Tuesday am

- Trial design and analysis
  - Easy exposition of principles of good experimental design
  - Easy introduction to the mixed model

Design and analysis of variety trials

The Three R’s

Replication

Randomisation

Restraint

Design and analysis of variety trials

Replication

We need a source of error for significance testing and to assign precision:

You need at least two observations to calculate a variance (more are preferable).

Design and analysis of variety trials

Randomisation

Simple explanation: we need to make fertility fluctuations fair:

equally likely for affect all varieties.

Subtle explanation: only randomisation will supply unbiased estimates of error:

Fertility	`HLHLHLHLHL`	
Vars	`1212121212`	low error and varieties apparently differ
Vars	`1111122222`	high error, varieties do not differ but “bad” trial.
Vars	`1121122122`	gorgeous.

Design and analysis of variety trials

Restraint

Blocking: assign varieties in groups to sections of the field where fertility effects are expected to be similar.

Still randomise within blocks to get unbiased estimate of error.

Simplest case is to randomise complete replications:

Randomised Complete Block Design.

A proportion of the error variance is sucked up by the difference between blocks and the precision of variety estimates is improved.

See course notes for examples.

Randomised Complete Block Design

Suitable for small numbers of varieties in uniform fields

Can design, randomise and analyse by hand.

Design and analysis of variety trials

Incomplete Block designs

Small blocks, so not all varieties are in all blocks.

Better control of variability in field trials

Model: $y_{ij} = \mu + v_i + b_j + e_{ij}$

Small blocks are usually still grouped into complete replications – the design is “resolvable.” Good if you don’t want to score the whole experiment for some traits.

Design and analysis of variety trials

Incomplete Block designs

Blocking in two dimensions:

row and column designs.

Model: $y_{ij} = \mu + v_i + b_j + r_{jk} + c_{jl} + e_{ij}$

Design and analysis of variety trials

Creating incomplete block designs:

Approach one:

1) Use a standard design in Cochran and Cox

2) If there isn’t a suitable one, drop varieties, or double up smaller designs.

This was state of the art 1940:

Described as “Procrustean design” (Mead).

Many breeders and testing authorities still work this way.

"it isn't wrong but we just don't do it“ Gordon the Blue Engine.

Design and analysis of variety trials

Creating incomplete block designs:

Approach two:

Using computer software, create a design to fit the experiment.

Alpha designs and alpha alpha designs are a good example.

Some options to create designs within GenStat.

Or better, use http://biometrics.hri.ac.uk/DesignOfExperiments/

or DIGGER http://www.austatgen.org/files/software/downloads/

Design and analysis of variety trials

Recovery of Interblock information

Variety effects can also be estimated by minimizing the variance between blocks. E.g. two varieties in different blocks:

(v1+b1) – (v2+b2) is an estimate of v1 – v2

The best estimate of variety effects is a weighted mean of the intra- and inter-block estimates

This leads to:

Random effects

Fixed effects

REML

Random & fixed effects

Trial analyses: varieties fixed blocks random.

Why not varieties random blocks random?

Often (but not always) this is justified and has advantages.

For varieties, the equivalent of the recovery of inter-block information is the incorporation of information from relatives.

However, even if all varieties are unrelated (or more strictly they are all of equal relationship), treating varieties as random has an effect:

Varieties are samples from a population with variance Vg.
Measurement error (between plots say) is Ve. No reps = r

There are two estimates of a variety effect: *Bayesian view*

Using information from relatives:

g_r (=0) with error Vg and weight 1/Vg *prior*

From the observed effect:

g_o, with error Ve/r and weight r/Ve *data*

The weighted effect is then:

$g_w = g_r .Vg/(Vg+Ve/r)$ *posterior*

ie the observed effect x the heritability of the mean

g_w (the BLUP) is shrunk compared to g_o (the BLUE)

Useful for selecting among varieties with unequal replicate numbers.

Design and analysis of variety trials

Plot size and block shape

What do you think?

Plot size:	statistical advice is as small as possible, otherwise long and thin.
Block size:	can decide empirically using historical data.
Starting point:	square blocks (with long and thin plots).

Design and analysis of variety trials

Unreplicated trials

Reasons:

1) Lack of seed

2) May give greatest response to selection

Options for design?

Regular checks

Partial replication – augmented designs

Pedigree relationships

Spatial analysis

Randomise.

Design and analysis of variety trials

Spatial Analysis

Exploits the observation that almost always, correlation between plot performance is higher the closer the plots are to each other.

Long history but better computing power has given opportunities for better modelling.

Championed by Australian statisticians:
*"*Researchers in Australia have promoted 'spatial' analysis of field experiments (Gilmour et al. 1997, JABES 2: 269-293) over the last 20 years because Australian fields are typically non uniform."

