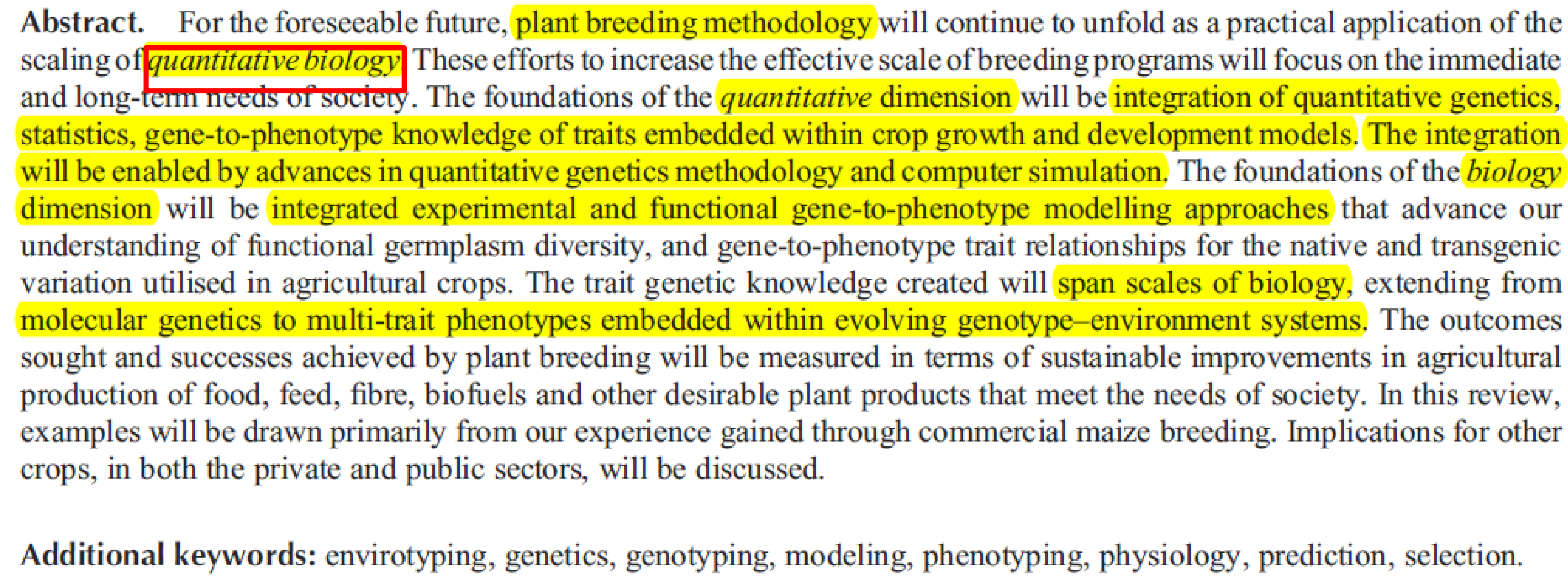
CSIRO PUBLISHING

Crop & Pasture Science, 2014, **65**, 311–336
http://dx.doi.org/10.1071/CP14007

Farrer Review

Predicting the future of plant breeding: complementing empirical evaluation with genetic prediction

Mark Cooper[A,D], Carlos D. Messina[A], Dean Podlich[B], L. Radu Totir[B], Andrew Baumgarten[B], Neil J. Hausmann[B], Deanne Wright[B], and Geoffrey Graham[C]



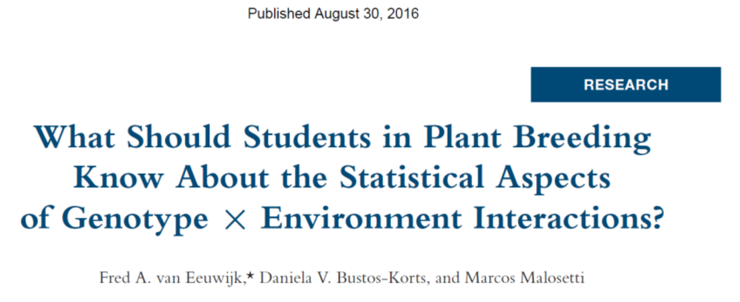
Abstract. For the foreseeable future, plant breeding methodology will continue to unfold as a practical application of the scaling of *quantitative biology*. These efforts to increase the effective scale of breeding programs will focus on the immediate and long-term needs of society. The foundations of the *quantitative* dimension will be integration of quantitative genetics, statistics, gene-to-phenotype knowledge of traits embedded within crop growth and development models. The integration will be enabled by advances in quantitative genetics methodology and computer simulation. The foundations of the *biology* dimension will be integrated experimental and functional gene-to-phenotype modelling approaches that advance our understanding of functional germplasm diversity, and gene-to-phenotype trait relationships for the native and transgenic variation utilised in agricultural crops. The trait genetic knowledge created will span scales of biology, extending from molecular genetics to multi-trait phenotypes embedded within evolving genotype–environment systems. The outcomes sought and successes achieved by plant breeding will be measured in terms of sustainable improvements in agricultural production of food, feed, fibre, biofuels and other desirable plant products that meet the needs of society. In this review, examples will be drawn primarily from our experience gained through commercial maize breeding. Implications for other crops, in both the private and public sectors, will be discussed.

Additional keywords: envirotyping, genetics, genotyping, modeling, phenotyping, physiology, prediction, selection.

Use phenotypic, genetic and genomic information optimally together to increase genetic gain

GxE concepts!

- *Reaction norm*
  - Genotype-specific relationship of the phenotype as a function of genotype, time, environmental gradient(s) and further phenotypes
- Separability
  - GxE = f(G;E)
- Genotypes
  - Adaptedness
  - Adaptability / *Sensitivity*
  - Stability / Risk / Resilience
  - Target population (TPG) / Sample
- Environments
  - *Environmental drivers* of GxE
  - *Environmental characterization*
  - Target population (TPE) / Sample

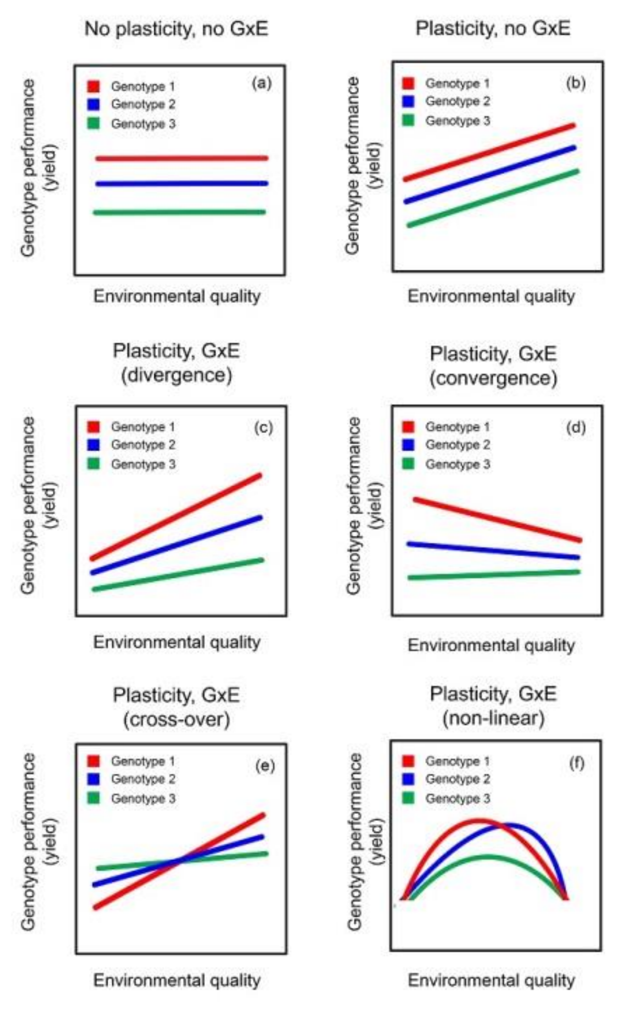


Published August 30, 2016

RESEARCH

What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions?

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



WAGENINGEN UR
For quality of life

Basic G2P models in plant breeding and GxE

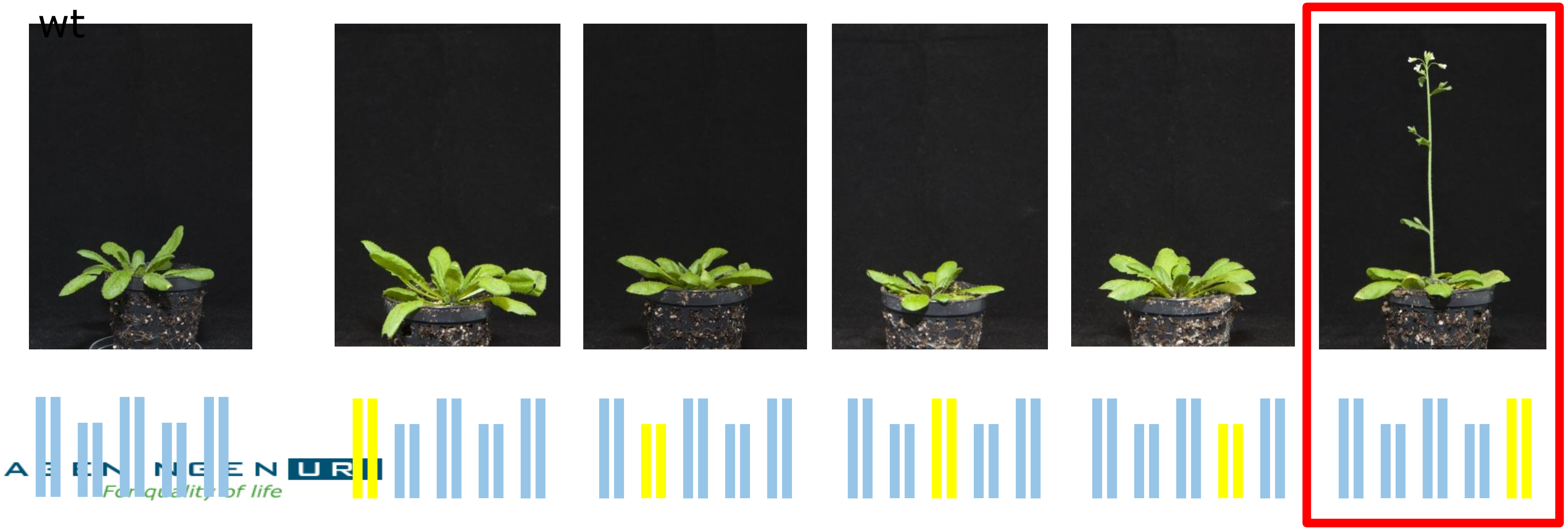
- Phenotype =
  - Genotype +
  - Environment +
  - Genotype by Environment Interaction +
  - Error
- Two-way ANOVA, fixed genotypes and environments, GxE fixed (lack of fit) term
  - $y_{ij} = \mu + g_i + e_j + ge_{ij} + \epsilon_{ij};\ \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$
  - Subscript *i* for genotype, *j* for environment
- Mixed model formulation, random genotypes, GxE as heterogeneity of genetic variances and correlations
  - $y_{ij} = \mu_j + g_{ij} + \epsilon_{ij};\ \mathrm{VCOV}(y_{ij}) = \boldsymbol{\Sigma}_{gge} + \boldsymbol{R}_\epsilon$

QTL models as a simple example of G2P models

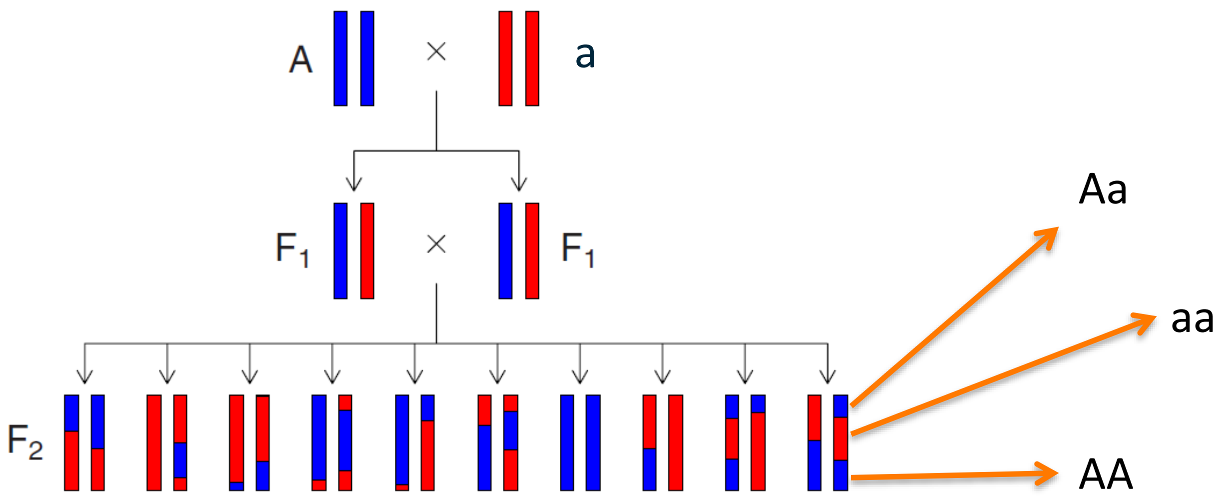
On what chromosome is the locus for early flowering?

Slide: Padraic Flood, WUR-Genetics

Columbia wt

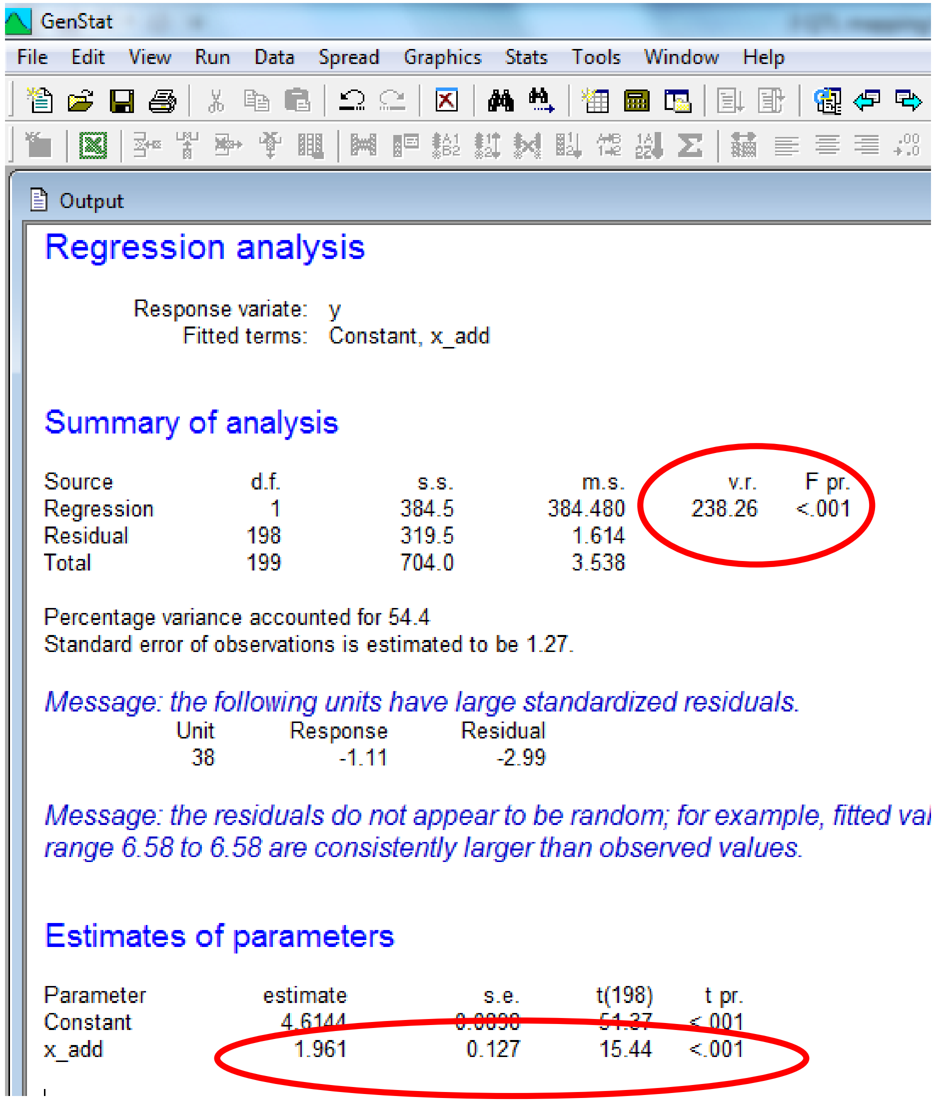


F2 population



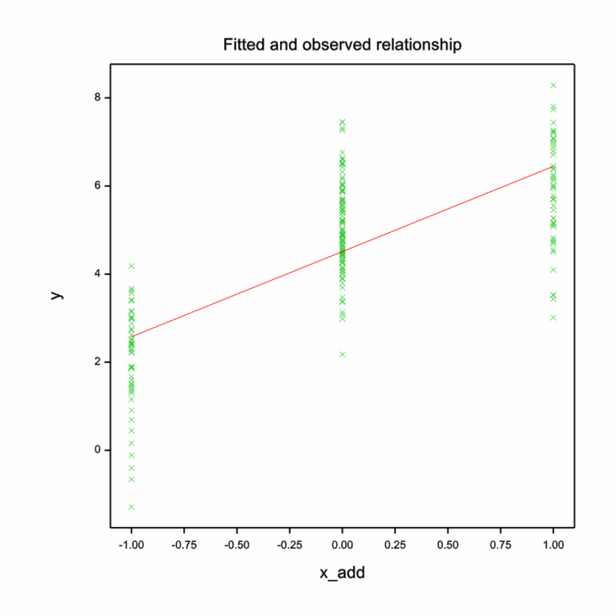
- The F1 is selfed one time
- All 3 possible genotypes are present: AA, Aa, and aa
- Again, short ‘history’ of recombination
- Allows to distinguish additivity from dominance

