1 3 2
AFEF 5
4 6

Estimating the genetic value of F1 apple progenies for irregular bearing during first years of tree production

Jean-Baptiste Durand[1,4], **Baptiste Guitton**[2], **Jean Peyhardi**[4,6], **Yan Holtz**[3,5], **Yann Guédon**[2,3,4], **Catherine Trottier**[6], **Evelyne Costes**[3,5]

Outline:

- Quantifying alternation of flowering and use in selection schemes.
- Quantification from data at tree and axis scales.
- Genetic and ontogenic effects.
- Prediction of yields and bearing behaviour.
- QTL detection
- Practical conclusions

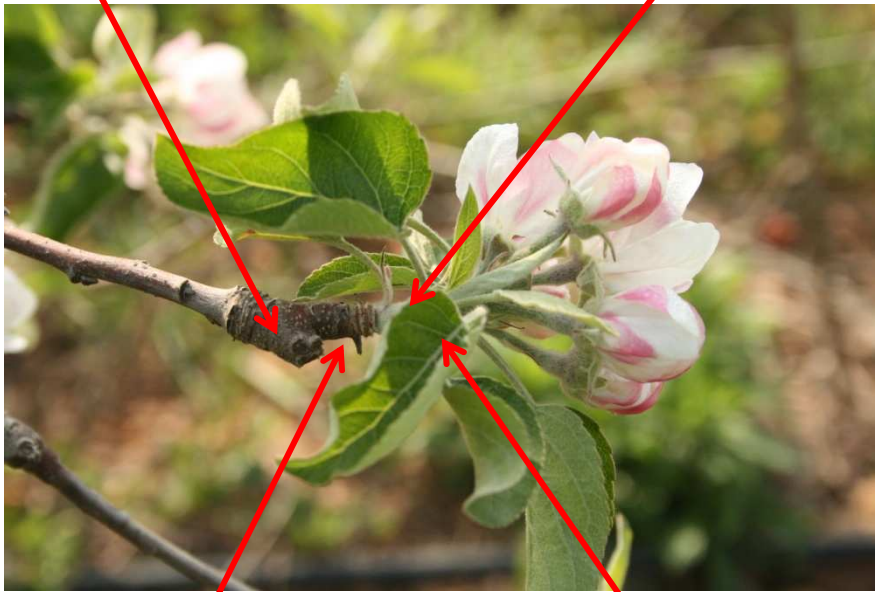# What is alternation / biennial bearing?
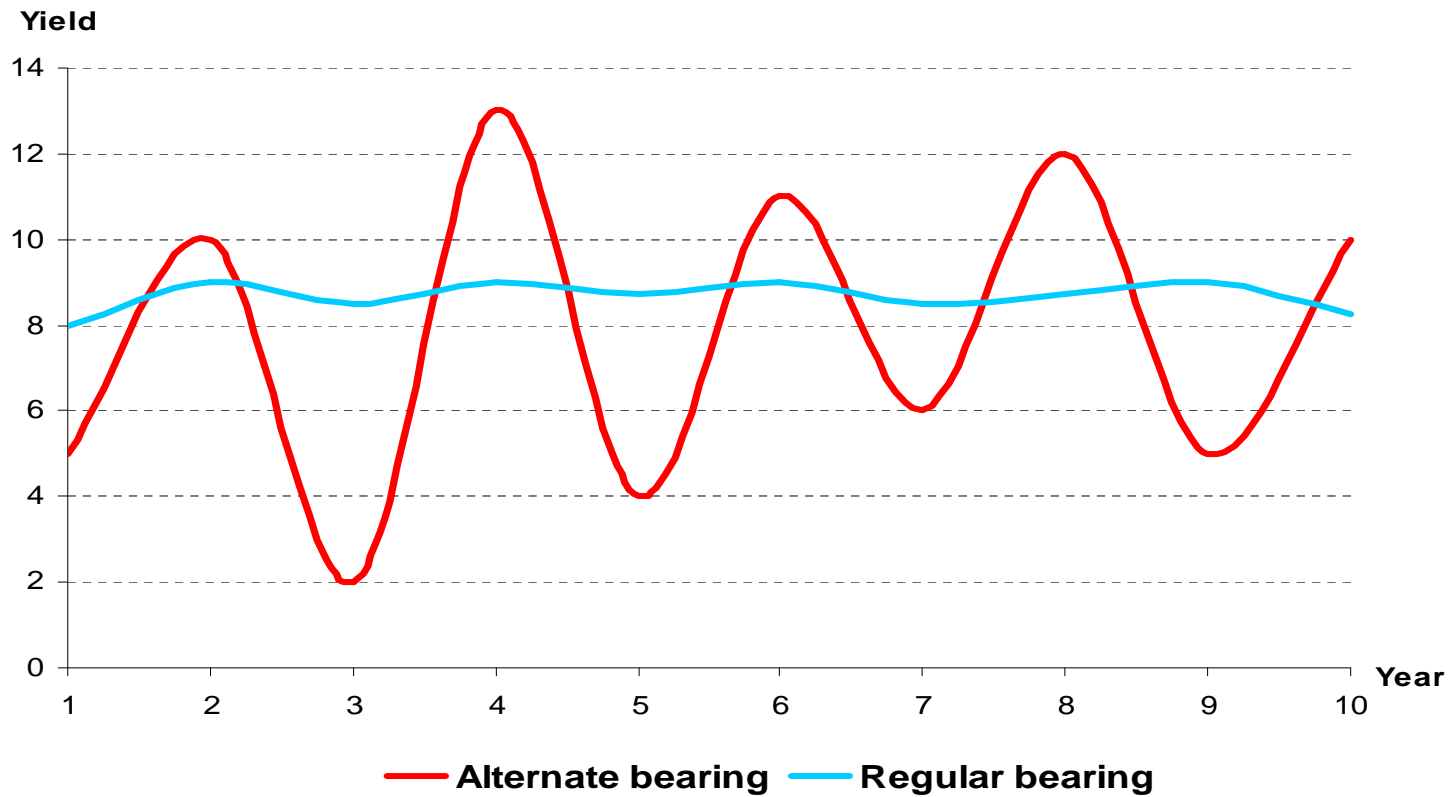
At shoot scale

Return bloom   *versus*   No return bloom

bourse (year *t*) = flowering growth unit

bourse (year *t* + 1)

bourse (year *t*)

vegetative growth unit
(year *t* + 1)

bourse shoot (year *t*)
vegetative growth unit

bourse shoot (year *t*+1)
(hidden behind the leaves)