Easy to examine – autocorrelation.

Design and analysis of variety trials

Spatial Analysis

Simple model for autocorrelation $r^i = k_i$

ie if two plots are adjacent (k=1) the correlation is r

if two plots are next but one (k=2) the correlation is r^2

etc.

Include this requirement in the computer program which analyses the data instead of the incomplete block structure

Design and analysis of variety trials

Spatial Analysis

Simple model for autocorrelation $r^i = k_i$

Trial analysis so far, error for each plot in matrix form is :

$\sigma^2\mathbf{I}$

With correlation of errors becomes:

$\sigma^2\ \mathbf{R}$

where elements of **R** are composed by $r_i = k^i$

Design and analysis of variety trials

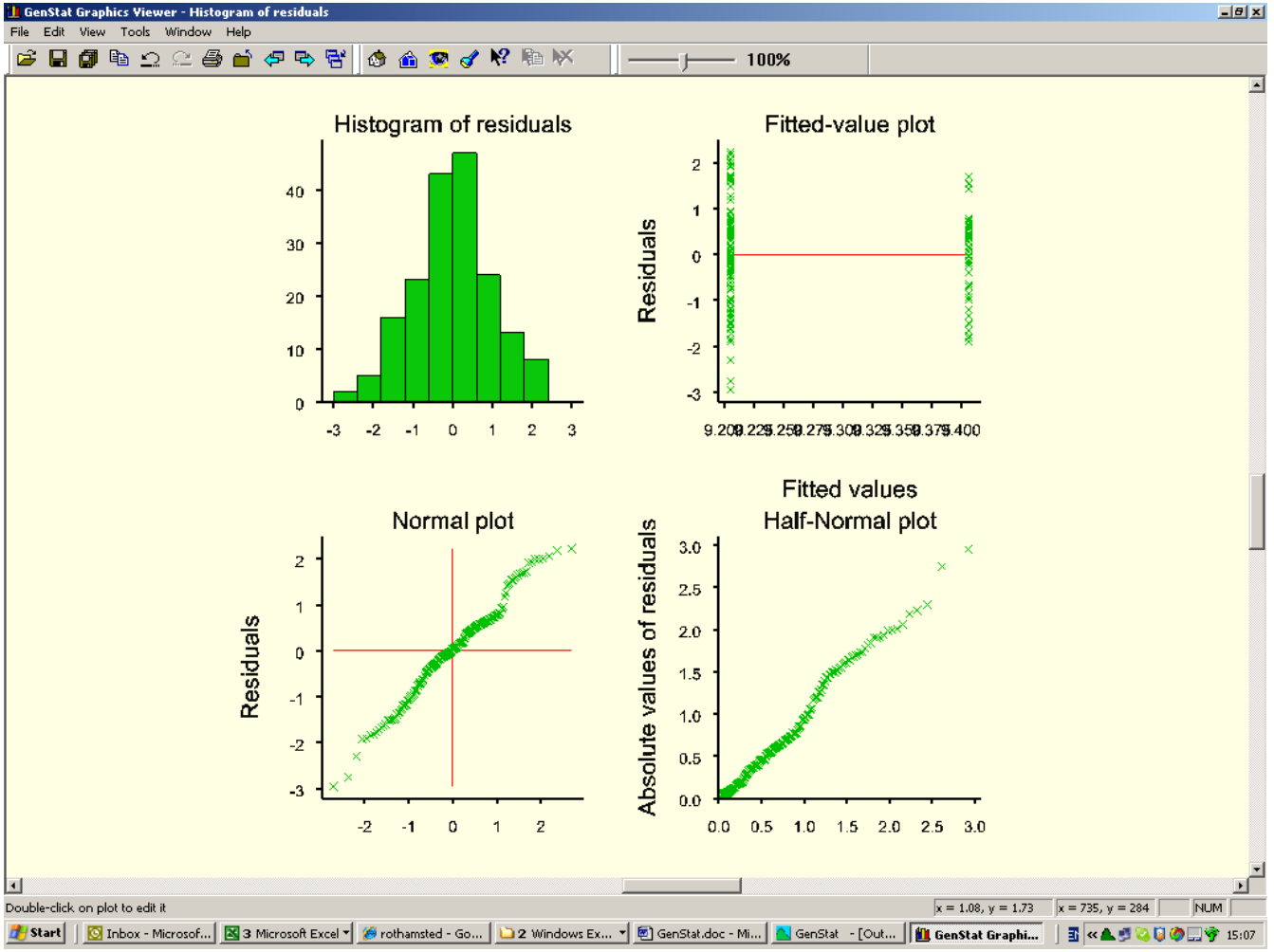
Spatial Analysis

Available directly within GenStat & ASREML

can have autocorrelation running in one direction or in two. More sophisticated options available too.