QTL detection by regression



```
Regression analysis

        Response variate:  y
            Fitted terms:  Constant, x_add

Summary of analysis

Source          d.f.       s.s.       m.s.     v.r.   F pr.
Regression         1      384.5    384.480   238.26   <.001
Residual         198      319.5      1.614
Total            199      704.0      3.538

Percentage variance accounted for 54.4
Standard error of observations is estimated to be 1.27.

Message: the following units have large standardized residuals.
            Unit        Response        Residual
              38           -1.11           -2.99

Message: the residuals do not appear to be random; for example, fitted val
range 6.58 to 6.58 are consistently larger than observed values.

Estimates of parameters

Parameter          estimate          s.e.     t(198)    t pr.
Constant             4.6144        0.0898      51.37    <.001
x_add                 1.961         0.127      15.44    <.001
```

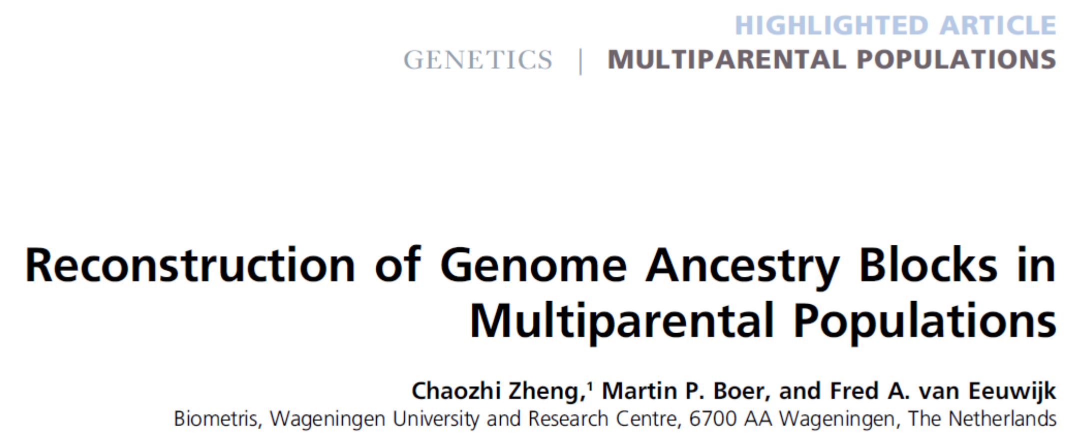
- F-test gives a significant effect
    - Reject Ho
    - There is a QTL at or close to this position
- The estimate of the additive effect is a = 1.961



A General Modeling Framework for Genome Ancestral Origins in Multiparental Populations

Chaozhi Zheng,[1] Martin P. Boer, and Fred A. van Eeuwijk
Biometris, Wageningen University and Research Centre, 6700AC Wageningen, The Netherlands

Genetics 2014



HIGHLIGHTED ARTICLE
GENETICS | MULTIPARENTAL POPULATIONS

Reconstruction of Genome Ancestry Blocks in Multiparental Populations

Chaozhi Zheng,[1] Martin P. Boer, and Fred A. van Eeuwijk
Biometris, Wageningen University and Research Centre, 6700 AA Wageningen, The Netherlands

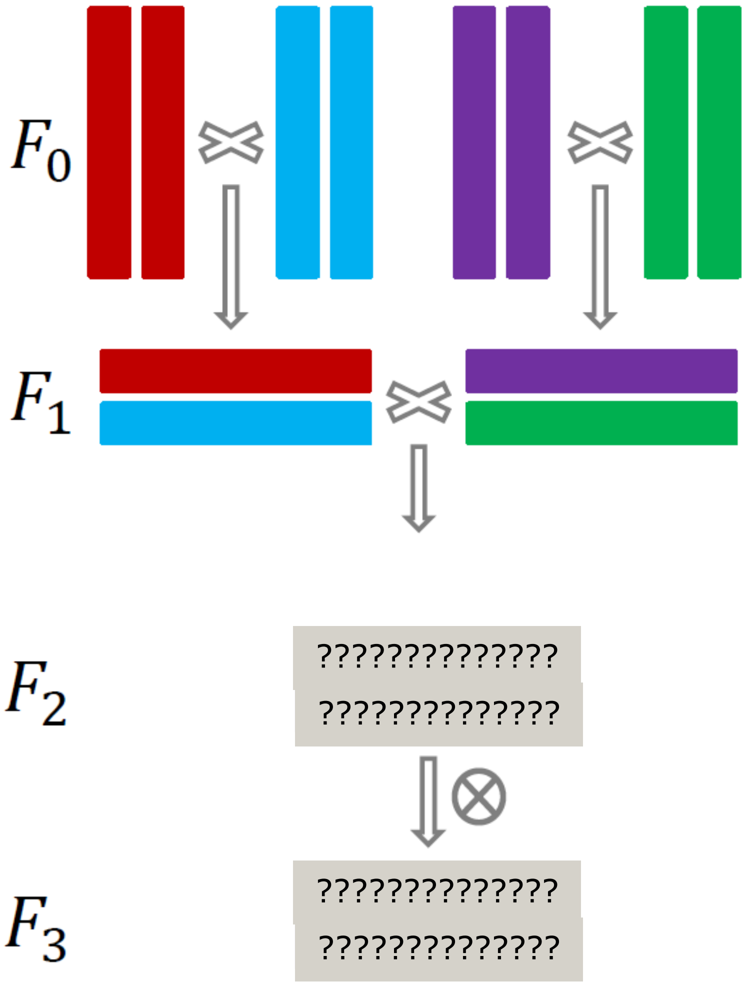
Genetics 2015

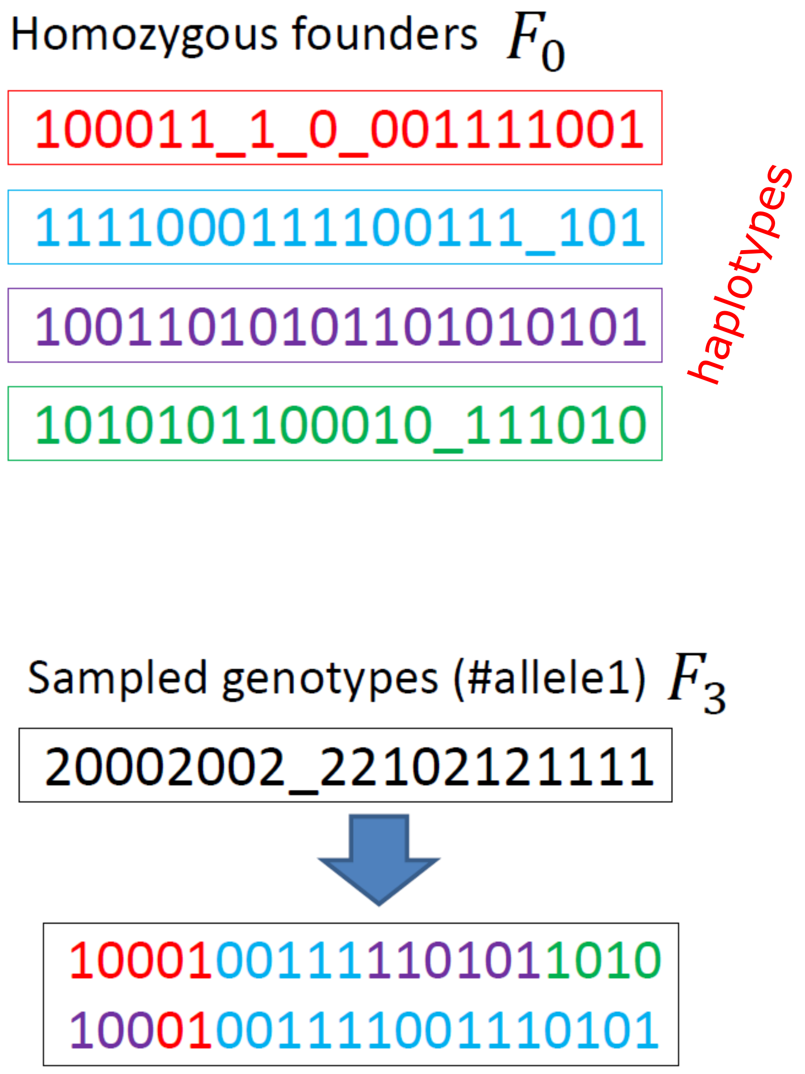
Tracking transmission of DNA from one generation to the next!

Identity by Descent versus Identity by State

MPP Haplotype reconstruction

Four-way RIL by selfing





Homozygous founders F_0

100011_1_0_001111001

1111000111100111_101

10011010101101010101

1010101100010_111010

Sampled genotypes (#allele1) F_3

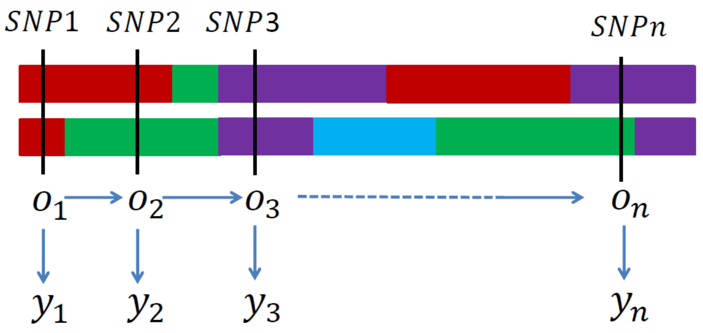
20002002_22102121111

10001001111101011010
10001001111001110101

Genetic predictors for QTL mapping

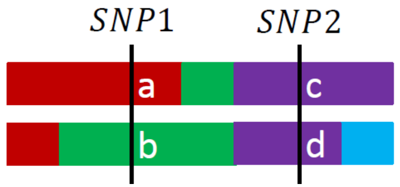
MPP Statistical framework

Hidden Markov model



- Hidden state $o_i = (p,m)$ at SNP$_i$, where p,m = **red**, **green**, **purple**, **blue**.
- Sample individual's genotype y_i = 00, 01, 11, or NN (= missing data), conditional independent.

The prior ancestral origin process



Modelling

$$p\left[o_1 = \binom{a}{b} \rightarrow o_2 = \binom{c}{d}\right]$$

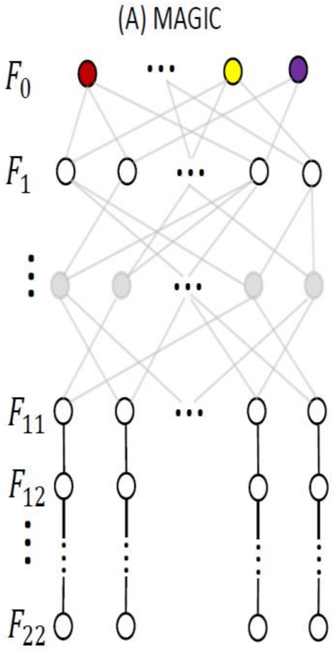
- *indepModel* <u>Indep</u>endent transition
  - $p(a \rightarrow c)\ p(b \rightarrow d)$
- *depModel* <u>Dep</u>endent transition
  - $p(b \rightarrow d \mid a \rightarrow c) = 1$
- *jointModel* <u>Joint</u> transition
  - $p(a \rightarrow c)\ p(b \rightarrow d \mid a \rightarrow c)$
  - ➢ *jointModel* ≈ *indepModel* if completely outbred (no inbreeding)
  - ➢ *jointModel* ≈ *depModel* if completely inbred

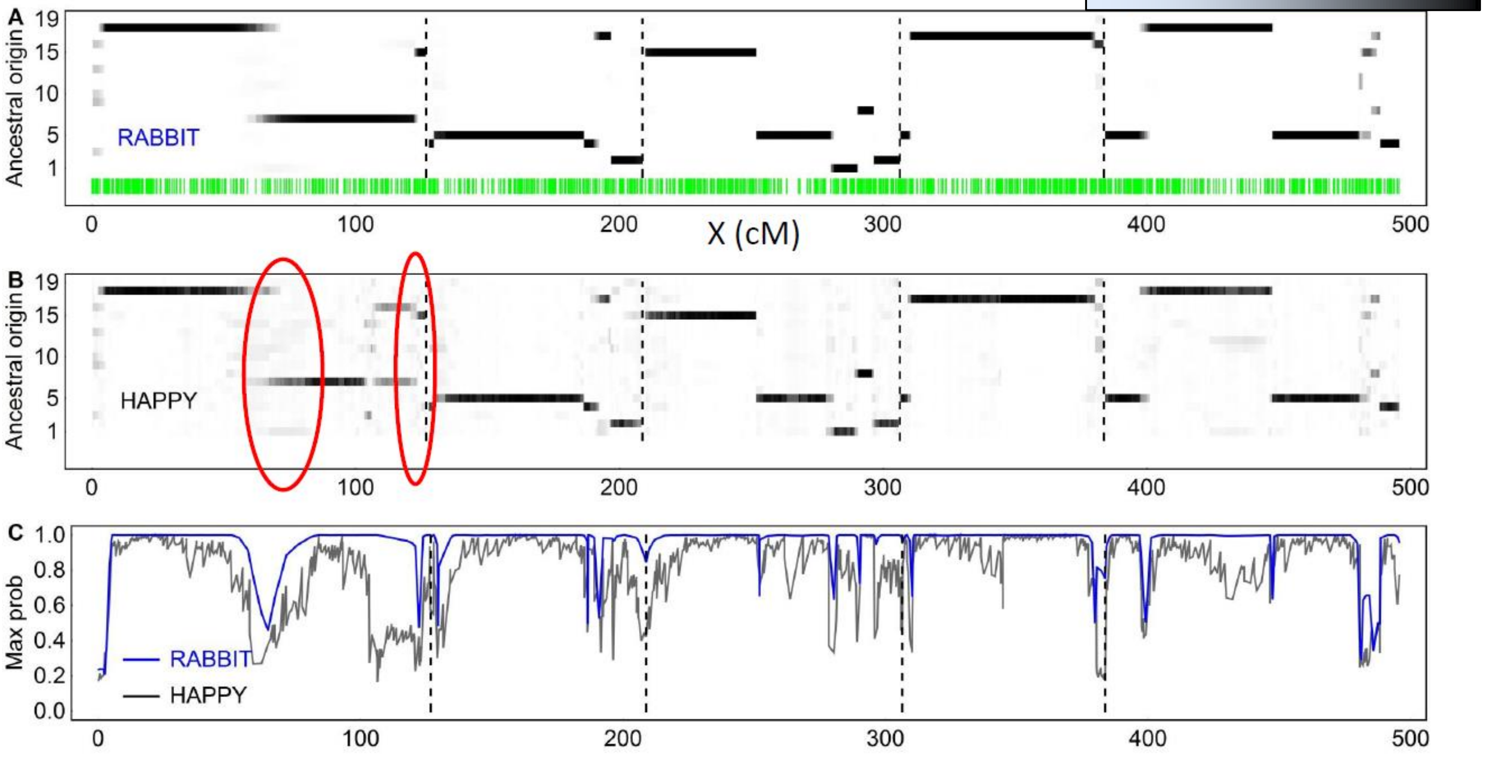
- The data model
  - Prior <u>true</u> gtp *homozygous* founder $p(00)=p(11)=0.5$
  - Prior <u>true</u> gtp sampled individual $p(00)=p(01)=p(10)=p(11)=0.25$
  - Allelic errors occur independently
  - Likelihood is conditional on missing patterns

MPP Individual’s Marginal Posterior Probability

Real data (Kover et al. 2009)
Arabidopsis MAGIC
19 founders

0 1
Grey level =
posterior probability

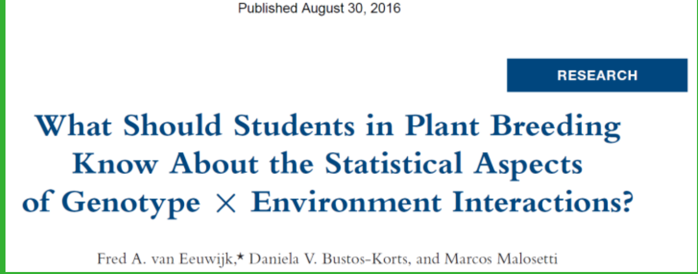




HAPPY are more noisy than RABBIT

G2P models: mixed models! & crop growth models!