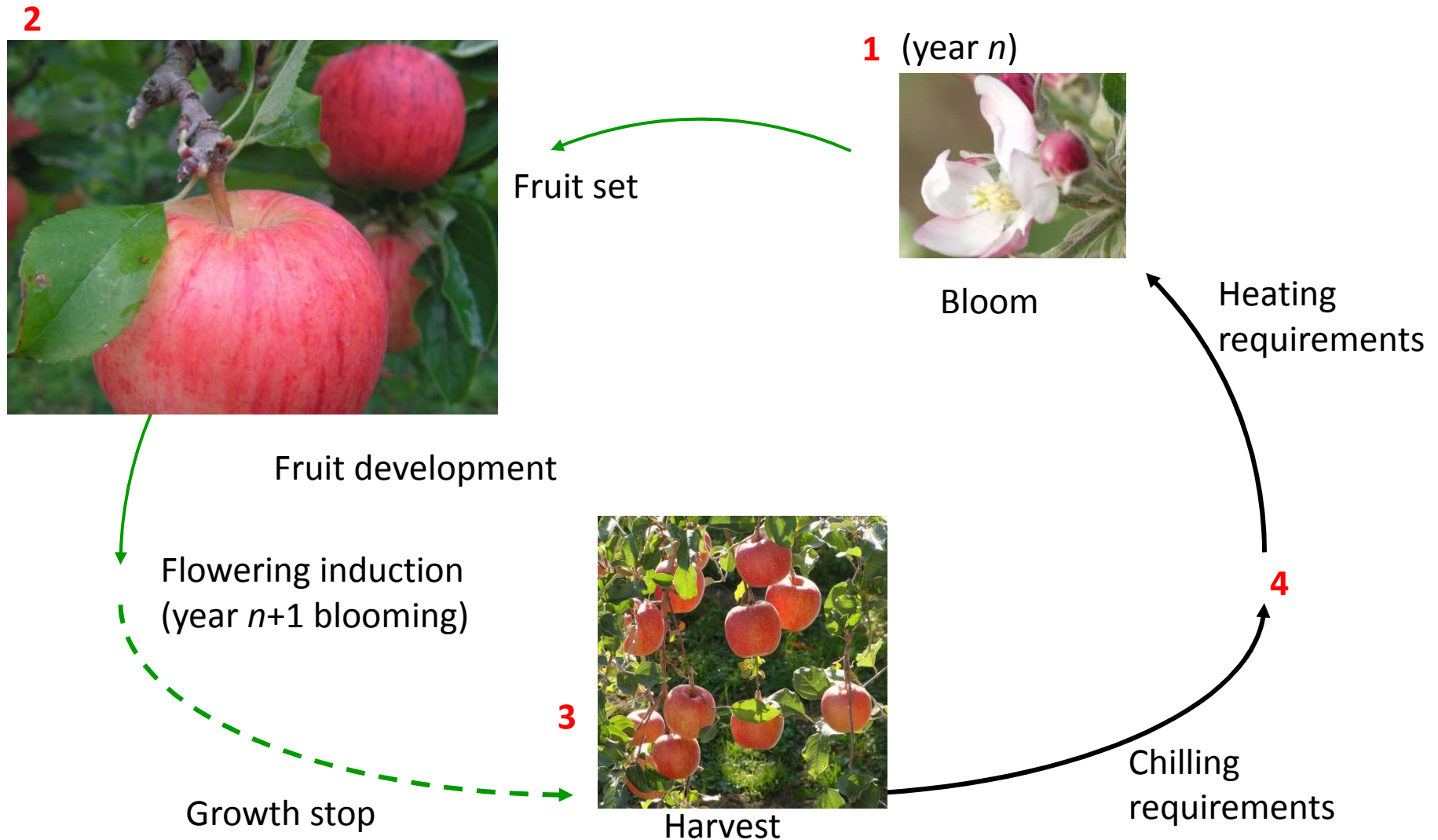bourse shoot (year *t*)

Cycle of floral induction

2

1 (year *n*)

Fruit set

Bloom

Heating requirements

Fruit development

Flowering induction (year *n*+1 blooming)

3

4

Growth stop

Harvest

Chilling requirements

Physiological hypotheses

- In apple, floral induction (FI) for next year occurs during fruit development of the current year (Foster *et al.,* 2003)

- Main putative causes of alternation
  (Monselise and Goldschmidt, 1982)
  – Carbohydrate resources: competition between « source » and « sink » organs
  – Hormonal substances originating from seeds inhibit FI
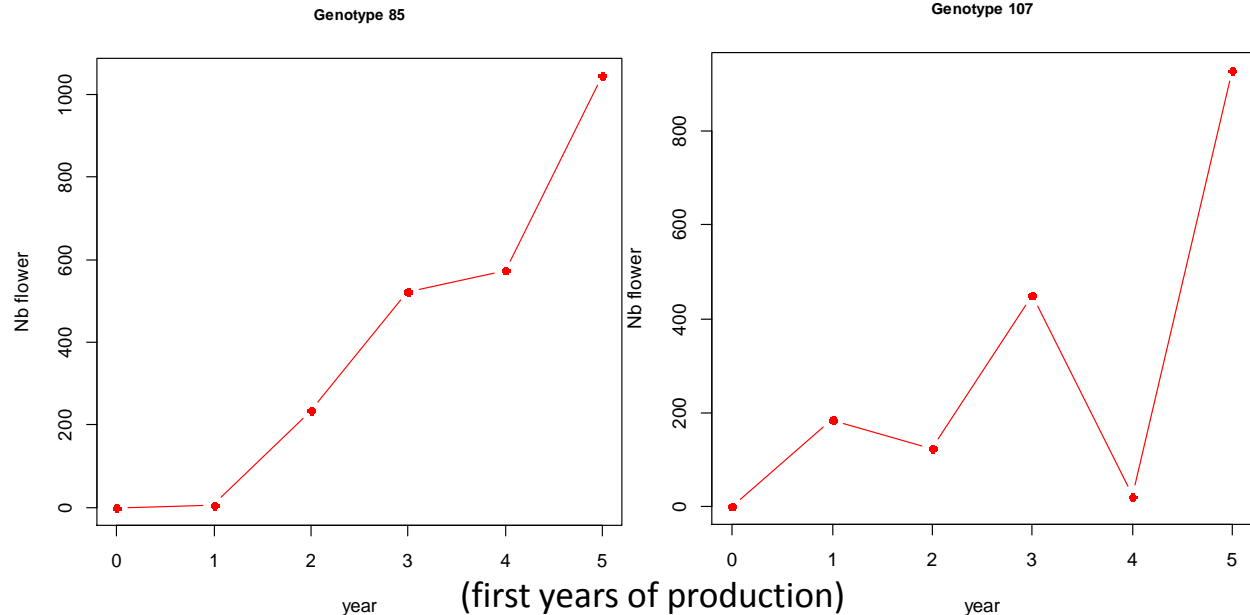   - Gibberellin
    - Auxin
    - Cytokinin

Alternation at shoot and tree scales

■ vegetative AS
■ flowering AS

● flower
[1] date / age
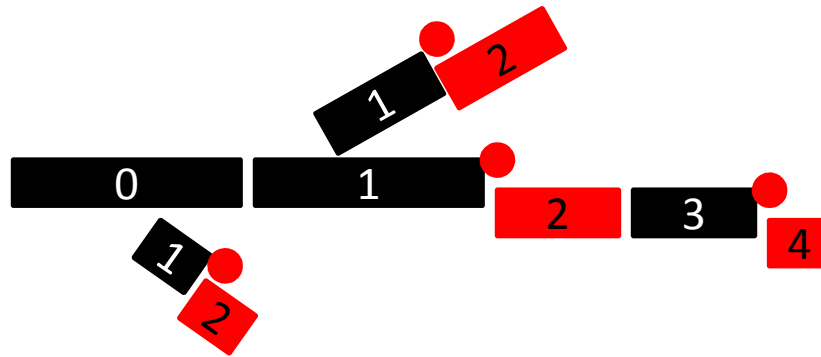
- Alternation at annual shoot (AS) scale

- Somehow related to alternation at tree scale?

Genotype 85

Genotype 107

(first years of production)

Nb flower / year

Main components of orchard production

➢Number of flowers at bloom

➔ Flowering regularity and return bloom

➢Fruit set

➢Number and position of fruits within the canopy

➔ Number of fruits per inflorescence

Alternation and agronomic issues

- Alternate and irregular bearing generates alternate and irregular incomes

- Negative impact on fruit size and quality

- Cost of fighting against alternation (thinning out):
    - manual / money (again…)
    - chemicals / environment (priceless)

- Is selection some possible way of obtaining "natural-born" regular cultivars?

Material

- F1 progenies from 'Strakrimson' x 'Granny Smith' cross (alternate v. regular bearing)

- 123 genotypes

- 2 replications per genotype

- measurements:
  - total number of inflorescences / fruits per year and per replication
  - about 20 sequences of AS (trunk, sequences along main axes or along short axes)

Some significant issues

- Annual yields subject to a trend (ontogeny)
- Quantifying regularity / alternation of yields requires detrending ($\Longrightarrow$ models)
- Quantifying the part of genetic v. individual variability in yields
- Information lost (/ gained?) in subsamples of sequences
- Connection between alternation at whole tree and axes / AS scales?
- How do irregular bearing genotypes behave?