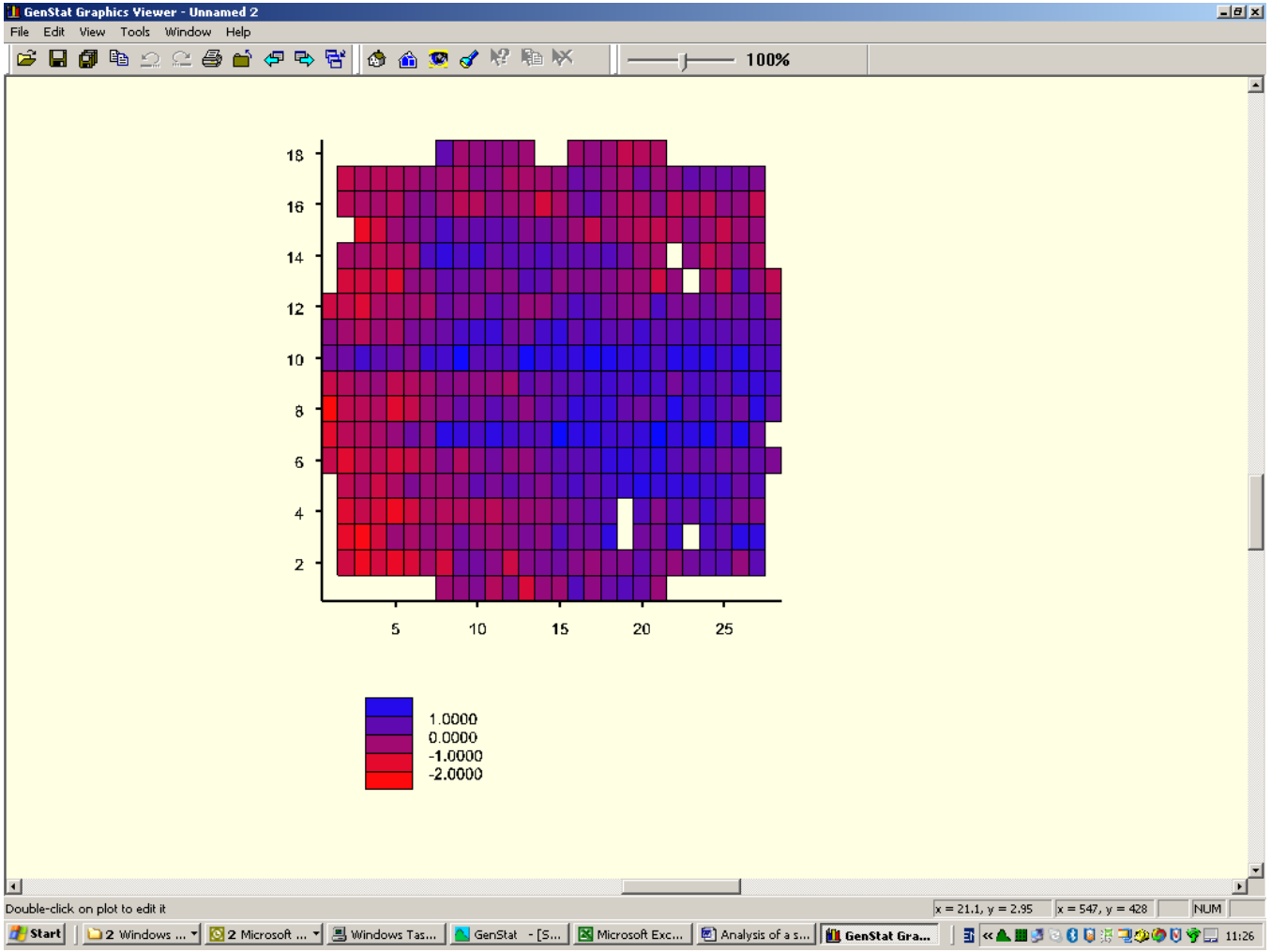
Design and analysis of variety trials

Inspecting residuals and outliers



Design and analysis of variety trials

Inspecting residuals and outliers



Tuesday pm

Quantitative Genetics

Books: Kearsey and Pooni.
Falconer and Mackay (no relation)
Felsenstein has a chapter on quantitative genetics.