WAGENINGEN UR
For quality of life



Published August 30, 2016

RESEARCH

What Should Students in Plant Breeding Know About the Statistical Aspects of Genotype × Environment Interactions?

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

Static G2P modelling and separability

- $\underline{y}_{ij} = \mu_{ij} + \underline{G}_{ij} + \underline{\epsilon}_{ij}$
- Statistical modelling van multi-environment (multi-trait) data is identifying separable and parsimonious formulations for
- Mean μ_{ij}
  - $\mu_{ij} = \sum_k x_{ik}\beta_{jk}$; $\mu_{ij} = \sum_k \alpha_{ik} z_{jk}$; $\mu_{ij} = \sum_k \theta_{ijk} x_{ik} z_{jk}$ (linear)
  - $\mu_{ij} = \sum_k \gamma_{ik}\delta_{jk}$ (bilinear)
- Variance-Covariance matrix $\underline{G}_{ij}$: $VCOV(\underline{G}_{ij})$
  - $VCOV(\underline{G}_{ij}) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{G} \otimes \boldsymbol{\Sigma}^{E}$
- Marker (***HTG***), pedigree, phenotyping platform, and ***HTP*** information enter as genotypic covariables / genotype VCOV structure
- Environmental information and characterisation enter as environmental covariables / env. VCOV structure

Multi-environment QTL/ GWAS/ genomic prediction model (Static)

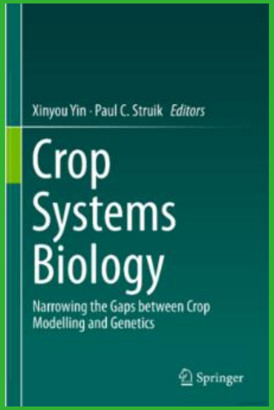
- $y_{i,j} = \mu_j + \sum_{q\in Q}\left(x_{i,q}^{GGE}\beta_{j,q}^{GGE}\right) + G_{i,j} + \epsilon_{i,j}$

- μ_j: environment specific intercept (environmental mean)
- $x_{i,q}^{GGE}$: count of reference allele for QTL q
- $\beta_{j,q}^{GGE}$: environment specific QTL allele substitution effect (fixed)
- $G_{i,j} \sim MVN(0, \boldsymbol{\Sigma}^G, \boldsymbol{\Sigma}^E)$: random (residual) polygenic effect
  - $\boldsymbol{\Sigma}^G$, relations between genotypes (pedigree, markers)
  - $\boldsymbol{\Sigma}^E$, relations between environments (physical, phenotypic)
- $\epsilon_{i,j} \sim MVN(0, \boldsymbol{I}^G, \boldsymbol{R}^E)$: random non-genetic residual

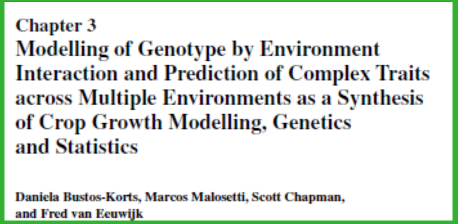
- Alternative: $y_{i,j} = \mu_j + \left\{\left(\sum_{q_1\in Q_1} x_{i,q_1}^{G}\beta_{j,q_1}^{G}\right) + G_i\right\} + \left\{\left(\sum_{q_2\in Q_2} x_{i,q_2}^{GE}\beta_{j,q_2}^{GE}\right) + G_{i,j}\right\} + \epsilon_{i,j}\ ;\ \ Q_2 \subset Q_1$

Mixed models with multiple structured random genetic effects

- It often makes sense to include multiple structured random effects, where the structuring matrices represent different parts of the genome. These can be chromosomes (GCTA), but also individual loci or series of loci belonging to a functional unit (haplotype, gene, exon)
- Haplotype definitions following from exon structure. (EU-Whealbi project)
- $y_i = \mu + G_i^1 + G_i^2 + \epsilon_i$ with $G_i^1 \sim N(0, \boldsymbol{K}_1 \sigma_{a,1}^2)$ and $G_i^2 \sim N(0, \boldsymbol{K}_2 \sigma_{a,2}^2)$
- $y_{i,j} = \mu_j + \sum_{q \in Q} x_{i,q}^{GGE} \beta_{j,q}^{GGE} + G_{ij}^1 + G_{ij}^2 + \epsilon_{i,j}$
- By creating GRMs on the basis of different types of data (SNPs, gene expression, metabolites), these models are also suitable for ***data integration***


Xinyou Yin · Paul C. Struik *Editors*
Crop Systems Biology
Narrowing the Gaps between Crop Modelling and Genetics
Springer

G2P model: physiological, dynamical, hierarchical

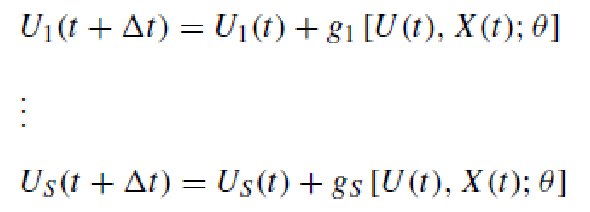

Chapter 3
Modelling of Genotype by Environment Interaction and Prediction of Complex Traits across Multiple Environments as a Synthesis of Crop Growth Modelling, Genetics and Statistics

Daniela Bustos-Korts, Marcos Malosetti, Scott Chapman, and Fred van Eeuwijk

$\underline{y}_{ij}^{f}, \underline{y}_{ij}^{c}$

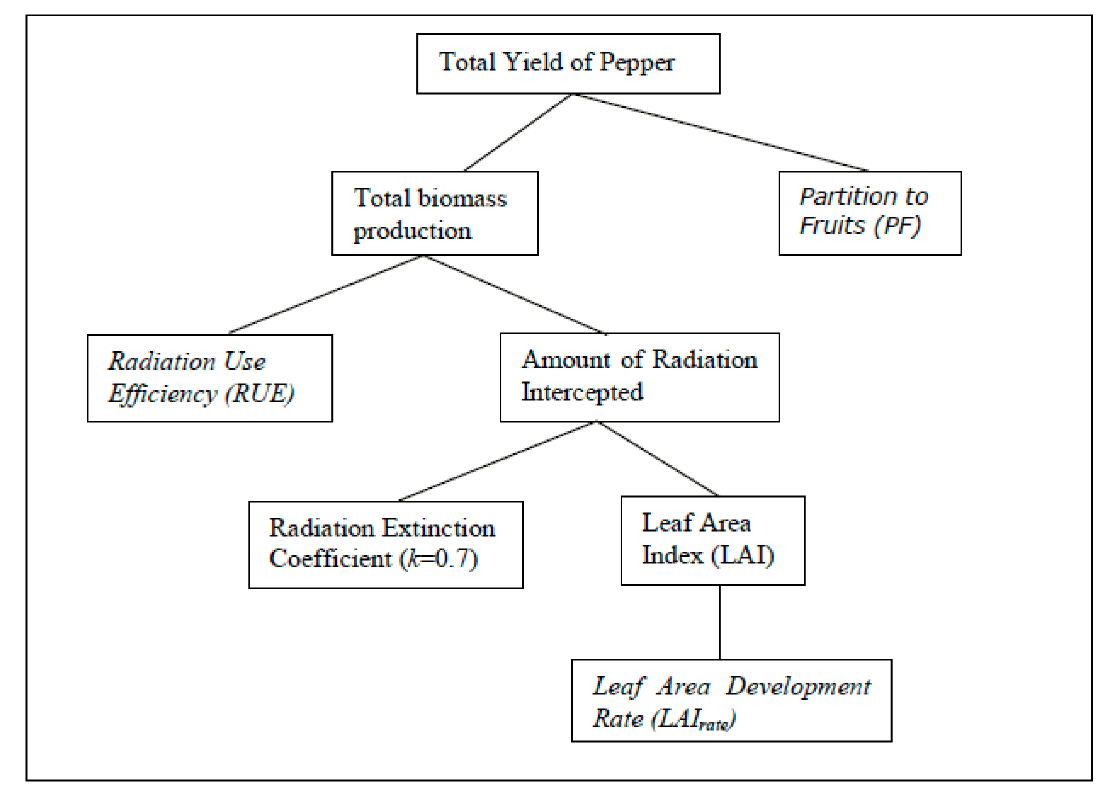
$$\underline{y}_{ij}^{f} = \int f_t\left(\underline{y}_{i}^{p}, \underline{y}_{ij}^{c}; \boldsymbol{z}_j\right) dt$$

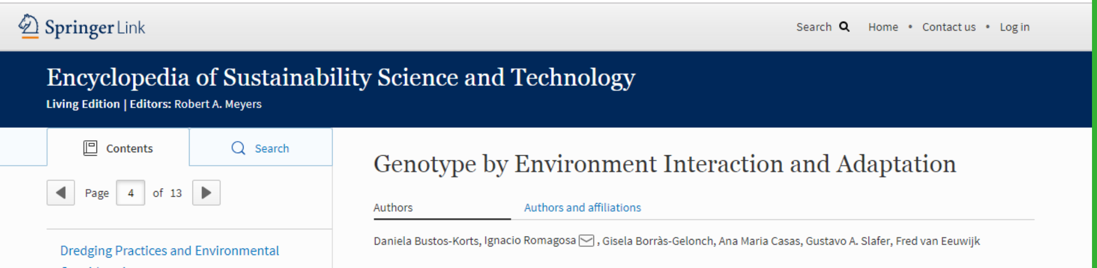
$$\underline{y}_{ij}^{c} = \int g_t\left(\underline{y}_{i}^{p}; \boldsymbol{z}_j\right) dt$$



$$\begin{aligned} U_1(t+\Delta t) &= U_1(t) + g_1\left[U(t), X(t); \theta\right] \\ &\vdots \\ U_S(t+\Delta t) &= U_S(t) + g_S\left[U(t), X(t); \theta\right] \end{aligned}$$

- $\underline{y}_{ij}^{f}$ focal or target trait for genotype i in environment j
- $\underline{y}_{i}^{p}$ physiological parameter
  - Resource capture and conversion, partitioning
- $\boldsymbol{z}_j$ environmental input
- $\underline{y}_{ij}^{c}$ intermediate / integrative / component trait


Total Yield of Pepper
Total biomass production
Partition to Fruits (PF)
Radiation Use Efficiency (RUE)
Amount of Radiation Intercepted
Radiation Extinction Coefficient (k=0.7)
Leaf Area Index (LAI)
Leaf Area Development Rate (LAI_{rate})



Encyclopedia of Sustainability Science and Technology
Living Edition | Editors: Robert A. Meyers

Dredging Practices and Environmental

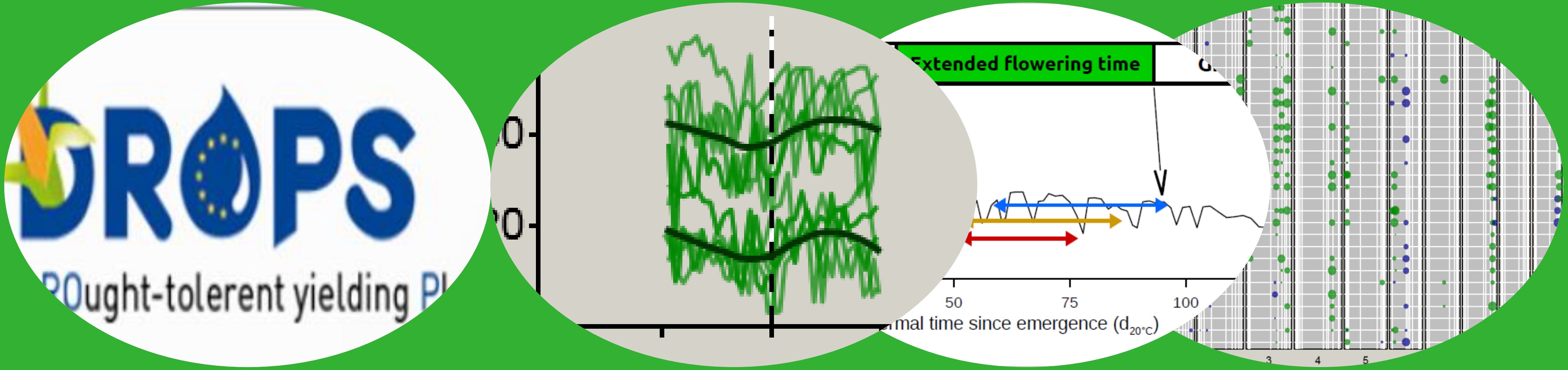
Genotype by Environment Interaction and Adaptation

Daniela Bustos-Korts, Ignacio Romagosa, Gisela Borràs-Gelonch, Ana Maria Casas, Gustavo A. Slafer, Fred van Eeuwijk

Example multi-environment GWAS

GWAS maize

Emilie Millet & DROPS



DROPS
DROught-tolerent yielding PlantS

Integrating physiology, environmental characterization and statistics

Genome-wide analysis of maize yield in European fields under drought and heat; EU-DROPS, project leader *Francois Tardieu*

Allelic effects of QTLs vary with climatic scenarios defined by drought and heat stresses