Usual indices for alternation

Hoblyn (1936)

Wilcox (1944)

- Biennial Bearing Index / BBI for yield $Y_{g,r,t}$

$$\text{BBI} = \frac{1}{\sum_r (T_{g,r} - 1)} \sum_r \sum_{t=2}^{T_{g,r}} \frac{\left| Y_{g,r,t} - Y_{g,r,t-1} \right|}{(Y_{g,r,t-1} + Y_{g,r,t})/2}$$

⇨ =0 for constant series, =2 for perfect alternation

⇨ $BBI(T) \approx \dfrac{\ln T}{(T-1)}$ for affine series (not appropriate for trends)

⇨ not appropriate for noisy series (dependent on mean and variance of $Y_{g,r,t}$)

Huff (2001)

Usual indices for alternation

- Number of sign changes in $Y_{g,r,2} - Y_{g,r,1}, ..., Y_{g,r,t} - Y_{g,r,t-1}$
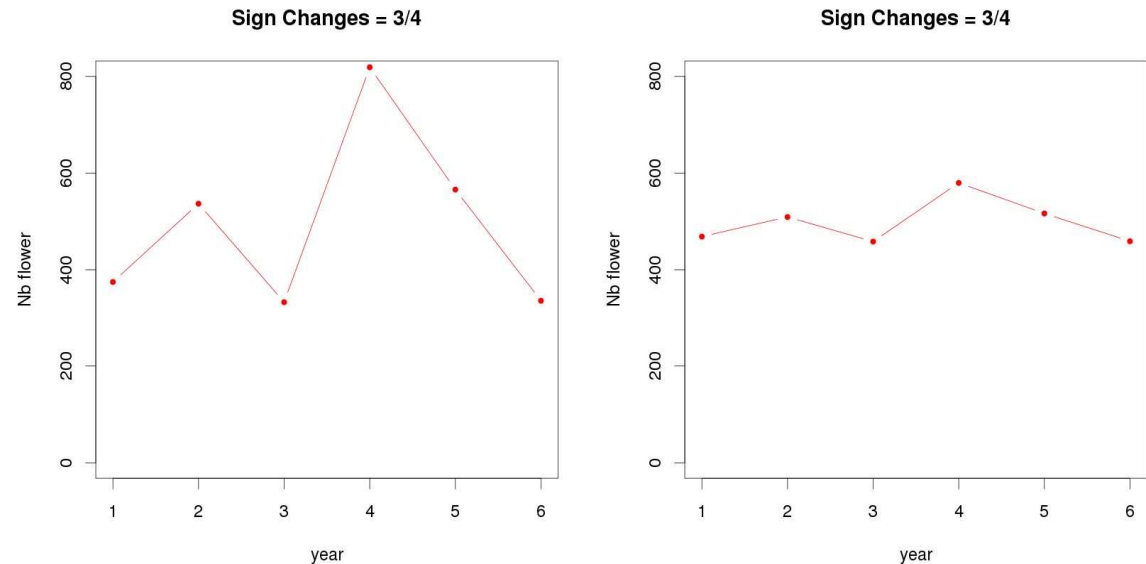
Cilas *et al.* (2011)

🙂
⇒ easily interpreted

🙁
⇒ sensitive to noise (what if $\left\{ Y_{g,r,t} \right\}_t$ is just noise?)

⇒ relative amplitudes of changes with respect to $Y_{g,r,t}$
       are not considered

Index for synchronism

- Alternation synchronism

Lauri *et al.* (1995,1997)

Among all cases of veg./fl. ▮▮ : ratio of occurrences in
   odd vs. even years (0: balance, 1: asynchronism)
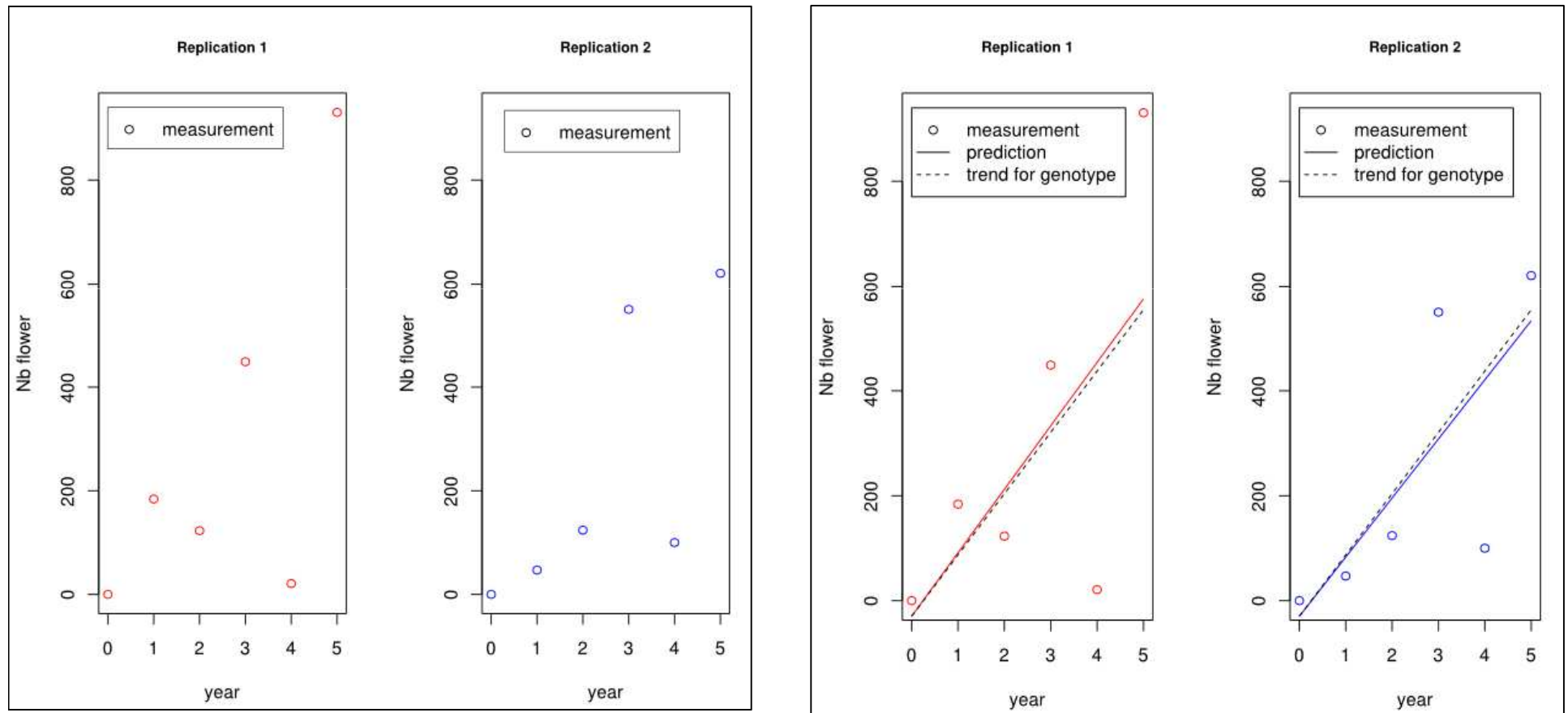☺
⇨considers both alternation and synchronism

(refinement in studies of alternation)

☹
⇨ considers both alternation and synchronism!
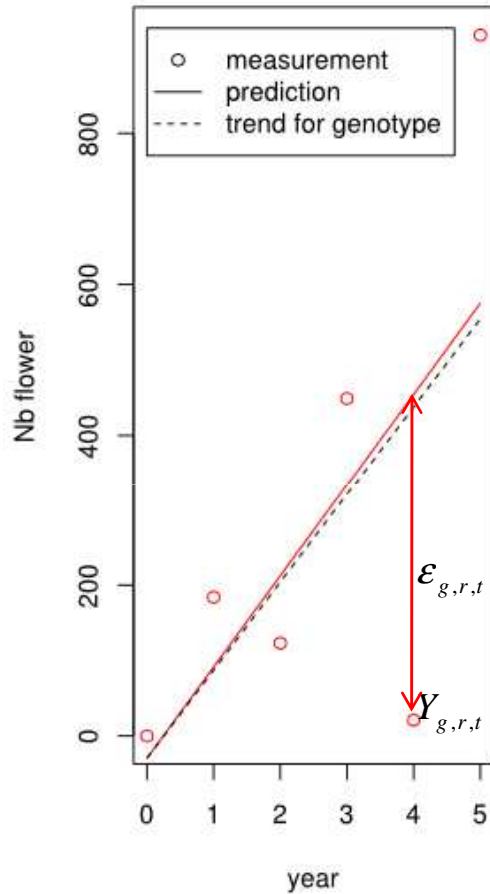
(does not quantify alternation as such)