Quantitative Genetics

Revision: means and variances

$$\text{mean} = m + \Sigma(p-q)a + 2\,\Sigma\,pqd$$

$$Vg = \Sigma 2pq[a+d(q-p)]^2 + \Sigma 4p^2q^2d^2$$

$$Va = \Sigma 2pq[a+d(q-p)]^2$$

$$Vd = \Sigma 4p^2q^2d^2$$

$$Vg = Va + Vd$$

$$Vp = Vg + Ve = Va + Vd + Ve$$

Quantitative Genetics

Means and variances: the F2

$p = q = ½$

mean = m defined with reference to the F2 only

$Va = ½ a^2$

$Vd = ¼ d^2$

Quantitative Genetics

Means and variances: no dominance

$$\text{mean} = m + \Sigma(p-q)a$$

$$Vg = Va = \Sigma 2pqa^2$$

Quantitative Genetics

Effect of inbreeding on the mean and variance

A_1A_1	A_1A_2	A_2A_2
$p^2+pq(1+F)$	$2pq(1-F)$	$q^2+pq(1+F)$

$$\text{mean} = m + (p-q)a + 2pqd(1-F)$$

d = 0:

$$Vg = Va(1+F)$$

F =1:

$$Vg = 2Va$$

Quantitative Genetics

Parent offspring regression

Covariance of offspring on parent = ½ Va

Variance of parental mean = ½ Vp

Regression of offspring on parent mean = Va/Vp

is called the heritability h^2_n

Heritability is the proportion of the total variation which is genetic.

No need to estimate by o/p regression – any experimental design which gives estimates of Va and Ve will do.

Quantitative Genetics

Revision: the breeders' equation $R = h^2S$.

$$R = ih^2\sigma_p$$

$$R = ih\sigma_g$$

i is the standardised mean of the selected group:

$$i = \Phi(x)/p$$

p is the proportion selected

$\Phi(x)$ is the p.d.f. $\Phi(x) = (2\pi)^{-1/2}\ e^{-½\ x\text{^}2}$

Unfortunately, it has a slight dependency on n, too.

look up in tables
simulate
substitute p = (k+ ½) / (n+ k/2n)
use the odds and sods spreadsheet

Quantitative Genetics

Selection within a generation (inbreds / DH / F1s / clones):

$$Vp_{(inbred)} = Vg_{(inbred)} + Ve/r$$

$$R = i.h^2_{inbred}.Vp_{(inbred)}$$

Optimise for i, no. of lines, and no. of replicates.

Selecting among inbreds: more complex cases:

Vp	plots
Vc	centres
Vy	years
Vv	varieties
Vcy	centres x varieties
Vcy	centres x years
Vyv	years x varieties
Vcyv	centres x years x varieties

The variance of a variety mean is then:
Vcv / c + Vyv / y + Vcyv / cy + Vp / ncy

Alternative: assume centres are nested within years
(ie sites are different each year)

s	sites each year
Vsv	varieties x sites within years

The variance of a variety mean is then:
Vyv / y + Vsv / sy + Vp / nsy

Quantitative Genetics

Estimating variances.

Published estimates

Other crops?

Use REML on existing data.

Other family types

	between	within	mean
individuals	Va + Vd	N/A	Σ [m+(p-q)a+2pqd]
clones	Va + Vd	0	Σ [m+(p-q)a+2pqd]
full sibs	½ Va + ¼ Vd	½ Va + ¾ Vd	Σ [m+(p-q)a+2pqd]
half sibs	¼ Va	¾ Va + Vd	Σ [m+(p-q)a+2pqd]
S1 progenies *	1½ Va	¼ Va	Σ [m+(p-q)a+pqd]
S2 progenies *	1¾ Va	1/8 Va	Σ [m+(p-q)a+½pqd]
fully inbred lines *	2Va	0	Σ [m+(p-q)a]
DH lines	2Va	0	Σ [m+(p-q)a]
F1s	Va + Vd	0	Σ [m+(p-q)a+2pqd]
4-way crosses.	½ Va + ¼ Vd	½ Va + ¾ Vd	Σ[m+(p-q)a+2pqd]

$Va = \Sigma\ 2pq[a+d(q-p)]^2$

$Vd = \Sigma\ 4p^2q^2d^2$

* Assume no dominance.

Quantitative Genetics

Estimating genetic variances and means – F2 derived populations

	mean			variance within		
	m	[a]	[d]	Va	Vd	Ve
P1	1	1	0	0	0	1
P2	1	-1	0	0	0	1
BC1	1	½	½	½	1	1
BC2	1	- ½	½	½	1	1
F1	1	0	1	0	0	1
F2	1	0	½	1	1	1
F3	1	0	¼	1½	¾	1
F∞	1	0	0	2	0	1

$Va = \Sigma a^2$

Quantitative Genetics

Estimating variances.

Simplest:

Use variation among plants in the F1 or the inbred parents as an estimate of Ve and get Vg by subtraction from Vp for F2.

More common:

Equate variances between and within families – full-sibs, half-sibs etc. with their genetic and environmental expectations and estimate from ANOVA or REML.