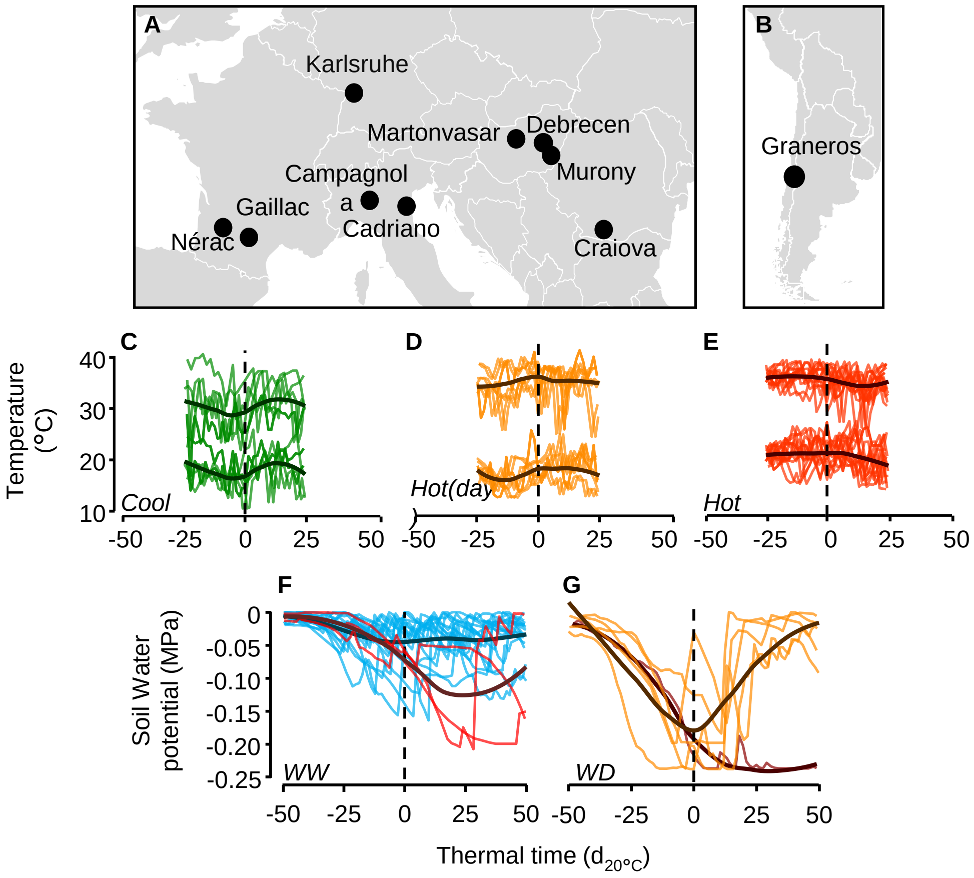
253 maize genotypes
333k SNPs
29 environments = management x location x year
15 environmental characterizations/ indices

Computing time: 7h (i7 PC)

Genome-Wide Analysis of Yield in Europe: Allelic Effects Vary with Drought and Heat Scenarios[1][OPEN]

Emilie J. Millet, Claude Welcker, Willem Kruijer, Sandra Negro, Aude Coupel-Ledru, Stéphane D. Nicolas, Jacques Laborde, Cyril Bauland, Sebastien Praud, Nicolas Ranc, Thomas Presterl, Roberto Tuberosa, Zoltan Bedo, Xavier Draye, Björn Usadel, Alain Charcosset, Fred Van Eeuwijk, and François Tardieu*

Plant Physiology, 2016



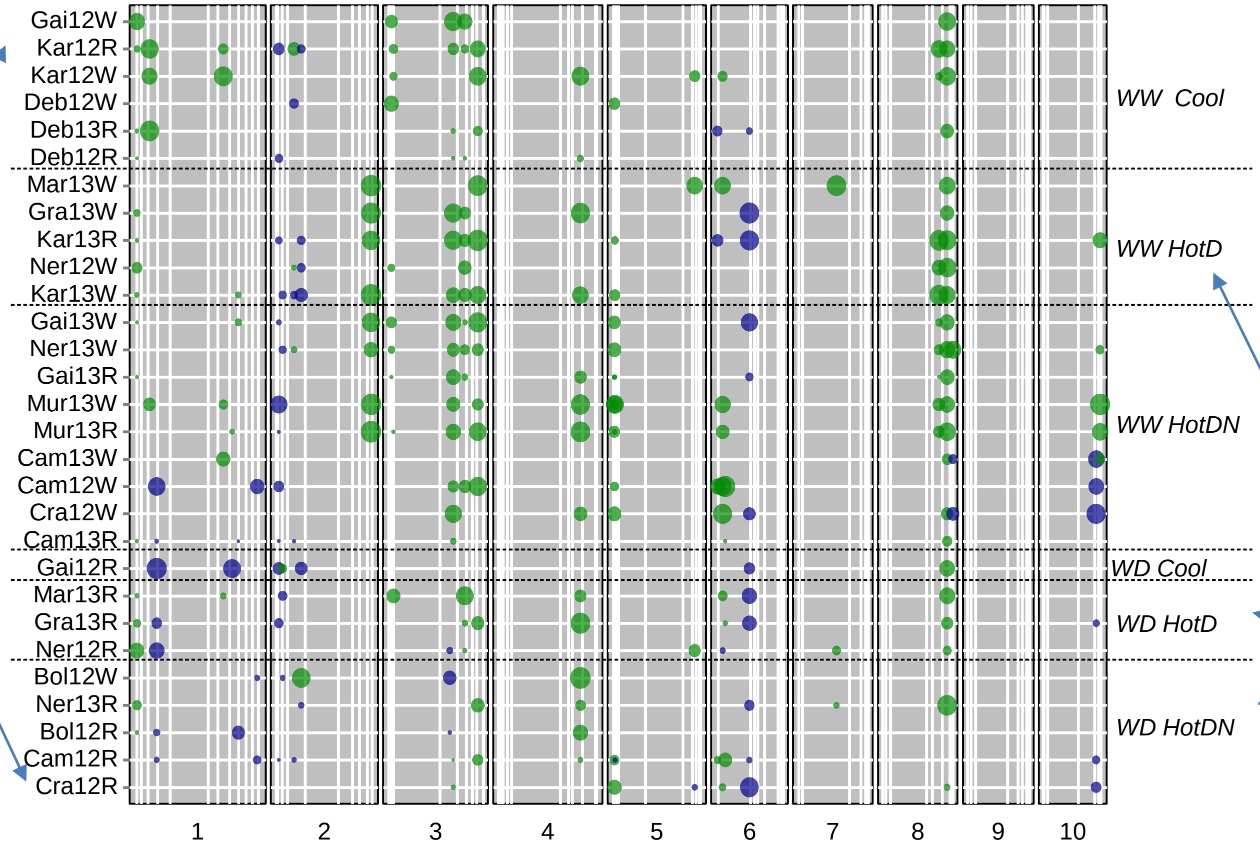
- **Environmental characterization (scenarios)**

- Biology
    - Run CGM over historical env. records

- Statistics
    - Cluster stress responses over growing season

Figure 1. Time courses of soil water potential and temperature in each environmental scenario.
C-G Each line represents a time course corresponding to one experiment of the field network in Europe, A, plus one Chilean location, B. C-G, time courses were centred (time 0) on the day of anthesis of the reference hybrid (vertical dashed line). Dark lines represent smoothed mean values.
Thermal time is in equivalent days at 20°C.
C, experiments with cool temperatures during both day and night (*Cool*);
D, experiments with hot temperatures during the day but cool temperatures during the night (*Hot(day)*);
E, experiments with hot temperatures (*Hot,* mean maximum temperatures >33°C and mean night temperatures >20°C).
C-E, the upper series of lines represents maximum temperature and the lower series represents mean night temperature. F, Well-watered experiments (*WW)* with experiments that displayed mean soil water potential > -0.10 Mpa (*WW cycle*, blue lines) and experiments with well-watered condition during flowering time and water deficit during grain filling (*late Term*, red) ;
G, Water deficit experiments (*WD*) with early deficit followed by recovery at flowering time (*Rec*, yellow lines) and experiments with water deficit from vegetative stage to maturity *(early Term*, dark red lines).

Multi environment GWAS



Environments/ trials

Illustration of results of multi-environment QTL analysis showing clear dependence of QTL expression on environment

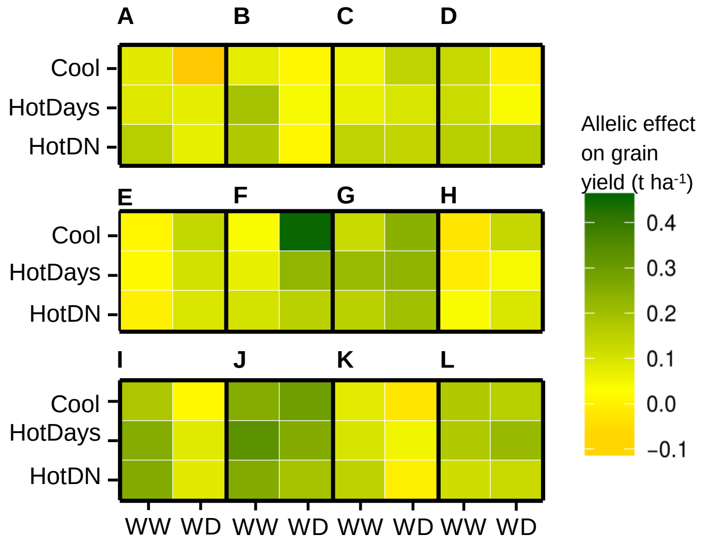
Environmental scenarios from Crop Growth Model simulations

Fig. 5 Final set of significant QTLs for grain yield in the different experiments.
Circle diameters are proportional to the absolute value of allelic effect: the highest the value, the largest the circle. Colour of point indicating direction of the effect: green when the reference hybrid allele increases grain yield, blue when the other allele increases grain yield. Physical positions of the markers are based on the RefGen_v2 map. Each horizontal line contains QTLs of one experiment, organized by scenarios of water status and temperature, separated by dashed horizontal lines. Vertical white lines indicate bin position (bin are subdivisions of chromosome in maize). Allelic effects were estimated with model M3 and were drawn when they were estimated as significant in an experiment (ratio of estimate vs. standard error above 2).

$$y_{i,j} = \mu_j + \sum\nolimits_{q\in Q} x_{i,q}^{GGE} \beta_{j,q}^{GGE} + G_{i,j} + \epsilon_{i,j}$$

Multi environment QTL model

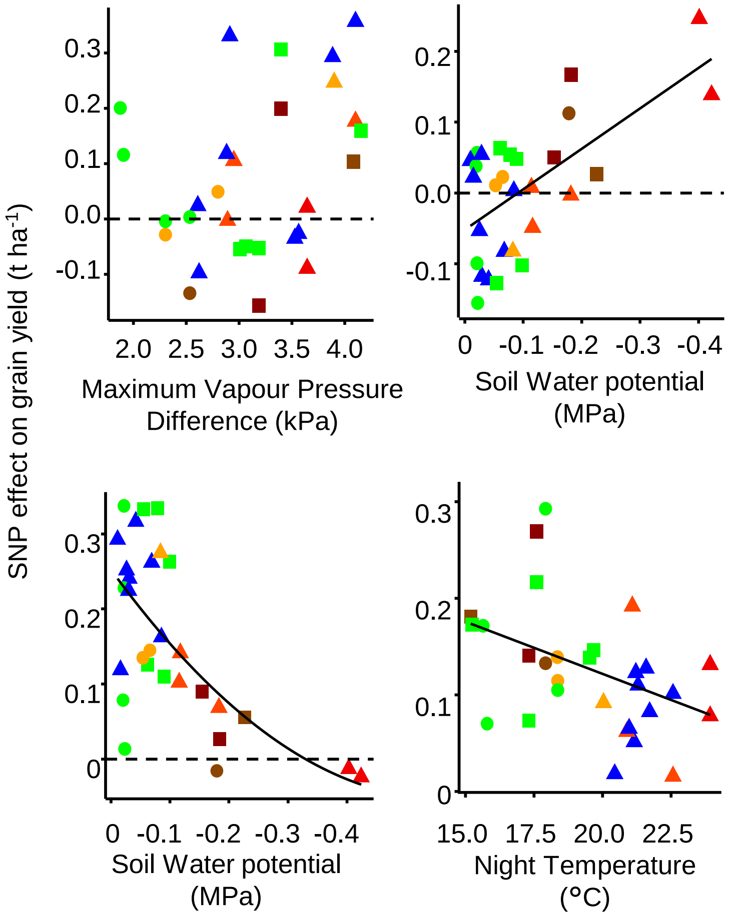
QTL effects in climatic scenarios for 12 largest QTLs



WW=well watered, WD=water deficit

$$\underline{y}_{i,j} = \mu_j + \sum_{q=1}^{Q} x_{i,q}\left(\gamma_q + \delta_q z_j\right) + \underline{G}_{i,j} + \underline{\epsilon}_{i,j}$$

Separability



Dependence of QTL effect on environmental indices / characterizations

Secondary phenotyping

Phenotyping technologies enabling to phenotype over time

Examples

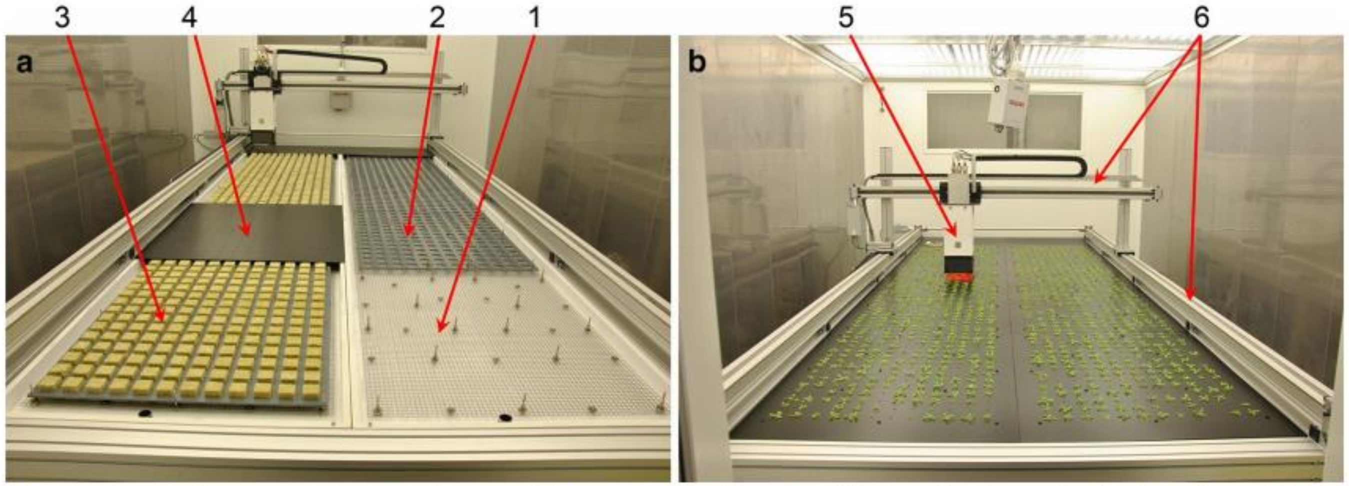


Fig. 2 The Phenovator. **a** The set-up of the growth system. *1* Support grid for the rockwool blocks with support pins for the rockwool block spacing plate and the top plate, *2* the rockwool block spacing plate to position and hold the blocks, *3* this grid with rockwool blocks in place, and *4* the black PVC top plate. **b** The Phenovator system in action: *5* the imaging head carrying the camera (the *red* light is the saturating pulse for ΦPSII measurement), *6* the X–Y rails used to position the imaging head