Indices for alternation at whole tree scale

- Trend model

linear model

Trend model

$$Y_{g,r,t} = \underbrace{\beta + \beta_g + \zeta_{g,r}}_{\text{Intercept}} + \underbrace{(\alpha + \alpha_g + \xi_{g,r})t}_{\text{Slope}} + \underbrace{\varepsilon_{g,r,t}}_{N(0,\sigma^2)}$$

α slope: common part to every tree

$\alpha + \alpha_g$ slope: specific part to genotype g

$\alpha + \alpha_g + \xi_{g,r}$ slope: specific part to replication (r, g) random with variance τ_ξ^2

β intercept: common part to every tree (statistically irrelevant here)

$\beta + \beta_g$ intercept: specific part to genotype g

$\beta + \beta_g + \zeta_{g,r}$ intercept: specific part to replication (r, g) (random, statistically irrelevant here)

Linear mixed model,
Verbeke & Molenberghs (2000)

Parameter	β	α	τ_ξ	σ
Estimate	43.1	90.7	20.9	160.4
	(-108,194)	(36,145)	(13.7,24.0)	(155,168)

Residual model
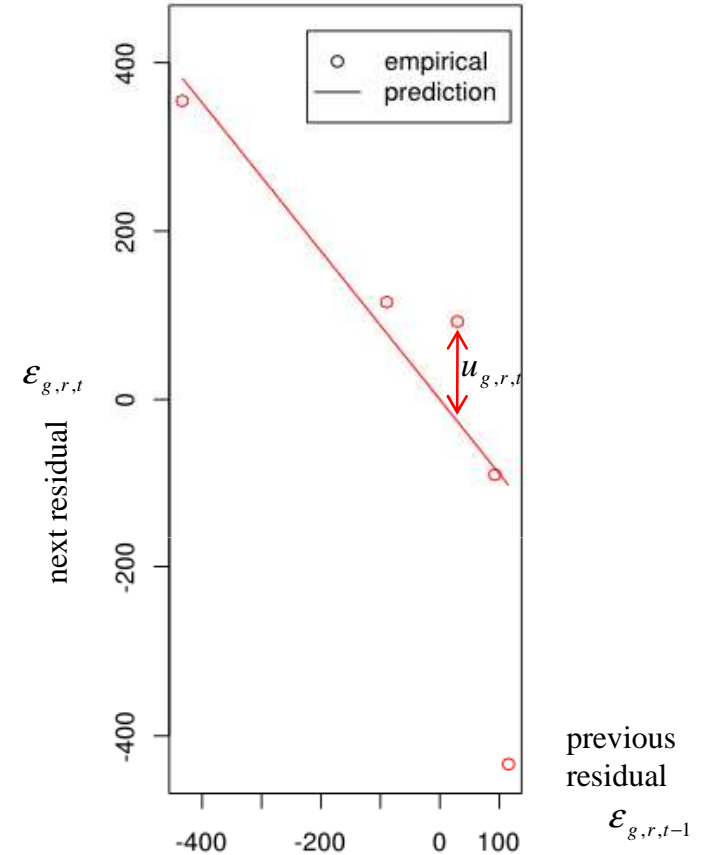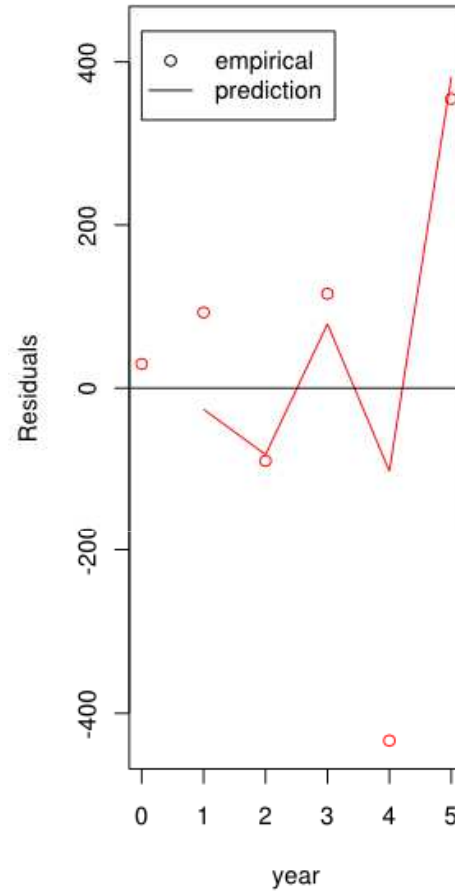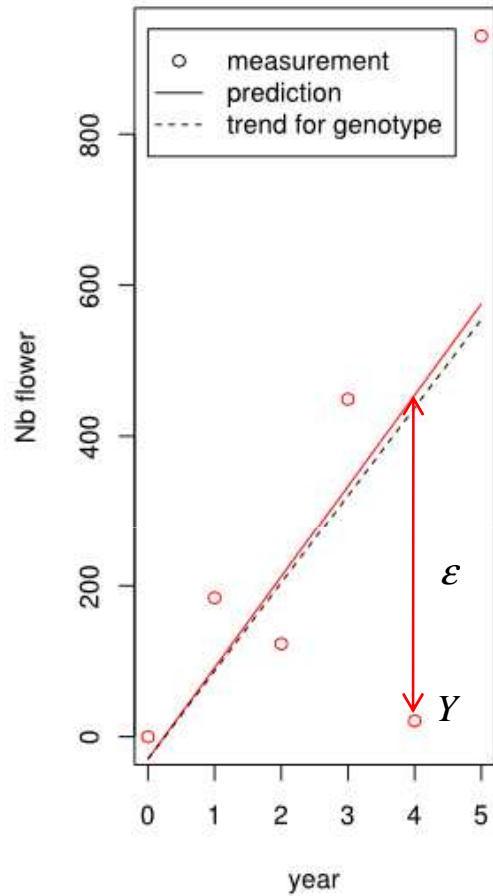
$$\varepsilon_{g,r,t} = (\gamma_g + \gamma_{g,r})\varepsilon_{g,r,t-1} + u_{g,r,t} \quad \text{w.} \quad u_{g,r,t} \sim N(0,\rho^2)$$

decomposition of slope: (genotype, replication)

$\gamma_{g,r}$ random but not relevant here

Autoregressive (AR) model on the residuals (dependencies)

$\gamma_g = -0.88$ here

Parameter	ρ
Estimate	150
	(134,172)

Pitfall of genotype AR coefficient γ_g

- The following two series would have same genotype AR coefficient
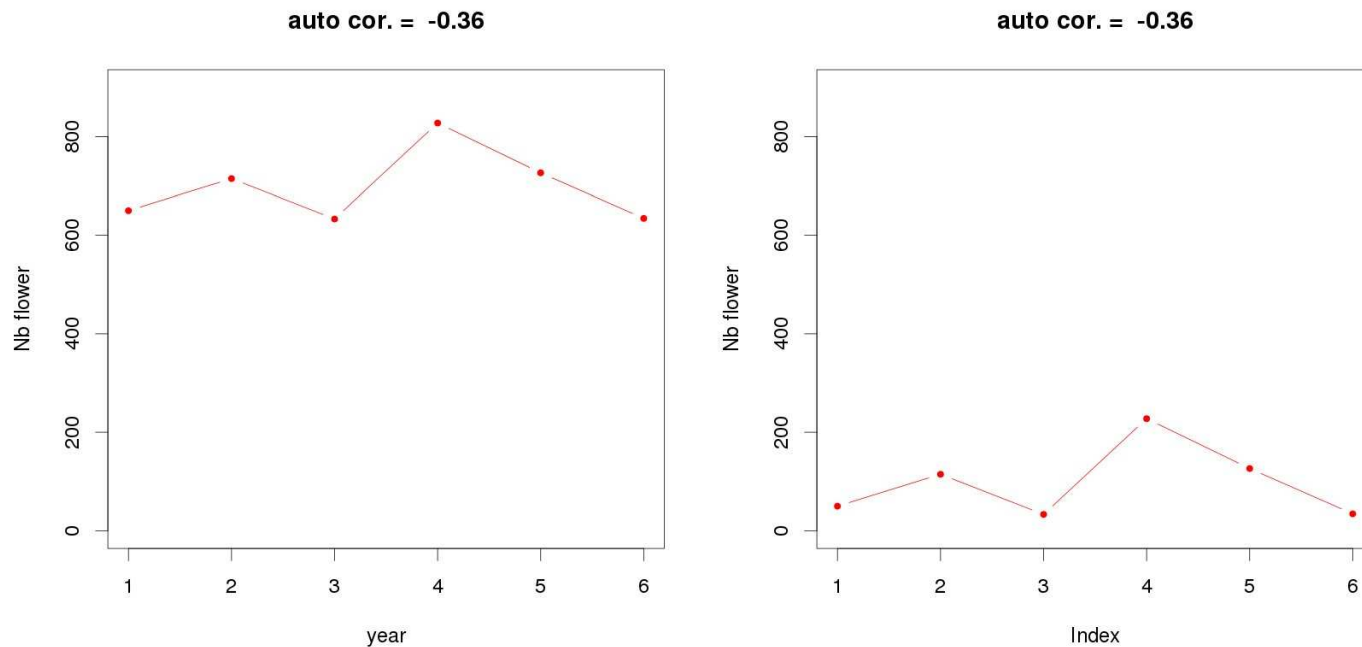
- Necessity to quantify **intensity** of alternation **too**