Quantitative Genetics

Model fitting:

1) Don’t overfit

2) Treat the complex models of others with caution (Type III error)

Quantitative Genetics

Heterosis:

Why does it occur?

overdominance or

dominance + dispersion

Heterotic subgroups:

Wahlund effect?

Any opinions anyone?

Does it matter?

Quantitative Genetics

Combining ability

GCA and SCA

Vgca = ½ Va

Vsca = Vd

Correlation between inbred performance and GCA

Why is it low?

Estimating GCA:

Use simple means across crosses if m and f lines are different.

If m and f lines are common, take care. See F&M

Quantitative Genetics

Heterosis Decreasing in Hybrids: Yield Test Inbreds

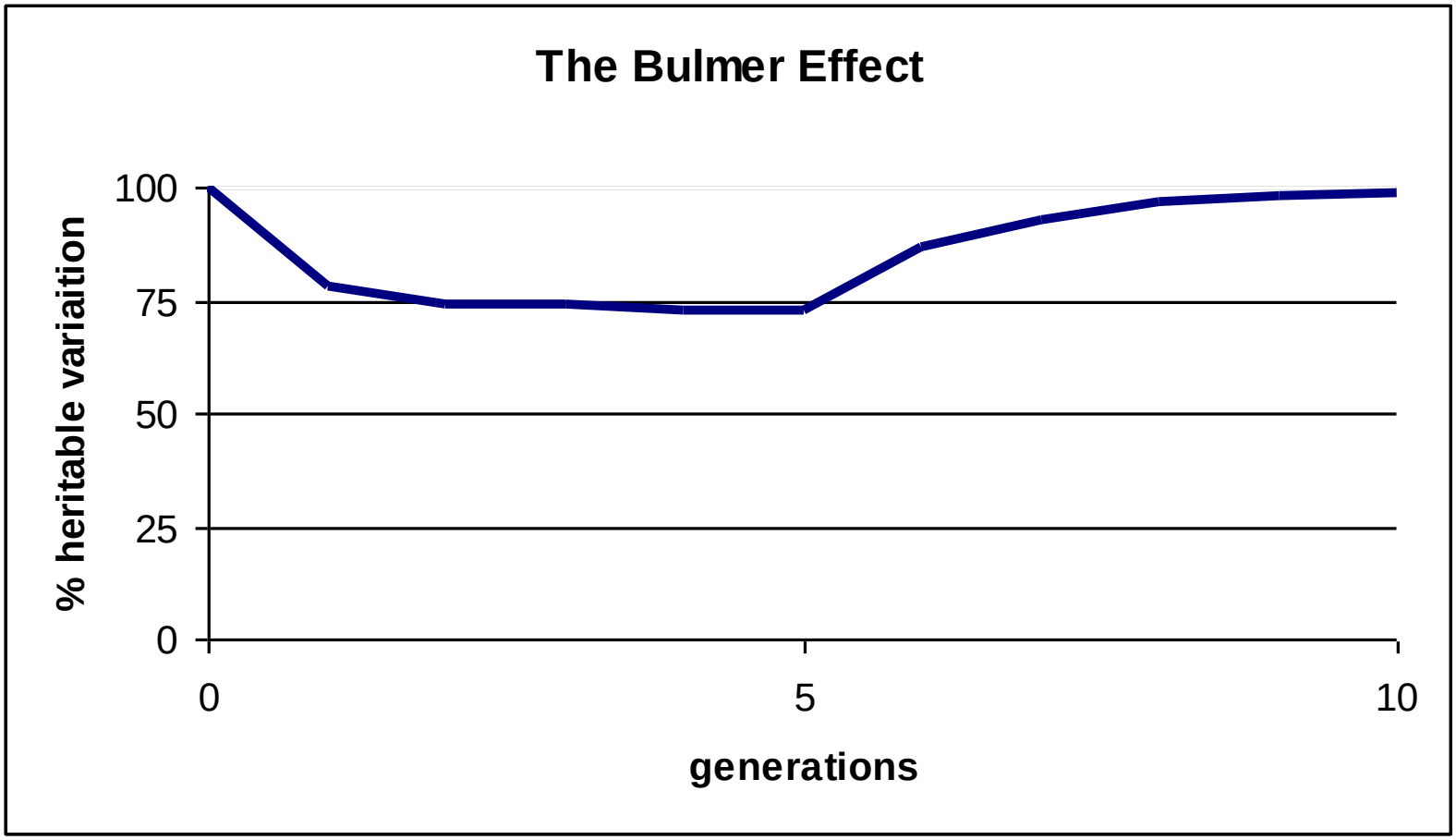
Troyer & Wellin

Crop Sci 49:1969-1976 (2009)

“Evaluation of more new inbreds will be conducive to increased genetic diversity that produces higher-yielding hybrids”.

Quantitative Genetics

Selection hides variation The Bulmer effect.