- Growth chambers for planophile species

Flood et al. Plant Methods (2016) 12:14

Phenotyping technologies enabling to phenotype over time

Examples

Field platforms

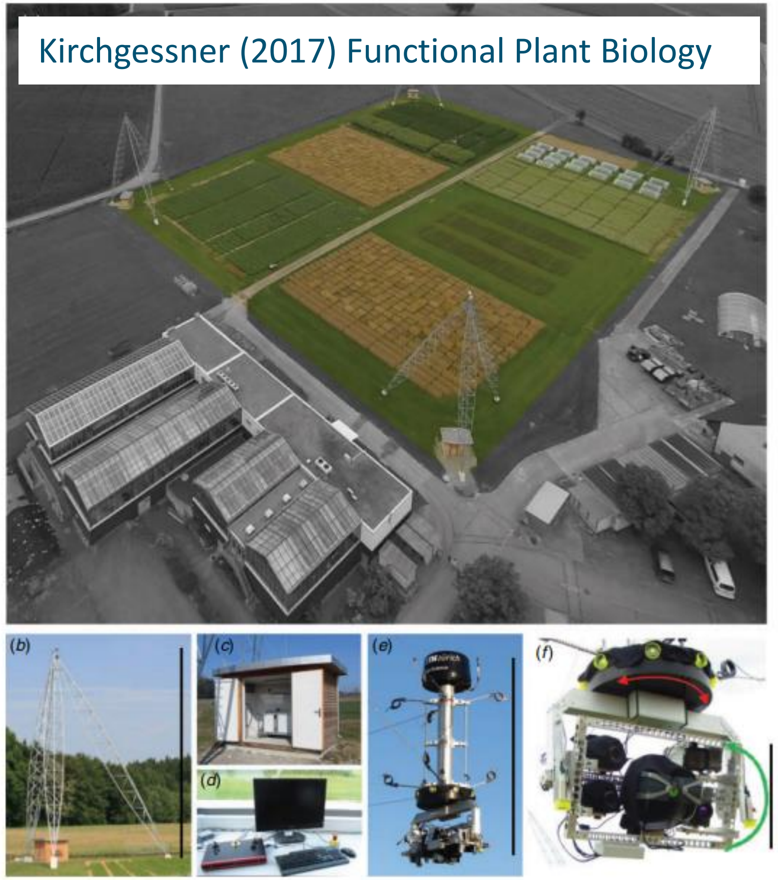
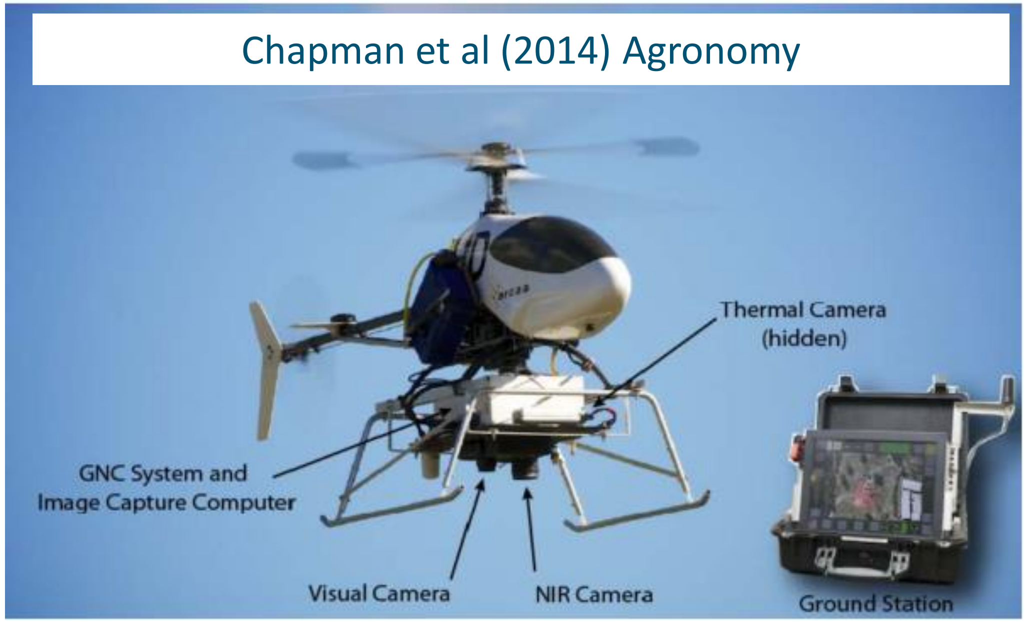

Kirchgessner (2017) Functional Plant Biology

Fig. 1. (a) Overview of the ETH Field Phenotyping Platform (FIP) at ETH plant research station Lindau-Eschikon (WGS84: 47.449 N, 8.682 E) acquired by UAV: D Constantin, M Rehak and Y Akhtman, EPFL ENAC TOPO. Field, poles and winch houses are well visible (scale bar: 24 m). (*b*) Close view of pole with winch house at its base in front view. (*c*) Back view of winch house with open doors and view on winch. (*d*) Console for interactive mode for position control by joysticks. (*e*) Sensor head, top: housing for batteries, computer, spectrometers (Ocean Optics USB 2000+, NIRQuest256–2.5) and electronics, middle: ultrasonic sensors for collision avoidance (yellow) bottom: sensors (scale bar: 1.9 m). (*f*) Sensors, left: two DSLR (Canon 5D Mark II, upper: ENDVI modified), middle: laser scanner (Faro Focus 3D), right top: operator camera, right bottom: thermal camera (Infratec VarioCAM head 600 h) the yellow rain protection of the ultrasonic sensors is easy to recognise. Rotations of pan- (red) and tilt-axis (green) are indicated (scale bar: 0.4 m).

Mobile devices


Chapman et al (2014) Agronomy

Australian Plant Phenomics Facility (CSIRO)

Canopy cover over time

Grieder et al (2014) *Functional Plant Biology*

Examples

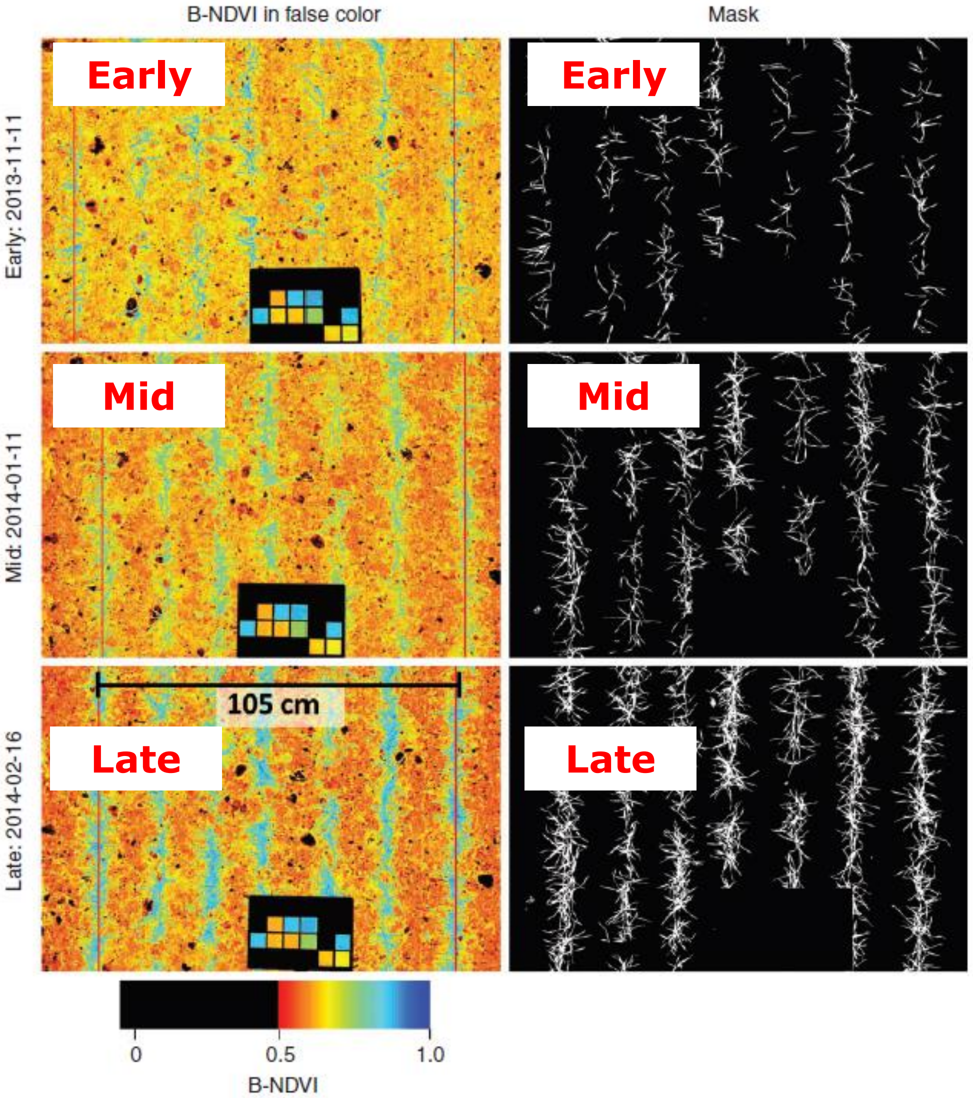


Fig. 1. Images of a wheat plot taken at three different dates during the growing period. Left: normalised difference vegetation index using the blue channel as visible light reference (B-NDVI) of single pixels displayed in false colour with the red lines displaying the recognised positions of the outermost rows. Right: mask obtained from segmentation of plant and background.

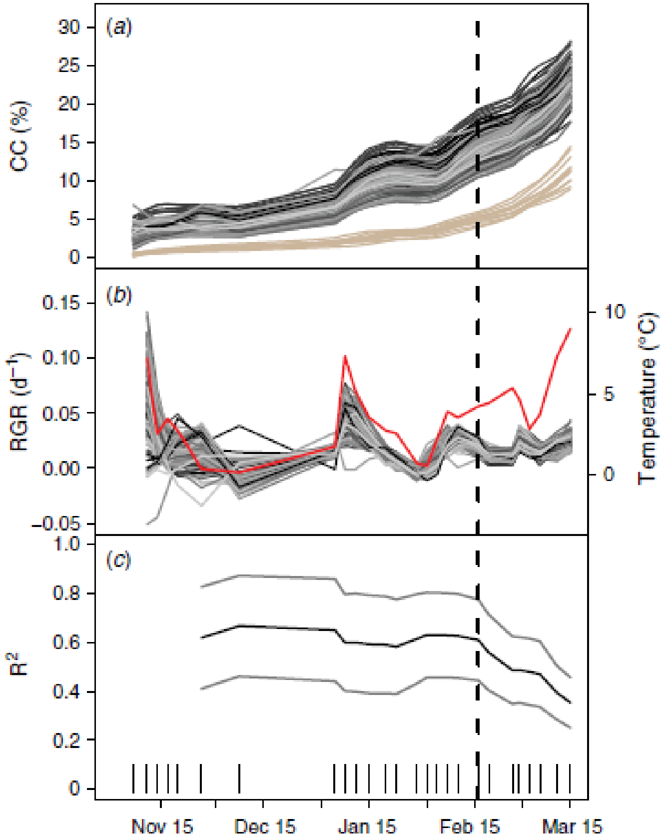
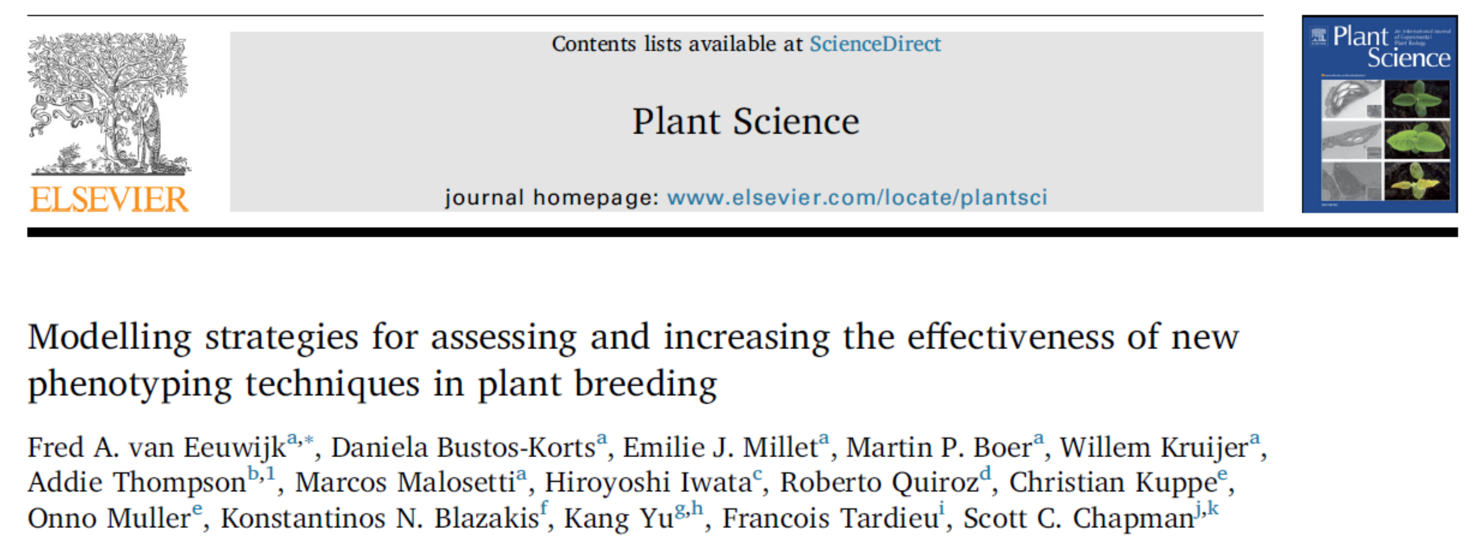


Fig. 2. Development over time of (*a*) canopy cover (CC), (*b*) relative growth rate of canopy cover (RGR, grey lines) and air temperature (red line) and (*c*) coefficient of determination (R^2) of the linear RGR-temperature relation in dependence of growth period duration. The vertical dashed line indicates the determined end of early development phase. In (*a*) and (*b*), curves represent single plot values of the 27 examined varieties (indicated by different grey levels), whereby bisque coloured lines (only in *a*) represent the two varieties Zenith and Probus, which have been excluded from further analysis. In (*c*), the black curve represents the mean R^2 value, grey curves delineate ± s.d. and black ticks depict the measurement dates.

Levels of secondary phenotyping!

- Feature extraction
- Correcting for design features and spatial trends
- Dynamical modelling (integrating over time) -> estimating rate parameters
- Modelling rate parameters on environmental gradients
- Prediction of target trait



Contents lists available at ScienceDirect

Plant Science

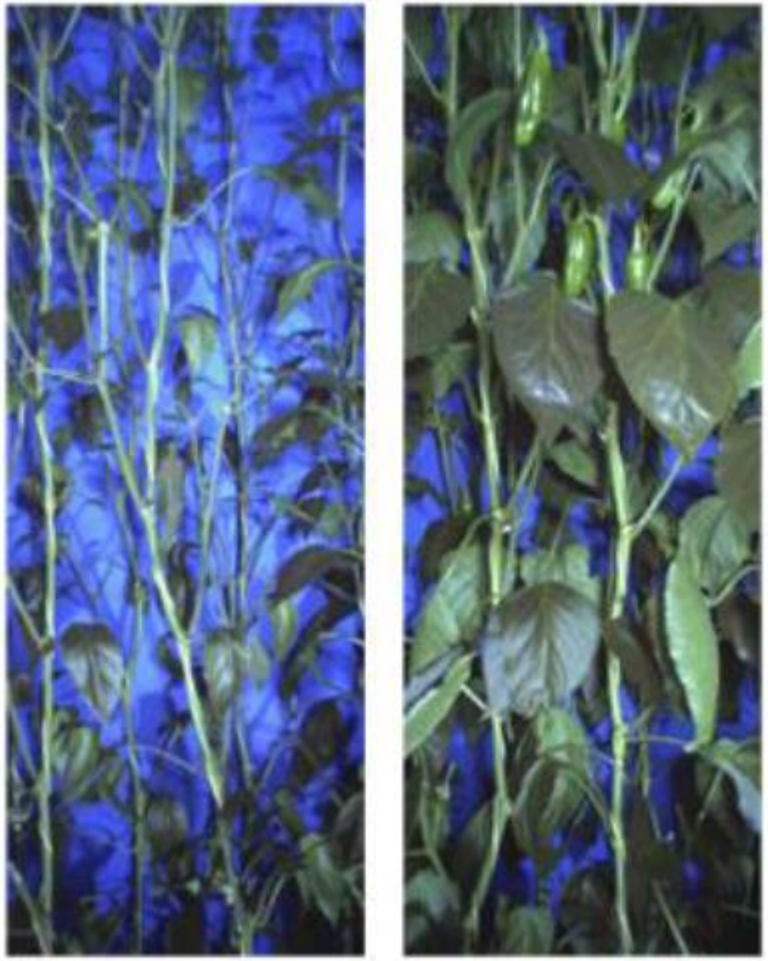
journal homepage: www.elsevier.com/locate/plantsci