A BBI on residuals

- Definition

$$\text{BBI\_res\_norm} = \frac{\sum_r \sum_{t=2}^{T_{g,r}} \left| \hat{\varepsilon}_{g,r,t} - \hat{\varepsilon}_{g,r,t-1} \right| / \sum_r (T_{g,r} - 1)}{\sum_r \sum_{t=1}^{T_{g,r}} Y_{g,r,t} / \sum_r T_{g,r}}.$$

- Comments:
  - to quantify intensity of alternation as dispersion of yields around the trend
  - variance should work too (but slightly less efficient in QTL detection)
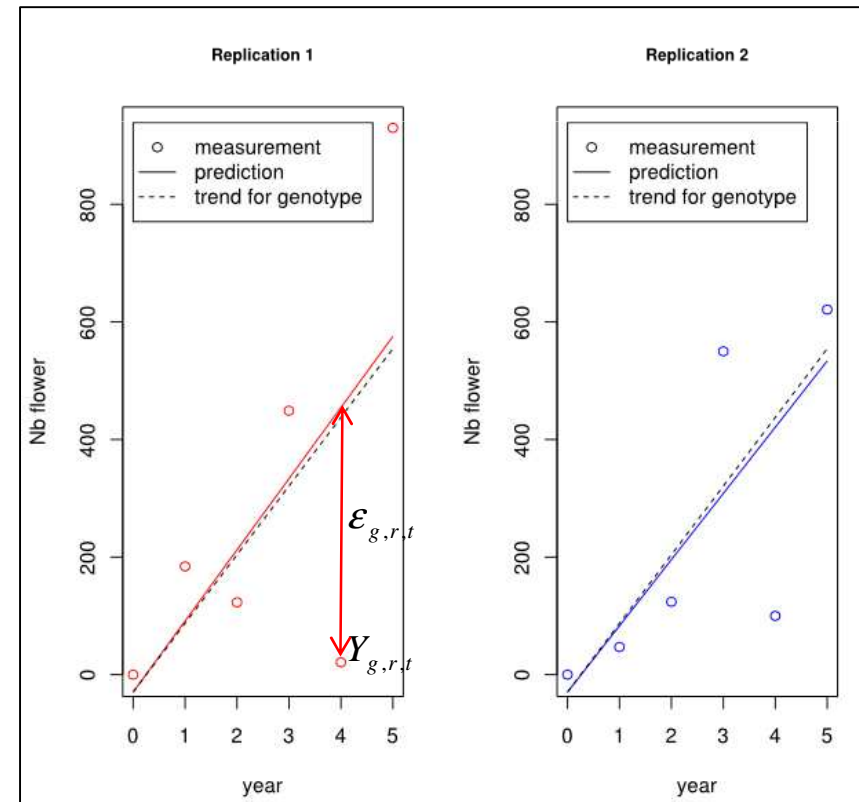
Examples on three genotypes
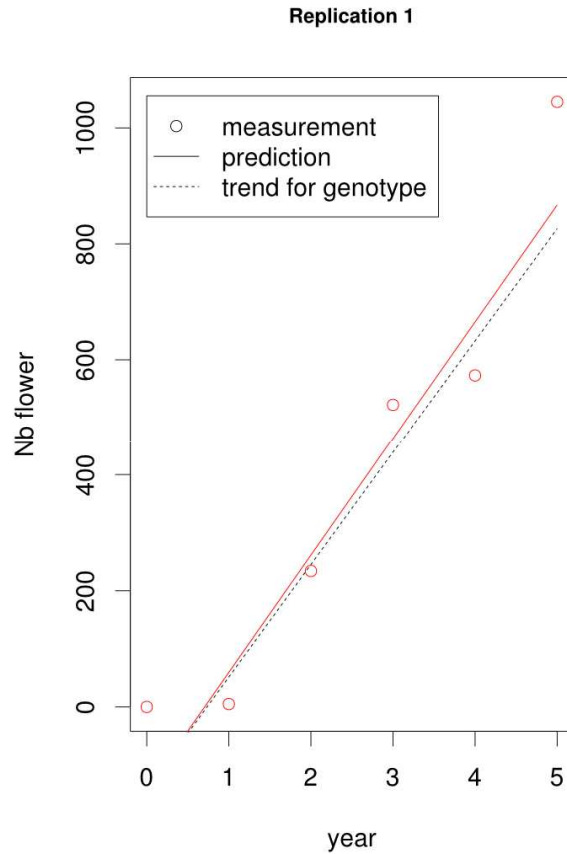
$\gamma_g = -0.07$
BBI_res_norm = 0.45

Regular
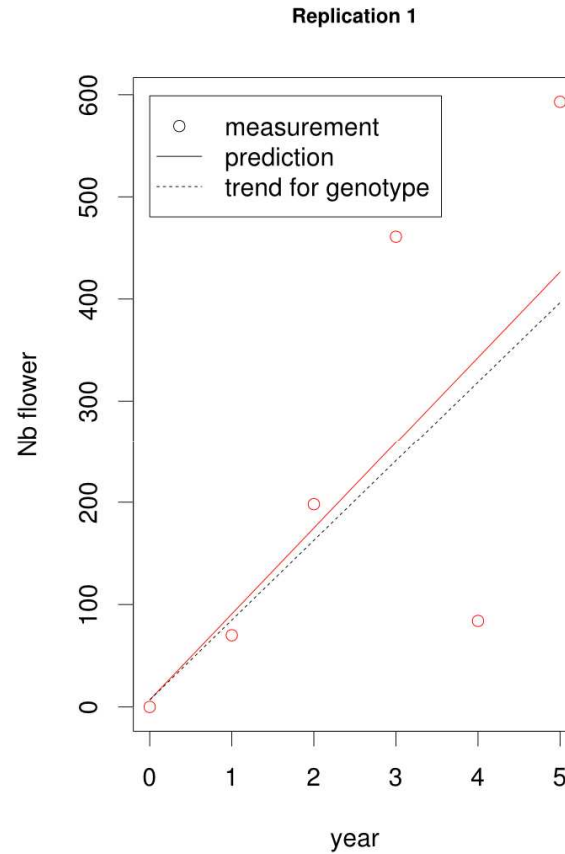
$\gamma_g = -0.88$
BBI_res_norm = 1.21

Biennial bearing
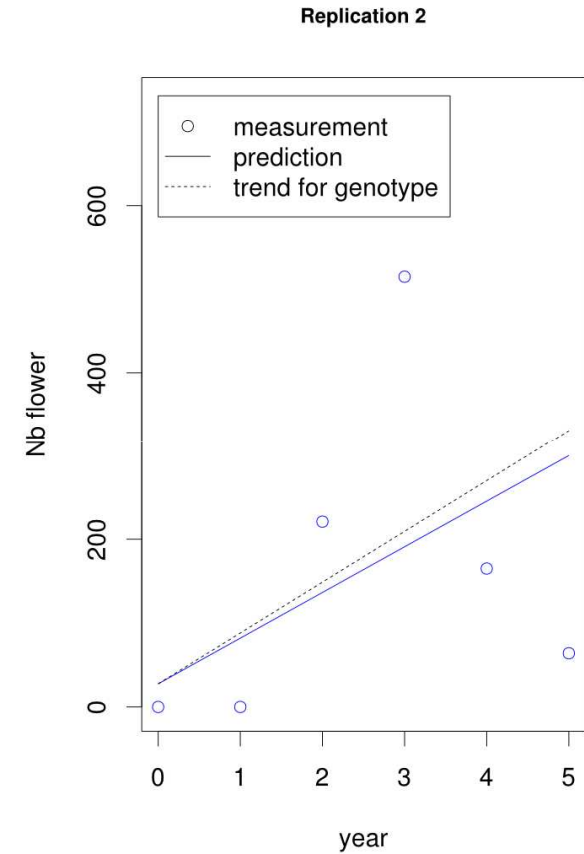
$\gamma_g = -0.28$
BBI_res_norm = 1.17

Irregular

Clustering of genotypes

- Using Gaussian Mixtures : BBI_res_norm and γ_g are assumed Gaussian within each cluster.