Quantitative Genetics

The Bulmer effect.

Within a generation (eg selecting the best inbred line)

Each stage of selection will:

increase the mean of the selected group.

reduce the variance.

Both are predictable:

$R = ih\sigma_g$

$Vp' = [1-i(i-z)]\ Vp$

$Vg' = [1-i(i-z)\ h^2\ ]Vg$

Can use this to design more efficient sequential testing programmes (eg the UK NL/RL list).

Quantitative Genetics

Where does the lost variance go?

genotype	AB	Ab	aB	ab	coeff of disequilibrium
phenotype	1.2	1.1	1.1	1	
freq	¼	¼	¼	¼	0
fitness~ pheno					
after seln:	1.2/4	1.1/4	1.1/4	1/4	
divide by total:	0.272	0.25	0.25	0.227	-0.00052

Selection generates negative disequilibrium between locus pairs

Quantitative Genetics

Bulmer effect over generations

Selection creates LD, but we know LD among loci decays at a rate of ½ per generation among unlinked loci.

$$Vg_{(t+1)} = \tfrac{1}{2} Vg_t[1-i(i-z)\, h^2\,] + \tfrac{1}{2}Vg_0$$

And at equilibrium:

$$Vg_{(t+1)} = Vg_{(t)}$$

NB – the loss of variation can be severe in crops.

Quantitative Genetics

Applications of the Bulmer effect

Pedigree breeding

	total	contribution from				
		F2	F3	F4	F5	F>5
F2	Va	Va				
F3	3Va/2	Va	Va/2			
F4	7Va/4	Va	Va/2	Va/4		
F5	15Va/8	Va	Va/2	Va/4		
F∞	2Va	Va	Va/2	Va/4	Va/8	Va/8

Use Bulmer to predict the reduction in variance carried forward to the next stage of selection then add-in the segregation variance.

Quantitative Genetics

Applications of the Bulmer effect

Insert a cycle of mating without selection in recurrent selection schemes?

This has been proposed / applied.

Depends how much extra time is required and linkage.

Inbreeding crops:

start with 4-way crosses rather than F1s?

How ubiquitous is the correlation between yield and quality?

Does the Bulmer effect provide a valid explanation?

Should it be used as a base model against which physiological explanations (= pleiotrophy) be judged?

Why hasn’t this been reviewed / studied?

Quantitative Genetics

Selection limits and changes in allele frequency at a single locus

$$R_{lim} = 2Ne\ iVa/\sigma_p$$

$$R_{ss} = 2Ne\ iVm/\sigma_p$$

Maximum long term response comes from 50% selection.

but in experiments, selection limits are not usually due to lack of Vg.

Mutation is an important source of new variation

In crop domestication and evolution, what proportion of variants were present in wild ancestors and what proportion are a result of new mutations over the last 10,000 years?

Quantitative Genetics

Multiple traits

Genetic and environmental correlations:

They can be different.

They can be zero even though there are strong relationships – constancy of gene action.

Estimation

Best is to use REML

Can work on sums of traits if necessary.

Quantitative Genetics

Correlated response to selection

Response on trait y to selection on x:

$$R = ih_x\sigma_{(gx)}\ cov_{g\,/}\ \sigma^2_{(gx)}$$

or

$$R = ir_g h_x h_y \sigma_{(py)}$$

Quantitative Genetics

GxE

$$y_{ijk} = m + g_i + s_j + ge_{ij} + e_{ijk}$$

This model is commonly fitted in two stages: first the means at each site are estimated, then an across sites analysis is carried out – using ANOVA or, increasingly, REML to account for missing varieties.

Easily extended to include sites and years and their interactions.

There is an enthusiasm among statisticians for a single stage analysis in which plot data at each site is included in the analysis. There are reports that this is more accurate, but it can be impractical – too much data to analyse in one step.

Aside from estimating the gi, variance components for each term can be used to optimise the allocation of resources in the breeding programme.

Quantitative Genetics

GxE

Can also analyse data across sites as if they are correlated traits:

Treat performance at each site as different traits. The correlation in performance across sites is all genetic so estimation of parameters is easy.

Not much used in plant breeding. Much more common approach in animal breeding?

Quantitative Genetics

Finlay & Wilkinson

Not much used in practice now.

Regress variety performance at each site on the site mean:

- Doesn’t require measurement of independent variables – rainfall and so on.
- Doesn’t need any physiology.
- b> 1 for varieties sensitive / responsive to the environment
- b< 1 for varieties stable or non-responsive.

Quantitative Genetics

AMMI

$$y_{ijk} = m + g_i + s_j + \sum_n^N u_n w_n v_n' + r_{ij} + e_{ijk}$$

$\sum_n^N u_n w_n v_n'$ this term represents the spectral value decomposition of the matrix of varieties x sites residuals (the GxE effects).