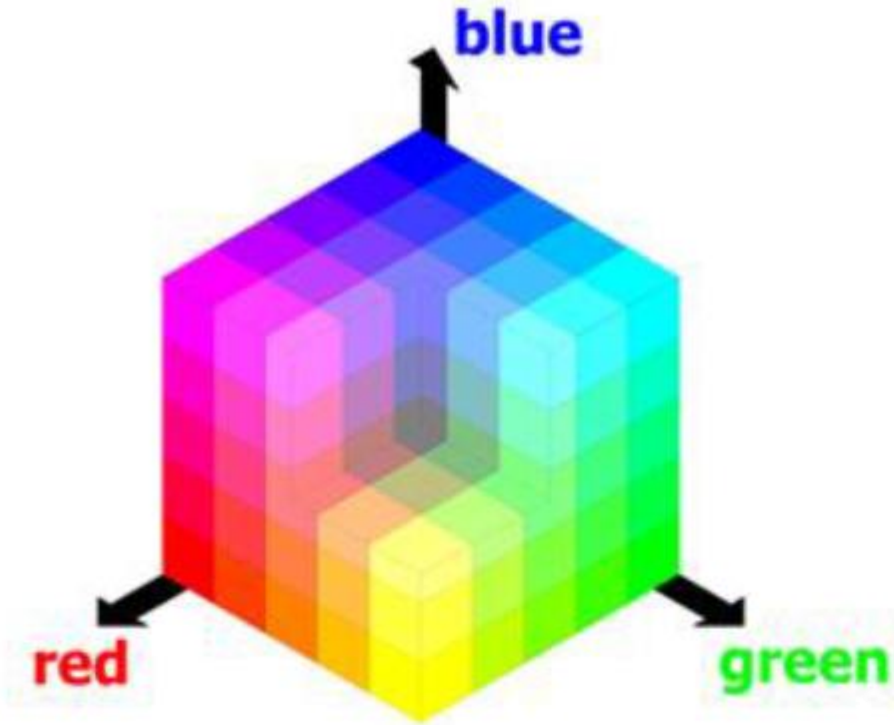
ELSEVIER

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

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

Feature extraction





Spatial model SpATS


Theor Appl Genet
DOI 10.1007/s00122-017-2894-4

ORIGINAL ARTICLE

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

Julio G. Velazco[1,2] · María Xosé Rodríguez-Álvarez[3,4] · Martin P. Boer[1] · David R. Jordan[5] · Paul H. C. Eilers[6] · Marcos Malosetti[1] · Fred A. van Eeuwijk[1]


Spatial Statistics 23 (2018) 52–71

Contents lists available at ScienceDirect

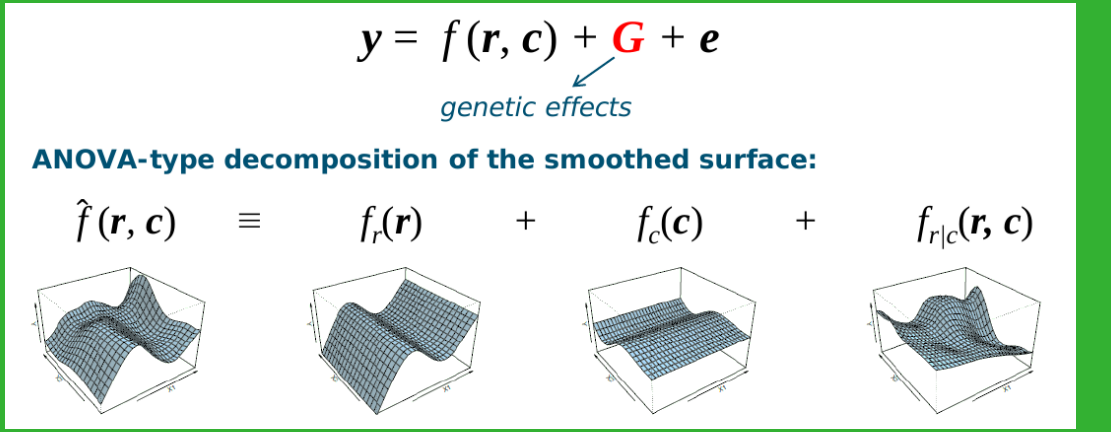
Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

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

María Xosé Rodríguez-Álvarez[a,b,*], Martin P. Boer[c], Fred A. van Eeuwijk[c], Paul H.C. Eilers[d]

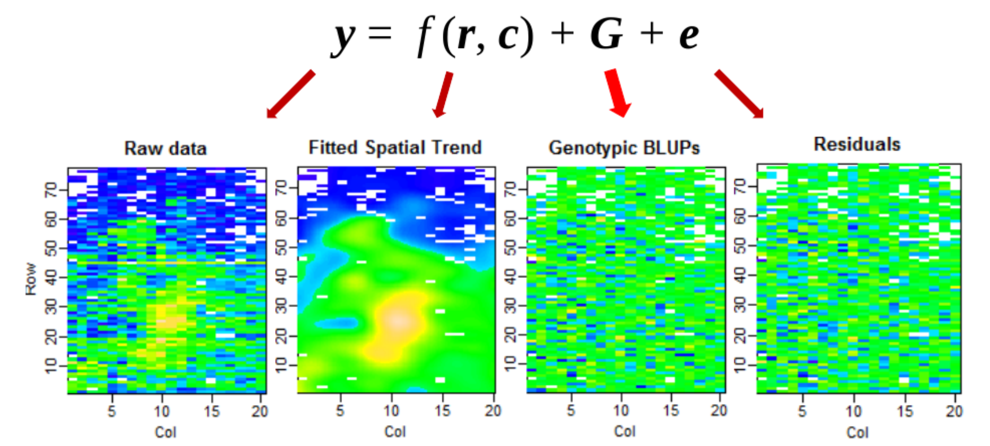
[a] BCAM – Basque Center for Applied Mathematics, Alameda de Mazarredo, 14. E-48009 Bilbao, Basque Country, Spain
[b] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[c] Biometris, Wageningen University & Research, Wageningen, The Netherlands
[d] Erasmus University Medical Centre, Rotterdam, The Netherlands

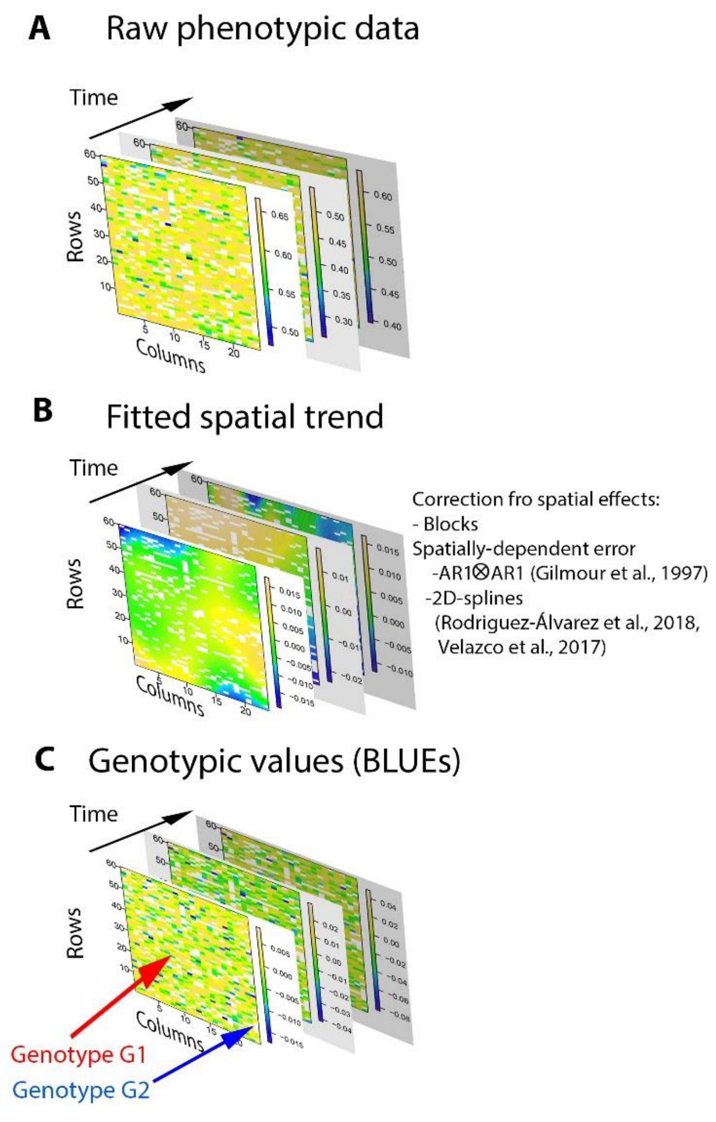

$$\boldsymbol{y} = f(\boldsymbol{r}, \boldsymbol{c}) + \boldsymbol{G} + \boldsymbol{e}$$

genetic effects

ANOVA-type decomposition of the smoothed surface:

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


$$\boldsymbol{y} = f(\boldsymbol{r}, \boldsymbol{c}) + \boldsymbol{G} + \boldsymbol{e}$$



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

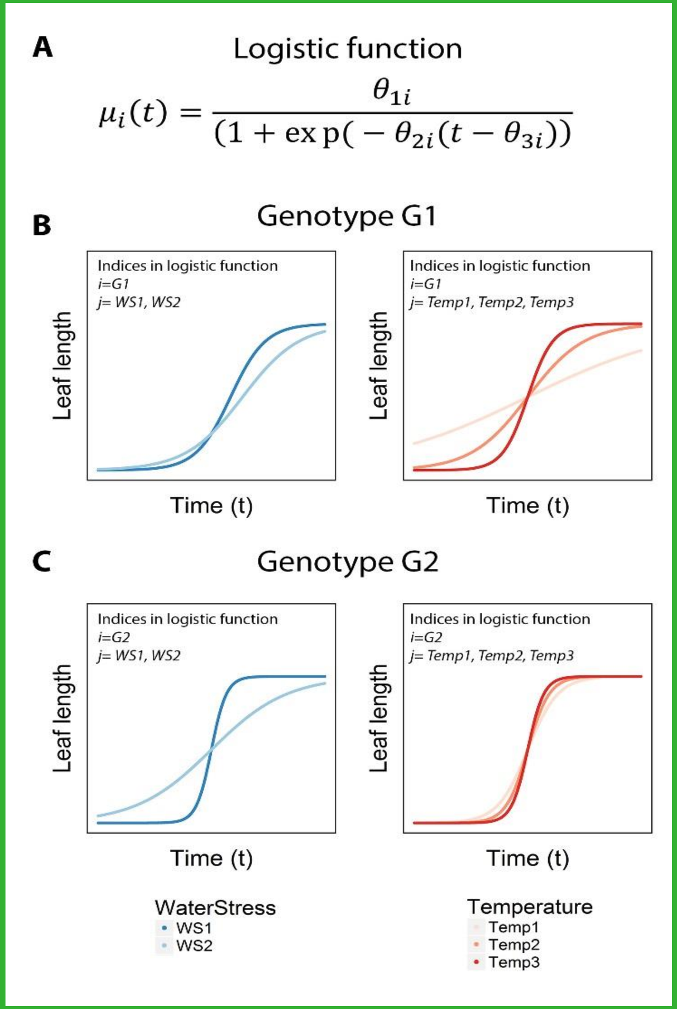
For repeated measures

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

3D smoothing in space and time

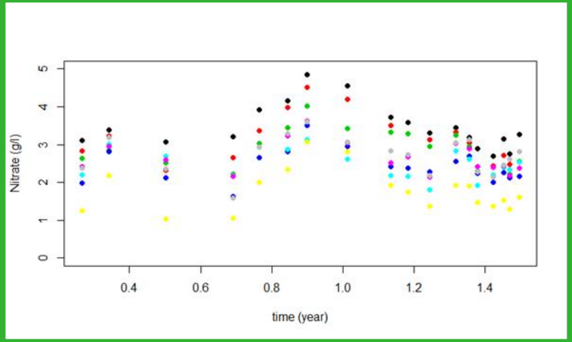
WAGENINGEN UR
For quality of life

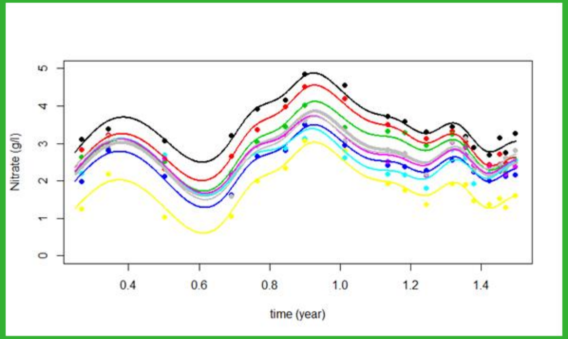
Modelling trait dynamics, parametrically and non-parametrically



Non-parametric modelling of target trait (Martin Boer)

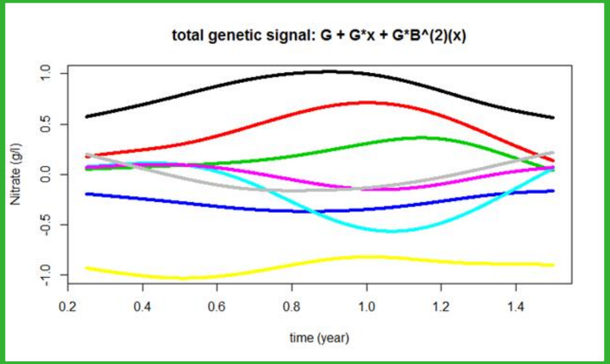
Nitrogen in lettuce data set: 8 genotypes, 18 environments





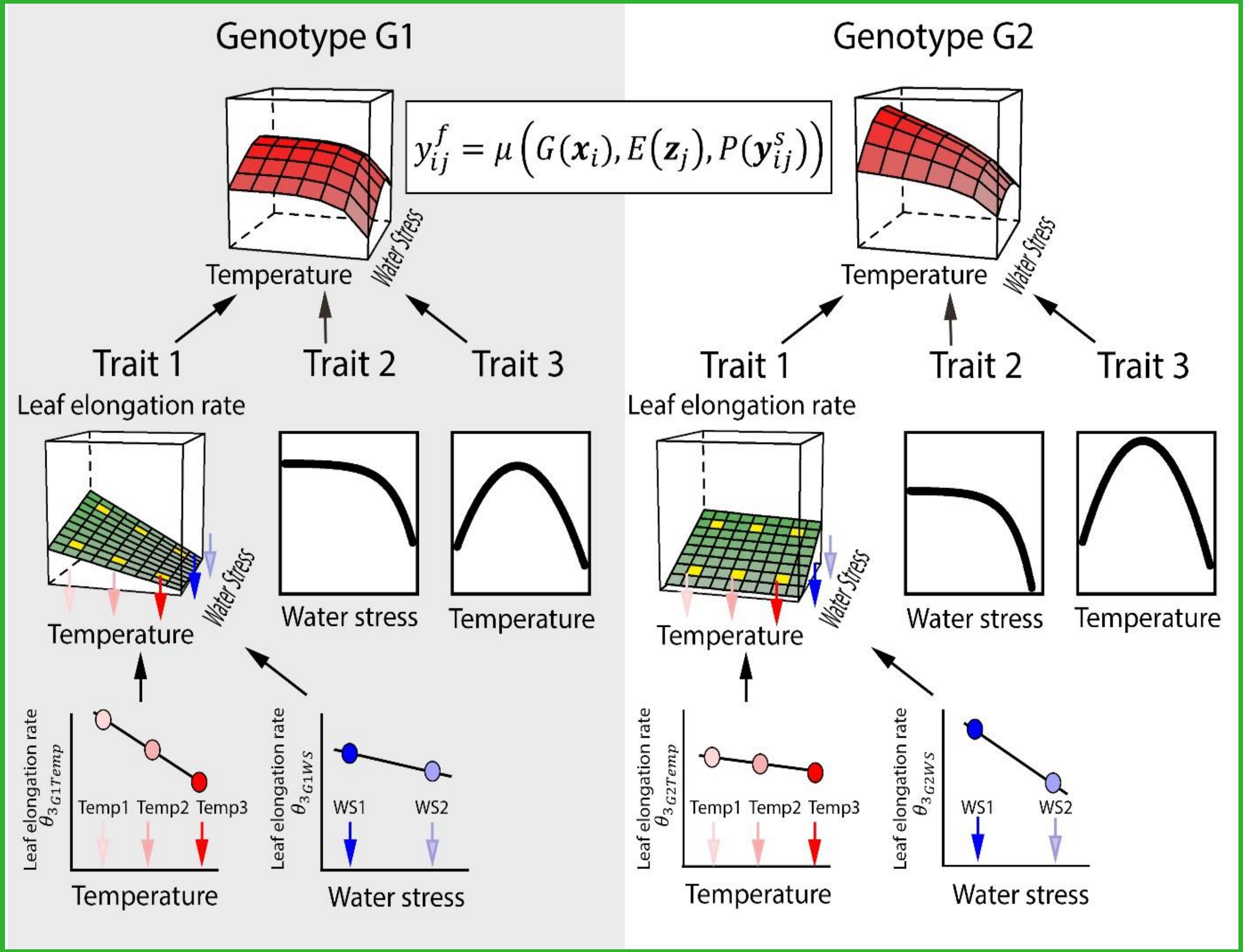
$y(t) = E(t) + G + GE(t) + \varepsilon(t)$