  McLachlan & Peel (2000)

- Their means depend on the cluster

Regular genotypes (low BBI_res_norm, $\gamma_g \approx 0$)

Biennial bearing genotypes (high BBI_res_norm, $\gamma_g \approx -1$)

Irregular genotypes (intermediate values of BBI_res_norm and γ_g)

Clusters in space (BBI.res.norm , auto.cov)
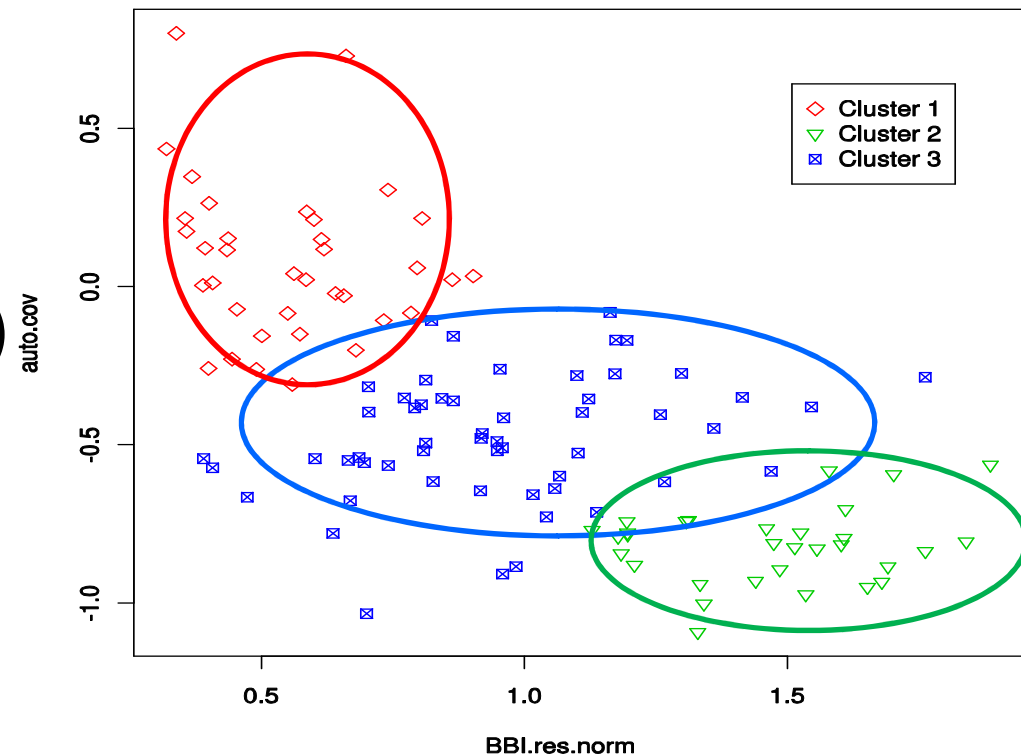
First conclusions

- Objective criteria to assess genotype bearing behaviours
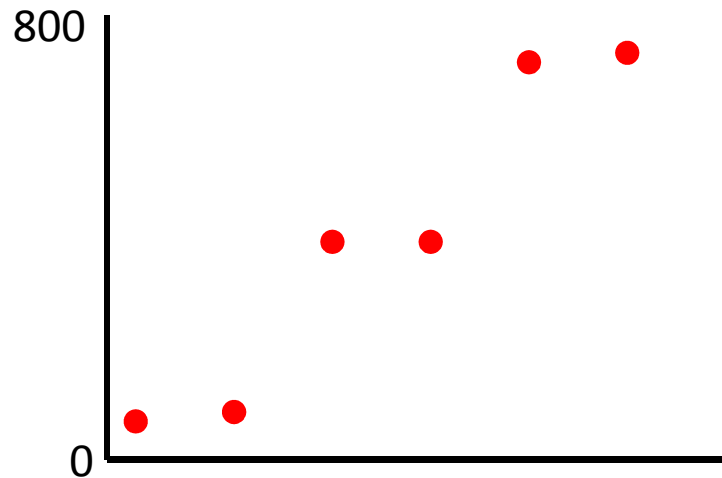
- Our models offer the possibility to predict future yields (nb flowers) and the behaviour of future genotypes

Model validation

- Reestimate the model parameters and clusters using every data except last year
- Predict yield (95% prediction interval) and clusters
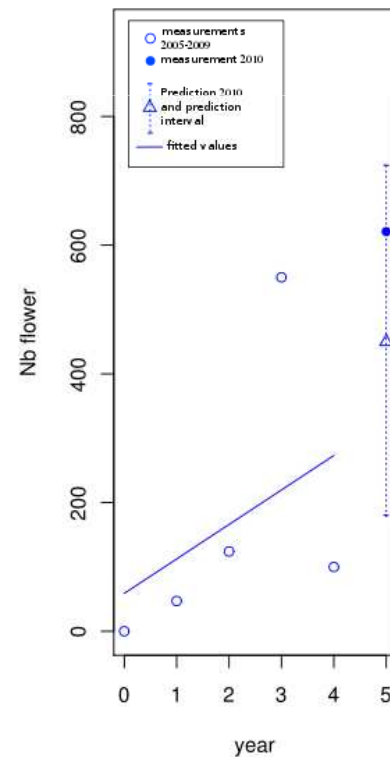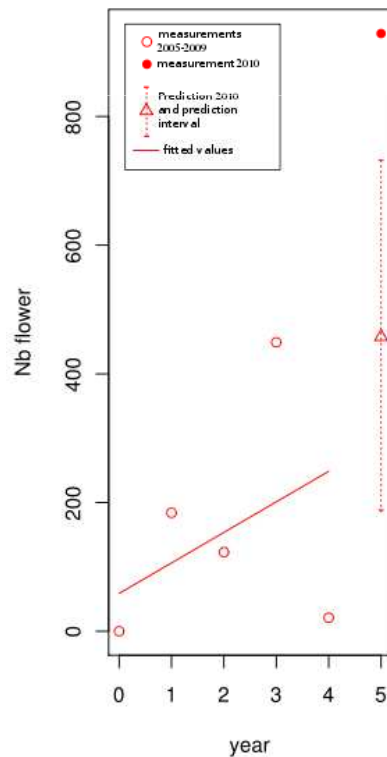
2. true value (not used to calibrate the models)

1. predicted value (not on trend due to dependent residuals)

Model validation: does it really work?

- True values in prediction intervals: 74%.
- Our models anticipate an increase of variance yields over years, but not sufficiently

$$\text{var}(\varepsilon_t) = \rho^2 \frac{\left(1 - (\gamma_g)^{2t}\right)}{1 - \gamma_g^2}; \quad \gamma_g^2 \in [0,1]$$

- Clustering: confusions between irregular and the other two bearing behaviours

		Cluster based on years 2005-2010 (truth)		
		Reg.	Bien. bear.	Irreg.
Cluster based on years 2005-2009 (prediction)	Reg.	28	0	5
	Bien. bear.	1	25	5
	Irreg.	7	6	44

Phenotyping at axis scale: what is at stake ?

- Our descriptors can be used for early selection of regular genotypes.

- But phenotyping the total number of inflorescences is much too costly (not realistic in practice for now).

- Descriptors have to be deduced from the sequences of AS.