In practice, use only the first one or two largest values of w (the latent roots).

Display results in a biplot.

Patterns may have an obvious biological interpretation.

Varieties and/or environments may cluster in an informative manner,

A couple of references on AMMI

Biplot Analysis of Genotype × Environment Interaction: Proceed with Caution

Yang R-C, Crossa J, Cornelius PL, Burgueno J, 2009 Crop Sci 49: 1564-76

Best Linear Unbiased Prediction (BLUP) for regional yield trials: a comparison to additive main effects and multiplicative interaction (AMMI) analysis

Piepho H-P (1994) Theor Appl Genet **89**:647-654

G x E: a personal view

Analysis of GxE analysis has a role in crop genetics but much less in breeding:

If GxE effects are large, run separate selection programmes for each target environment.

Future:

methods need to be more predictive and less descriptive:

modelling at the QTL level : QGENE

better long range weather forecasts

G x E analysis in very incomplete data.

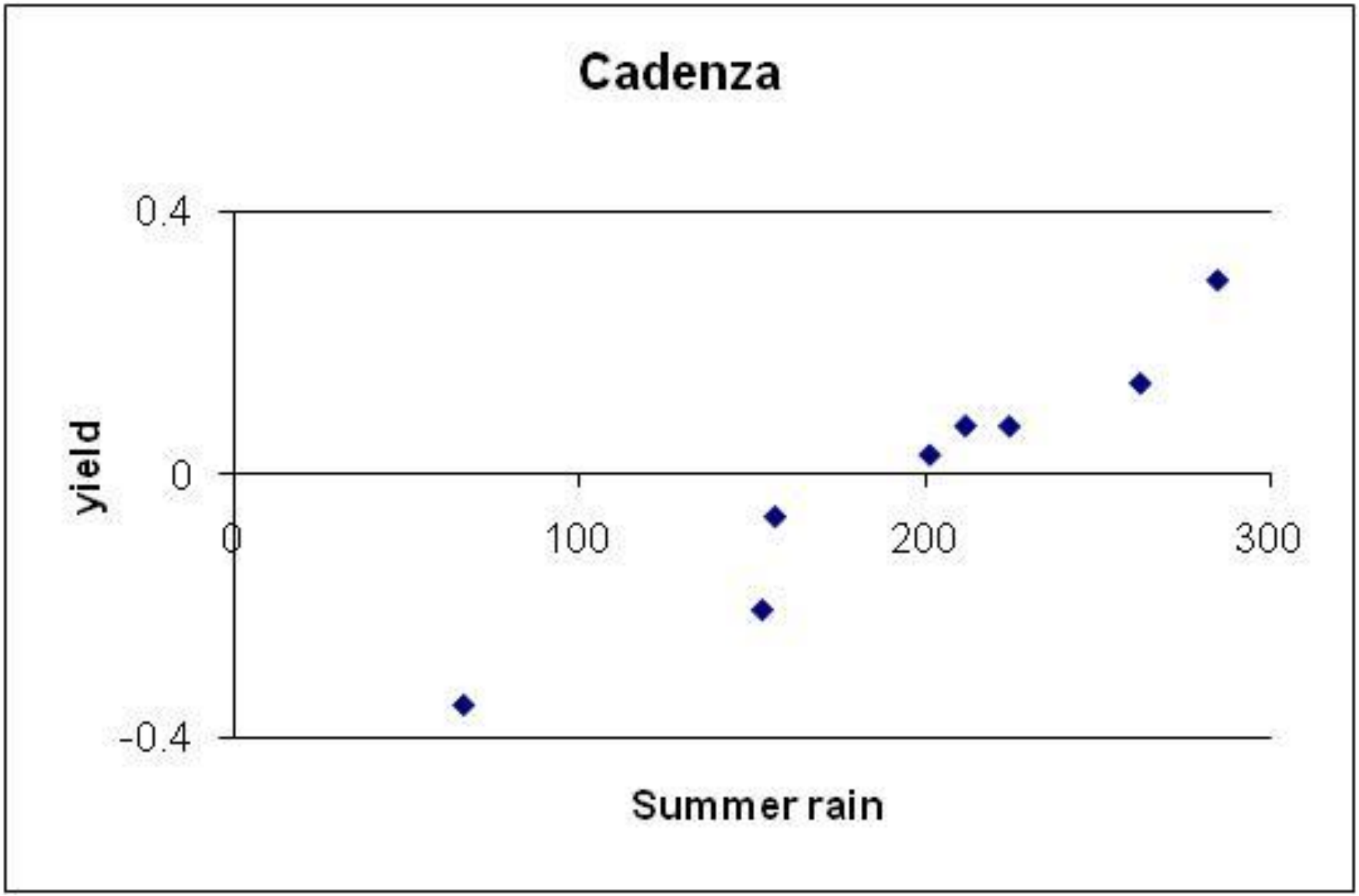
Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. *Mackay et al (2010) Theor Appl Genet* ***122****:225-238*