Summer Winter



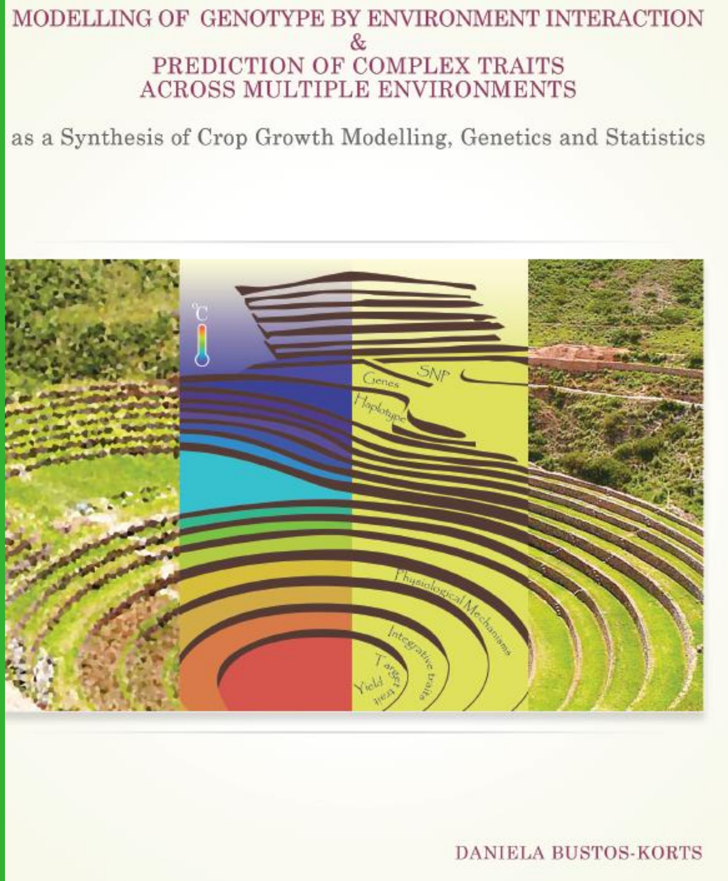
$y(t) = G + GE(t) + \varepsilon(t)$

Modelling rate parameters on environmental gradients & prediction of target trait



$$y_{ij}^{f} = \mu\left(G(\boldsymbol{x}_i), E(\boldsymbol{z}_j), P(\boldsymbol{y}_{ij}^{s})\right)$$

Dynamical modelling / crop growth modelling!



MODELLING OF GENOTYPE BY ENVIRONMENT INTERACTION
&
PREDICTION OF COMPLEX TRAITS
ACROSS MULTIPLE ENVIRONMENTS

as a Synthesis of Crop Growth Modelling, Genetics and Statistics

DANIELA BUSTOS-KORTS

Temperature and Solar Radiation Effects on Potential Maize Yield across Locations

R.C. Muchow, T.R. Sinclair,* and J.M. Bennett

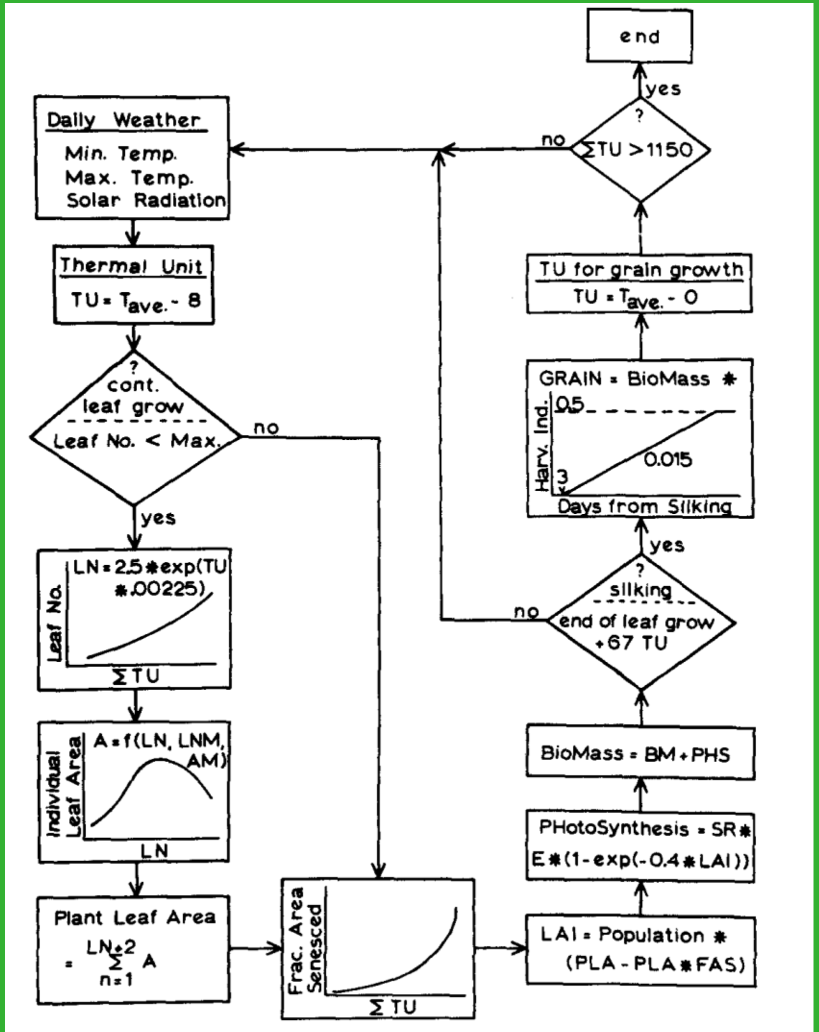


Fig. 1. Flow diagram showing operation of the maize growth model on a daily time step.

RESEARCH ARTICLE

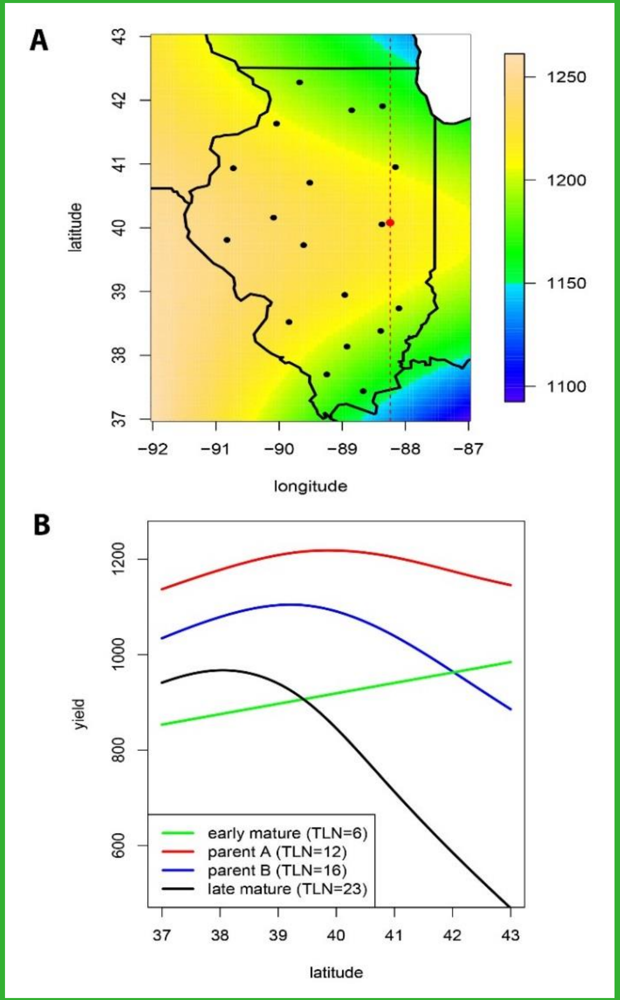
Integrating Crop Growth Models with Whole Genome Prediction through Approximate Bayesian Computation

Frank Technow[1] *, Carlos D. Messina[2], L. Radu Totir[1], Mark Cooper[2]

$$F(y_{TLN_i}, y_{SRE_i}, y_{AM_i}, y_{MTU_i}, \Omega_k)$$

$$\begin{aligned} y_{TLN_i} &= \mu_{TLN} + \mathbf{z}_i\mathbf{u}_{TLN} \\ y_{AM_i} &= \mu_{AM} + \mathbf{z}_i\mathbf{u}_{AM} \\ y_{SRE_i} &= \mu_{SRE} + \mathbf{z}_i\mathbf{u}_{SRE} \\ y_{MTU_i} &= \mu_{MTU} + \mathbf{z}_i\mathbf{u}_{MTU}, \end{aligned}$$

Assessing the granularity level for prediction



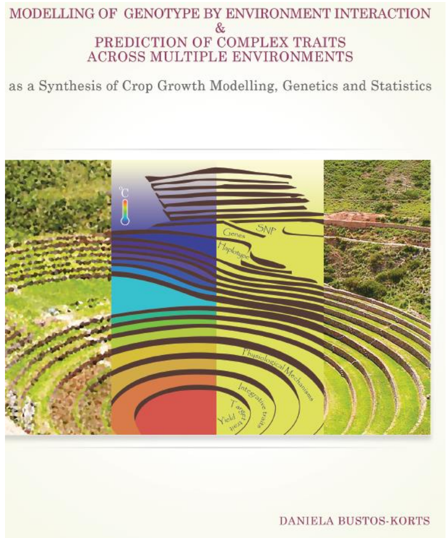
Martin Boer

WAGENINGEN UR
For quality of life

Dynamical modelling of secondary traits & genomic prediction (Daniela Bustos-Korts)

Single- and multi-trait genomic prediction

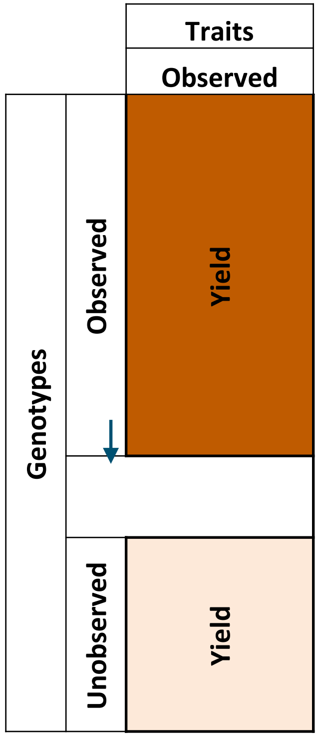
Early prediction of yield from biomass and canopy temperature


MODELLING OF GENOTYPE BY ENVIRONMENT INTERACTION
&
PREDICTION OF COMPLEX TRAITS
ACROSS MULTIPLE ENVIRONMENTS

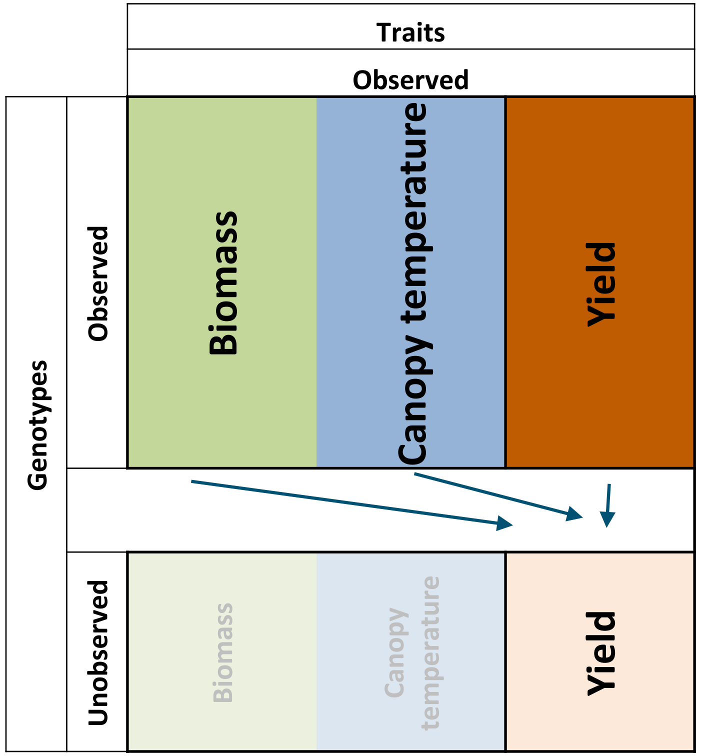
as a Synthesis of Crop Growth Modelling, Genetics and Statistics