Quantifying synchronism in flowering

- Rationale: *alternation at tree scale* should result from alternation at axis scale and synchronism

- *Regularity at tree scale* could result from complex configurations at axis scale

- A descriptor for synchronism at tree / genotype scale: entropy   Bishop (2006)

$$Ent_{g,r} = -\sum_{t} \left(f_{g,r,t,0} \log f_{g,r,t,0} + f_{g,r,t,1} \log f_{g,r,t,1} \right)$$

To average on replications

Unpredictability of flowering at year t :
0 if $f_{g,r,t,0}$ = 0 or 1 (perfect prediction) and maximal if $f_{g,r,t,0}$ = 0.5 (just toss a coin!)

Illustration: typical values of entropy

Genotypes		Year					Entropy
		2005	**2006**	**2007**	**2008**	**2009**	
g=85	**Number of GUs**	2	6	10	14	18	Total 50
	F$_{g,r,t}$	0.00	0.00	0.20	0.64	0.72	
	Ent$_{g,r}$	0.00	0.00	0.50	0.65	0.59	0.50
g=107	**Number of GUs**	2	5	9	11	11	Total 38
	F$_{g,r,t}$	0.00	0.60	0.0	1.0	0.0	
	Ent$_{g,r}$	0.00	0.67	0.00	0.00	0.00	0.09
g=108	**Number of GUs**	5	9	13	15	12	Total 54
	F$_{g,r,t}$	0.00	0.22	0.15	0.53	0.00	
	Ent$_{g,r}$	0.00	0.53	0.43	0.69	0.00	0.38

regular ⟶ *g*=85

biennial ⟶ *g*=107

irregular ⟶ *g*=108

$F_{g,r,t}$: frequency of flowering at AS scale

minimal values of entropy   maximal values of entropy   mean values of entropy

Using BBI_res_norm and γ_g on AS sequences

- Entropy can be computed from the sequences of AS

- Total number $Y^{loc}_{g,r,t}$ of inflorescences in sequences as an approximation of $Y_{g,r,t}$

- Computing BBI_res_norm and γ_g from $Y^{loc}_{g,r,t}$ yields two descriptors:

$$BBI\_res\_norm^{loc} \text{ and } \gamma^{loc}_g$$

- These are approximations of

$$BBI\_res\_norm \text{ and } \gamma_g \text{, respectively}$$

Approximation of descriptors at tree scale from descriptors at AS scale

- How good are they?

	coefficient γ_g	Local BBI_res_norm	Local genotype AR coefficient γ_g^{loc}	Mean entropy
BBI_res_norm	-0.66 (-0.75;-0.55)	**0.72** (0.61;0.80)	-0.58 (-0.69;-0.45)	-0.49 (-0.62;-0.34)
γ_g	1	**-0.61** (-0.71;-0.49)	0.55 (0.41;0.67)	0.33 (0.16;0.48)
Local BBI_res_norm			1	-0.63 (-0.73;-0.51)	-0.66 (-0.75;-0.54)
Local genotype AR coefficient γ_g^{loc}			1	0.33 (0.15;0.48)

Correlations (with confidence intervals)

- Multiple linear regressions

$$BBI\_res\_norm_g = \alpha_0 + \alpha_1 B_g^{loc} + \alpha_2 \gamma_g^{loc} + \cancel{\alpha_3} \overline{Ent}_g + \eta_g^{(1)} \qquad R=0.74$$

$$\gamma_g = \beta_0 + \beta_1 B_g^{loc} + \beta_2 \gamma_g^{loc} + \cancel{\beta_3} \overline{Ent}_g + \eta_g^{(2)} \qquad R=-0.65$$

Prediction of bearing behaviour from descriptors at AS scale

- Gaussian mixture *clustering* of genotypes from
$$BBI\_res\_norm^{loc} \text{ and } \gamma_g^{loc} \text{ (}unsupervised\text{)}$$

- Does not take profit from the knowledge of the "true" clusters (realistic practical situation)

1: regular
2: biennial
3: irregular

regular ↔ biennial
irregular ↔ other

		Cluster C_1 on global indices ("truth")		
		1	**2**	**3**
Cluster C_2 on local indices (approx.)	**1**	33	2	29
	2	0	18	4
	3	2	9	18

Global error rate: 40%

Assessing the validity of predictions obtained from descriptors at AS scale

- Clustering: no information available at tree scale / new cross?

- Classification: use information available at tree scale / new descendants from same parents?

1: regular
2: biennial
3: irregular

		"True" class C_1		
		1	2	3
Predicted class C_2	1	17	1	9
	2	0	20	5
	3	18	8	37

Global error rate: 36%

Classification: not much better than clustering here.

Graphical interpretation

- Plane where the 3 clusters have maximal separation

- Regular / biennial bearing: good separation

- Irregular genotypes are mixed between regular and biennial bearing genotypes

- NB. x-axis provides some scoring of the genotypes

FDA, Tabachnick & Fidell (2007)

1: regular
2: biennial
3: irregular

Clusters of genotypes in first FDA plane
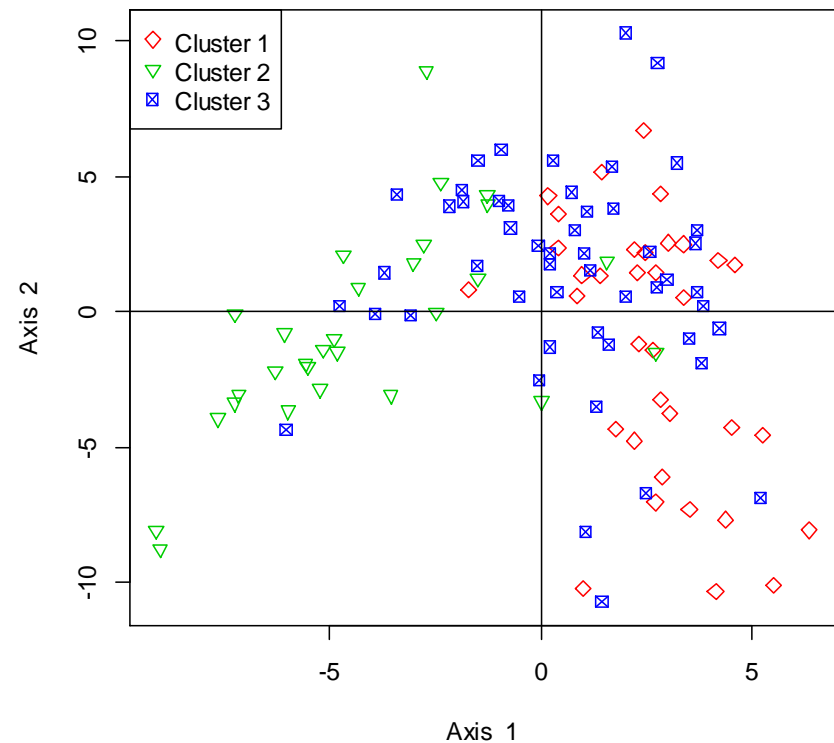
A practical scheme for selection

- Some genotypes start exhibiting irregular or biennial bearing at some age (unknown in advance).
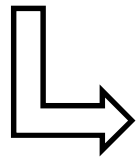
  1. Progressively suppress biennial or irregular genotypes after the first observation of a large decrease in flowering during the beginning of mature phase

  2. Confirm the regular fruiting behaviour of the pre-selected genotypes during stable mature phase.

Taking profit from information contained in sequences

- Between 4 and 45 sequences per genotype, containing information on alternation in flowering.

year *t*

	2006	2007	2008	…
axis 1	0	0	1	
axis 2		1	1	
axis 3		0	0	
axis 1	0	1	0	
…				

genotype *g*

replication *r* in *g*

Taking profit from information contained in sequences

- Between 4 and 45 sequences per genotype, containing information on alternation in flowering.

year t

	2006	2007	2008	…
axis 1	0	0	1	
axis 2		1	1	
axis 3		0	0	
axis 1	0	1	0	
…				

genotype g replication r in g

memory m=01   site ℓ

$$\log \frac{P(F_{g,r,t,m,\ell} = 1)}{P(F_{g,r,t,m,\ell} = 0)} = \lambda + \mu_m + \phi_t + \theta_{g,m} + \eta_{g,t} + \zeta_{g,r}$$