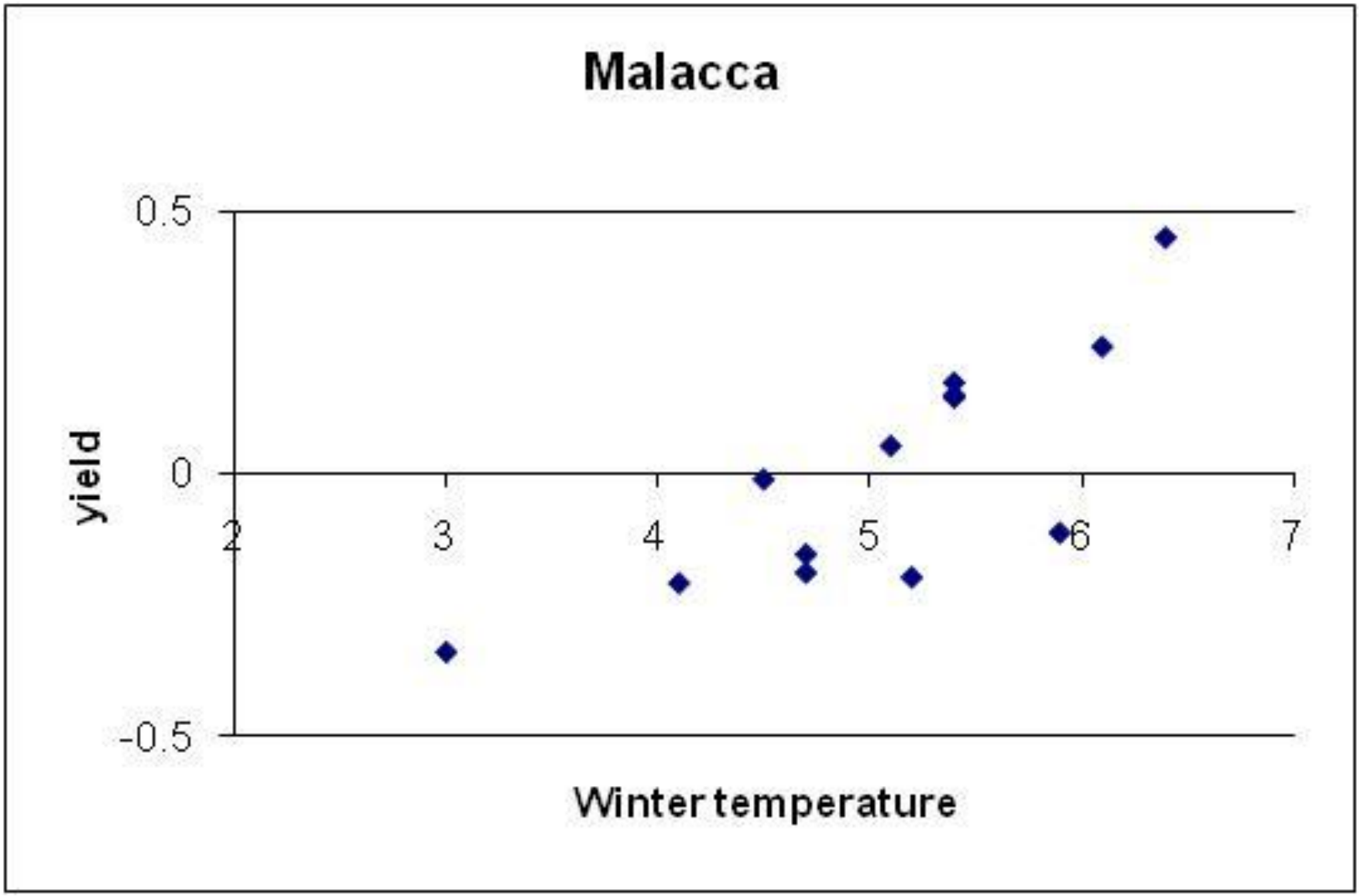
165 varieties x 26 years. ~80% missing data

Detect important climatic drivers through FDR

Table 4 Numbers of winter wheat varieties, out of 165 total with false discovery rate <0.5 for eight climatic variables

	Autumn	Winter	Spring	Summer
Rainfall	0	0	1	26
Temperature	0	27	8	2


Cadenza
0.4
0
-0.4
yield
0
100
200
300
Summer rain


Malacca
0.5
0
-0.5
yield
2
3
4
5
6
7
Winter temperature

Variety x disease interactions: a cautionary tale.