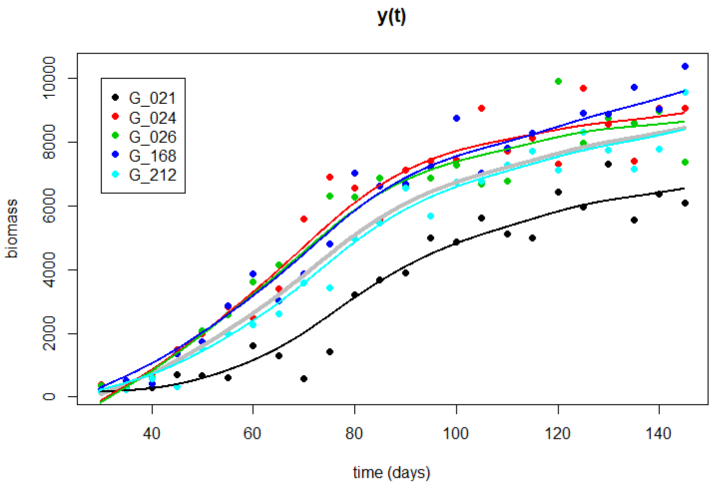
DANIELA BUSTOS-KORTS

Single-trait



Multi-trait

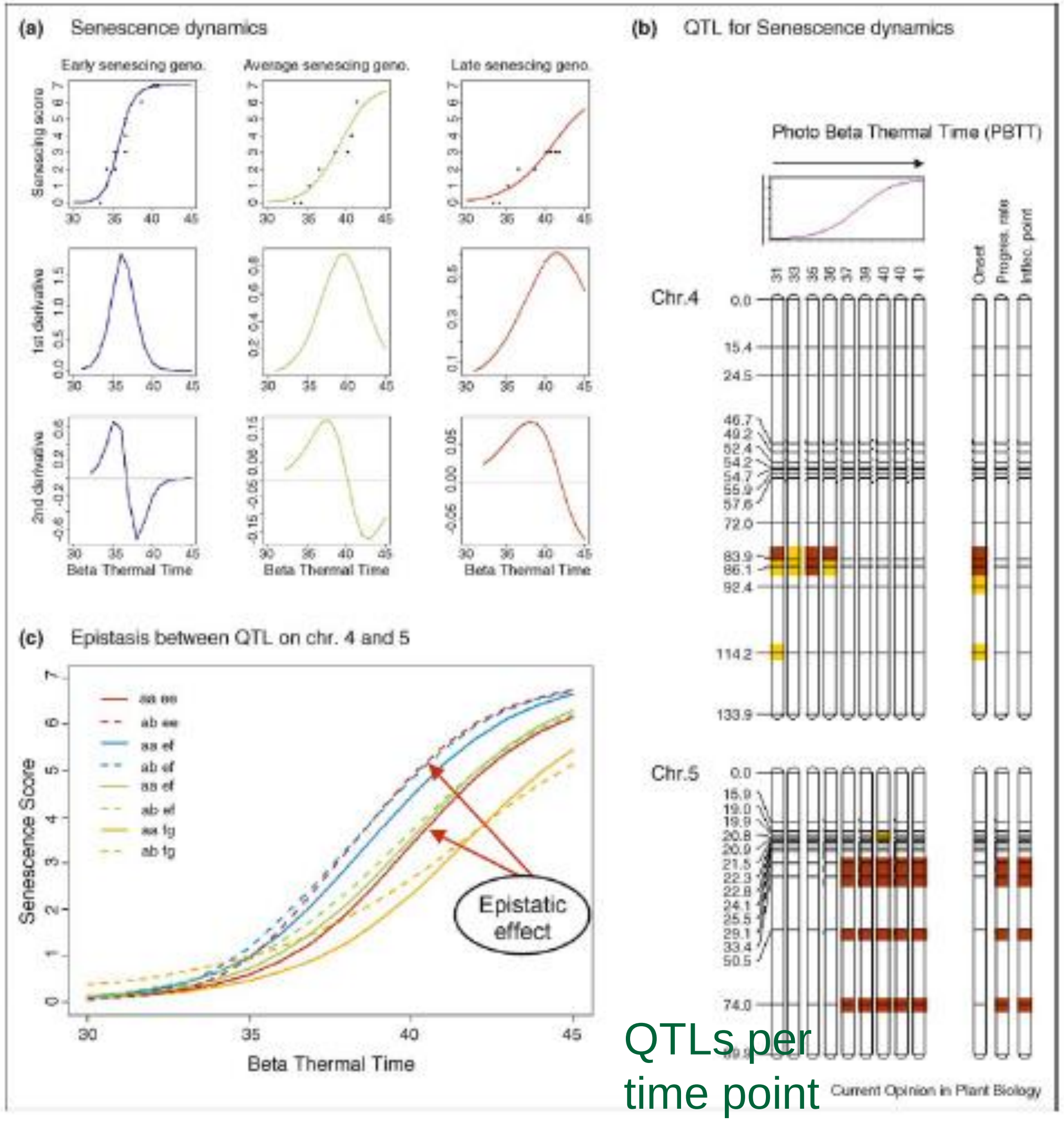




$$y(t) = E(t) + G + GE(t) + \varepsilon(t)$$

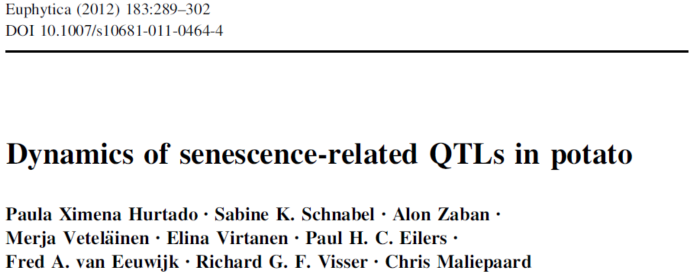
Model QTLs for response curve parameters, not trait at a particular time

Only the curve parameter QTLs can be used in prediction



QTLs for curve parameters

QTLs per time point


Euphytica (2012) 183:289–302
DOI 10.1007/s10681-011-0464-4

Dynamics of senescence-related QTLs in potato

Paula Ximena Hurtado · Sabine K. Schnabel · Alon Zaban ·
Merja Veteläinen · Elina Virtanen · Paul H. C. Eilers ·
Fred A. van Eeuwijk · Richard G. F. Visser · Chris Maliepaard


ELSEVIER

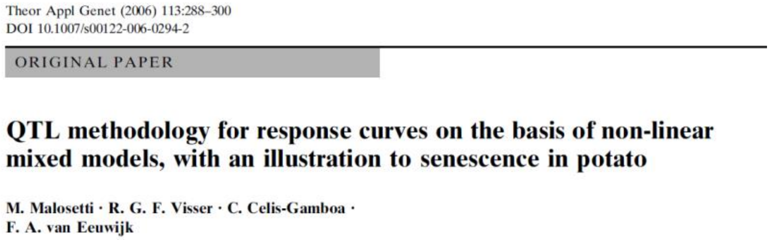
Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Plant Biology

Detection and use of QTL for complex traits in multiple environments

Fred A van Eeuwijk[1,2], Marco CAM Bink[1], Karine Chenu[3] and Scott C Chapman[4]


Theor Appl Genet (2006) 113:288–300
DOI 10.1007/s00122-006-0294-2

ORIGINAL PAPER

QTL methodology for response curves on the basis of non-linear mixed models, with an illustration to senescence in potato

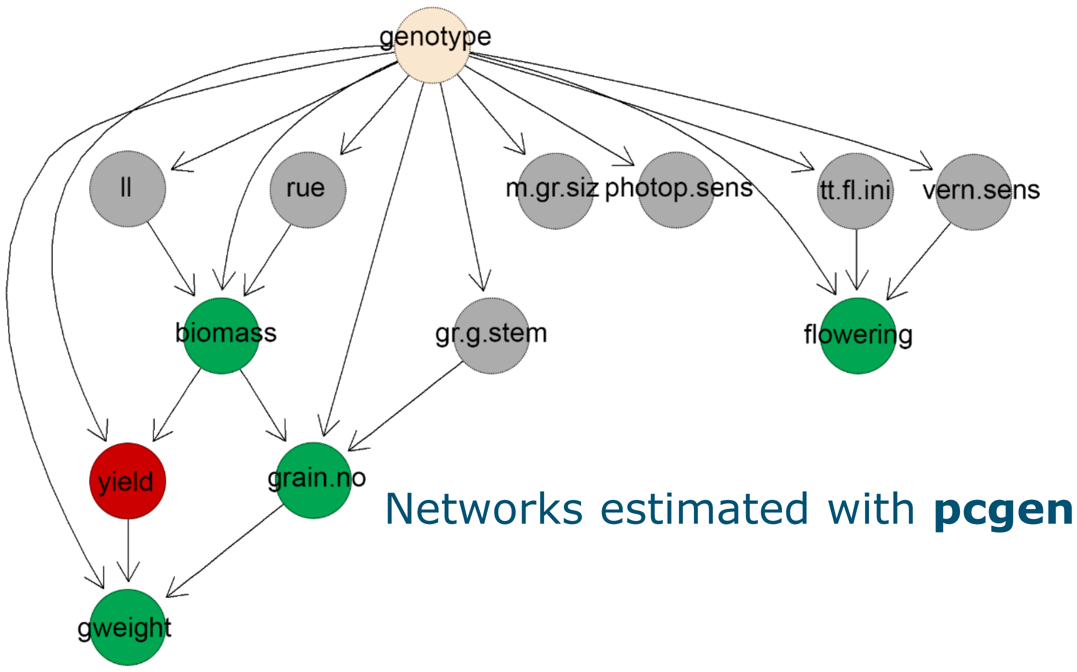
M. Malosetti · R. G. F. Visser · C. Celis-Gamboa ·
F. A. van Eeuwijk

Hierarchical, causal and crop growth modelling!

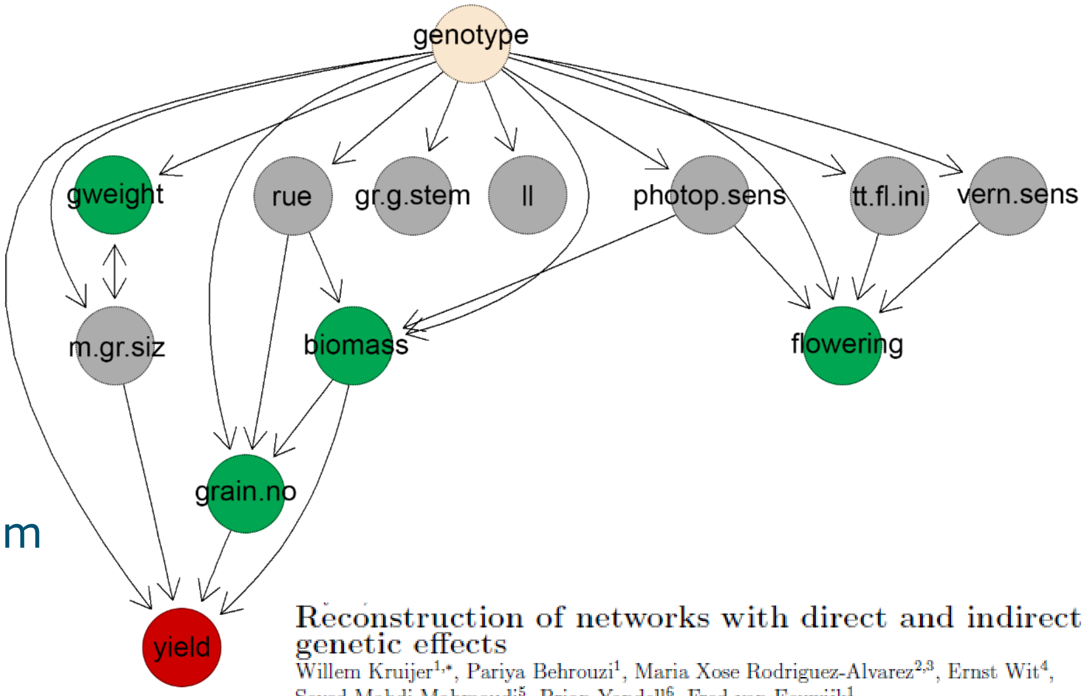
Network showing genetic basis of target trait, component traits and physiological parameters with trait hierarchy

- Networks help visualizing trait relationships (biomass plays a central role for yield)
- Adaptive mechanisms differ between environments (see role phenology)
- Trait organisation shows ‘modules’ (clusters of highly correlated traits, which one to phenotype?)

Drought



No drought



Networks estimated with **pcgen** algorithm

Reconstruction of networks with direct and indirect genetic effects
Willem Kruijer[1,*], Pariya Behrouzi[1], Maria Xose Rodriguez-Alvarez[2,3], Ernst Wit[4], Seyed Mahdi Mahmoudi[5], Brian Yandell[6], Fred van Eeuwijk[1]

G2P models

- Mixed models have served as an excellent class of models for modelling phenotypes as functions of genetic effects and simple environmental characterizations (locations, years, environmental covariables)
- Mixed models are useful for modelling GxE (fixed and random)
- New phenotyping and environmental data require a reconsideration of the type of mixed models for dynamical G2P modelling of multiple traits as functions of genetic and environmental inputs
- Options for G2P modelling: functional data analysis, splines, network models, synthesis of crop growth models and mixed models (statistical genetic)
- Mixed model framework needs extensions in the direction of dynamic and hierarchical modelling

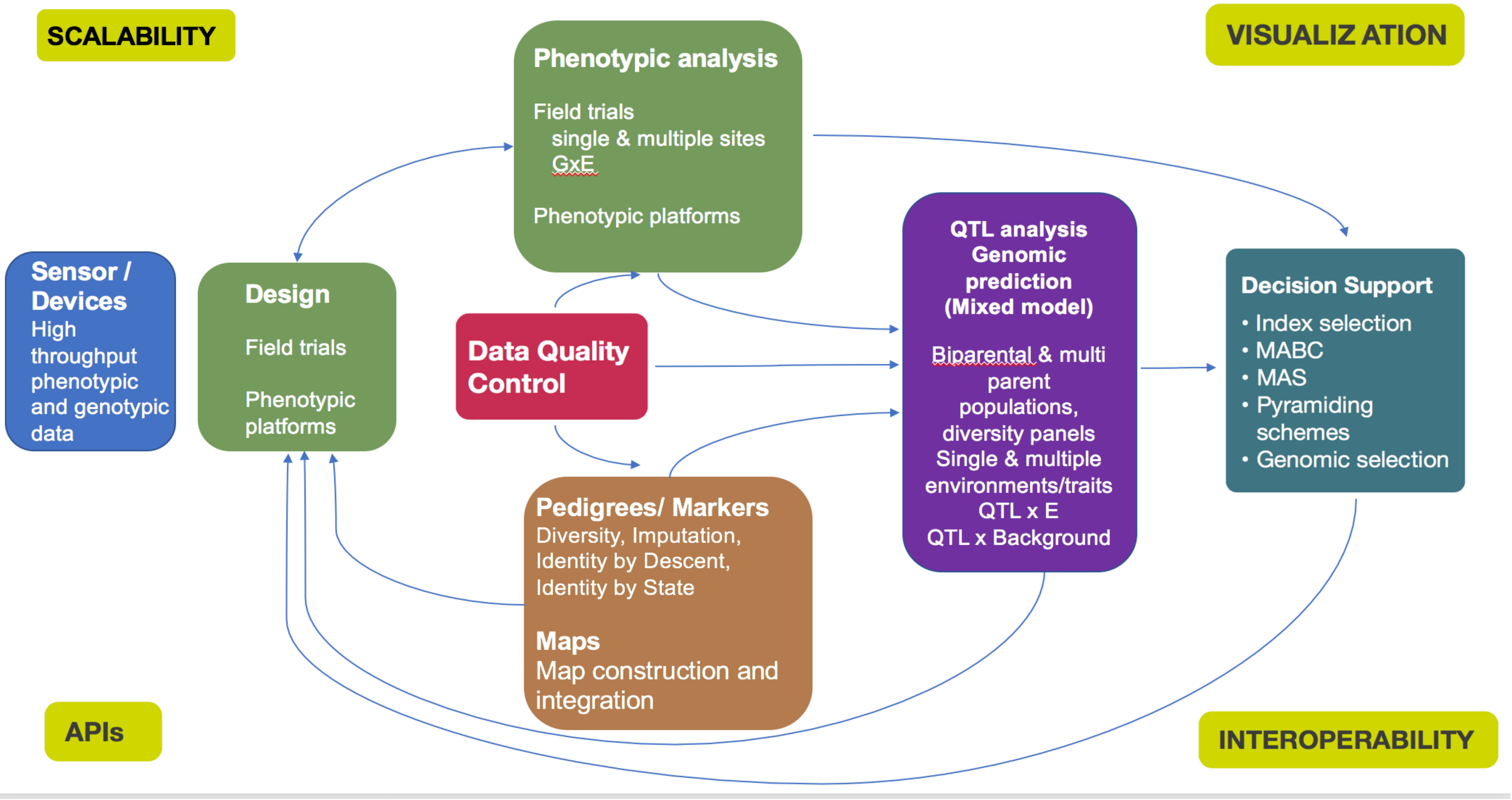
Statistical Genetics at Biometris (themes)

- Design and analysis of individual plant breeding experiments (field & platform)
  - Row x column designs, p-rep designs
  - Spatial correction by spline surfaces / mixed models
- Analysis of series of experiments and GxE
  - Linear & bilinear models (FW, AMMI, GGE)
  - Mixed models (heterogeneous genetic variances and correlations, stability variances)
- QTL / genomic prediction
  - Single / multi-environment & single / multi-trait
  - Bi-parental/ multi-parental / diversity panel
  - Inbreeders / outbreeders & diploids / polyploids
  - New types of breeding populations: MPPs / CSL
- New developments (secondary phenotyping)
  - Dynamical and functional modelling
  - Graphical modelling / causal modelling

Statistical Genetics at Biometris (Project Forms)

- STW (NWO-TTW)
- EU
- CGIAR / BMGF
- Bilateral Biometris - Companies
  - Collaborative development of methodology
  - Post doc & PhD projects
  - Consultation + education
  - Statistical genetic pipeline projects

A robust and versatile statistical genetic pipeline for plant breeding



SCALABILITY

VISUALIZ ATION

Phenotypic analysis

Field trials
single & multiple sites
GxE

Phenotypic platforms

Sensor / Devices
High throughput phenotypic and genotypic data

Design

Field trials

Phenotypic platforms

Data Quality Control

QTL analysis
Genomic prediction
(Mixed model)

Biparental & multi parent populations,
diversity panels
Single & multiple environments/traits
QTL x E
QTL x Background

Decision Support

- Index selection
- MABC
- MAS
- Pyramiding schemes
- Genomic selection

Pedigrees/ Markers
Diversity, Imputation, Identity by Descent, Identity by State

Maps
Map construction and integration

APIs

INTEROPERABILITY

P0151: A Robust and Versatile Statistical Genetic Pipeline for Plant Breeding

WAGENINGEN UR
For quality of life

VSNi

The James Hutton Institute

Dutch Wisdom: Johan Cruijff

- In theory, the theory is complicated and the practice is simple, but in practice the practice is complicated and the theory is simple