2nd-order Markov chain model

Markov chain as GLMM

$$\log \frac{P(F_{g,r,t,m,\ell} = 1)}{P(F_{g,r,t,m,\ell} = 0)} = \lambda + \underbrace{\mu_m + \phi_t}_{\text{fixed effects}} + \underbrace{\theta_{g,m} + \eta_{g,t} + \zeta_{g,r}}_{\text{random effects}}$$

μ_m common effect of memory m ($m \in$ {00, 01, 10, 11}) / Markovian part

ϕ_t common effect of year t ($t \in$ {2006, 2007, 2008, 2009})

$\theta_{g,m}$ correction to transition probability from memory m due to genotype g

$\eta_{g,t}$ correction to probability of flowering due to interaction year t x genotype g

$\zeta_{g,r}$ correction to probability of flowering due to replication r of genotype g

Nature	Intercept	Fixed effects : memory			Fixed effects : year			Variances		
Parameter	λ	μ_{10}	μ_{01}	μ_{11}	ϕ_{2007}	ϕ_{2008}	ϕ_{2009}	$\theta_{g,m}$	$\eta_{g,t}$	$\zeta_{g,r}$
Estimate	-1.68	**2.69**	0.81	**1.29**	0.54	**2.21**	0.17	0.70	1.48	**0.36**
p-value	1e-12	1e-16	1e-7	1e-10	0.05	1e-16	0.6	-	-	-

Behaviour at AS scale: biennial

- lowest probability to flower at year t after flowering at year t - 1 ($\textcolor{red}{\theta_{g,01}}$ and $\theta_{g,11}$)
- lowest probability to flower in 2009 ($\eta_{g,2009}$) → "off" year for the majority of biennial bearing genotypes
- Note that some biennial bearing genotypes have odd "on" years and even "off" years
- lowest entropy (synchronism at axis scale)

Behaviour at AS scale: regular

- highest probability to flower at year *t* after flowering at year *t* - 1 ($\theta_{g,01}$ and $\theta_{g,11}$)

- highest probability to flower in 2009 ($\eta_{g,2009}$) →"off" year

- highest entropy (asynchronism and some kind of irregularity at axis scale)

Behaviour at AS scale: irregular

- Intermediate values of descriptors at both scales ($\theta_{g,01}$, $\theta_{g,11}$, $\eta_{g,2009}$) + (BBI_res_norm, γ_g)
- Exception: high entropy (as high as for regular genotypes)
- Suggests that axes have a mixed behaviour between regular and biennial bearing (each axis, or half the axes for example?)

QTL detection: descriptors at tree scale

- BBI_res_norm $\rightarrow$ 2 QTL in 2 separated genomic regions (linkage groups 1 and 8, LOD scores 6.17 and 5.27).

  Trait variability explained: 17.7-20%

- Corroborates zones previously identified in Guitton *et al.* (2012). QTL cluster on LG1 co-locating with those for inflorescence yield at year *t* and fruit yield at *t*-1.

- No QTL found for γ_g.

QTL detection: descriptors at AS scale

- $BBI\_res\_norm^{loc} \rightarrow 1$ QTL in LG 8, LOD score 4.32 (same LG as BBI_res_norm)

- $\gamma_g \rightarrow 2$ QTL in LG11 and LG14 (LOD scores 7.24 and 4.55). Unreported zones for flowering.

  Trait variability explained: 22.5% and 13.5%

- $\theta_{g,01} \rightarrow 2$ QTL in LG1 and LG8 (same LG as BBI_res_norm)

- $\theta_{g,00} \rightarrow 2$ QTL in LG10 and LG13 (precocity)

- No QTL found for entropy, $\theta_{g,10}$, $\theta_{g,11}$.

Summary of main results

- New descriptors at two scales to characterize alternation (relevance for other species).
- Possible selection strategy from subsamples, with elimination of biennial and some irregular bearing genotypes.
- Interpretation of regularity and irregularity as the result of behaviour at axis scale.
- Confirmation of the role of some genomic loci.
- New genomic loci (to be investigated, more genotypes...)

Current work and perspectives for future research

- Possibility of using local descriptors on new progenies (Y. Holtz's Master Thesis, PhD and projects in AFEF team)

- Investigate the two physiological hypotheses on alternation (competition? hormones?)

- Connection between branching and desynchronisation? (P. Fernique's PhD, VP team)

Relation between biennial bearing and tree architecture

- » Type I and II: high tendency to biennial bearing
- » Type III and IV: low tendency to biennial bearing
- » Architecture traits:
  - » Internode length
  - » Branching intensity
  - » Extinction rate (Lauri *et al*, 1995 et 1997)

Type I   Type II   Type III   Type IV

Literature cited

- **Bishop CM. 2006**. Pattern Recognition and Machine Learning. Springer Verlag, 2006.
- **Cilas C, Montagnon C, Bar-Hen A**. **2011**. Yield stability in clones of *Coffea canephora* in the short and medium term: longitudinal data analyses and measures of stability over time. *Tree Genetics & Genomes* **7**: 421-429.
- **Foster T, Johnston R, Seleznyova A**. 2003. A Morphological and Quantitative Characterization of Early Floral Development in Apple (*Malus × domestica* Borkh.). The Annals of Botany **92**(2): 199-206.
- **Guitton B, Kelner J-J, Velasco R, Gardiner SE, Chagné D, Costes E**. **2012**. Genetic control of biennial bearing in apple. *Journal of Experimental Botany* **63**: 131-149.
- **Hoblyn TN, Grubb NH, Painter AC, Wates BL**. **1936**. Studies in biennial bearing. *International Journal of Pomology and Horticultural Science* **14**: 39-76.
- **Huff A**. **2001**. A significance test for biennial bearing using data resampling. *Journal of Horticultural Science & Biotechnology* **76**: 534-535.

- **Lauri P-É, Térouanne E, Lespinasse J-M**. **1997**. Relationship between the early development of apple fruiting branches and the regularity of bearing - An approach to the strategies of various cultivars. *Journal of Horticultural Science* **72**: 519-530.

- **Lauri P-É, Térouanne E, Lespinasse J-M, Regnard J-L, Kelner J-J**. **1995**. Genotypic differences in the axillary bud growth and fruiting pattern of apple fruiting branches over several years - an approach to regulation of fruit bearing. *Scientia Horticulturae* **64**: 265-281.

- **McLachlan GJ, Peel D. 2000.** *Finite Mixture Models*. New York, Wiley

- **Monselise SP, Goldschmidt EE**. **1982**. Alternate bearing in fruit trees. *Horticultural Reviews* **4**: 128-173.

- **Tabachnick B, Fidell L**. **2007**. *Using multivariate statistics*. New York, USA: Harper and Row Publishers.

- **Verbeke G, Molenberghs G**. **2000**. *Linear mixed models for longitudinal data*. New York, USA: Springer.

- **Wilcox J. 1944**. Some factors affecting apple yields in the Okanagan Valley: tree size, tree vigor, biennial bearing, and distance of planting. *Scientific Agriculture* **25**: 189.

Bonus

Physiological hypotheses

- In apple, floral induction (FI) for next year occurs during fruit development of the current year (Foster *et al.,* 2003)

- Main putative causes of alternation
  (Monselise and Goldschmidt, 1982)
  - Carbohydrate resources: competition between « source » and « sink » organs
  - Hormonal substances originating from seeds inhibit FI
    - Gibberellin
    - Auxin
    - Cytokinin

  Mutations in flowering genes may induce different sensibility to hormonal signals and different regulatory networks leading to FI
  - Flowering integrator genes
  - Meristem identity genes

QTL detection: descriptors at AS scale (1)

- *BBI_res_norm^{loc}* $\rightarrow$ 1 QTL in LG 8, LOD scores 4.32 (same LG as BBI_res_norm)

- γ_g $\rightarrow$ 2 QTL in LG11 and LG14 (LOD scores 7.24 and 4.55). Unreported zones for flowering.

  Trait variability explained: 22.5% and 13.5%

QTL detection: descriptors at AS scale (2)

- $\theta_{g,01} \rightarrow 2$ QTL in LG1 and LG8 (same LG as BBI_res_norm).

- $\theta_{g,00} \rightarrow 2$ QTL in LG10 and LG13 (precocity). LOD scores 8.46 and 4.66).

- No QTL found for entropy.