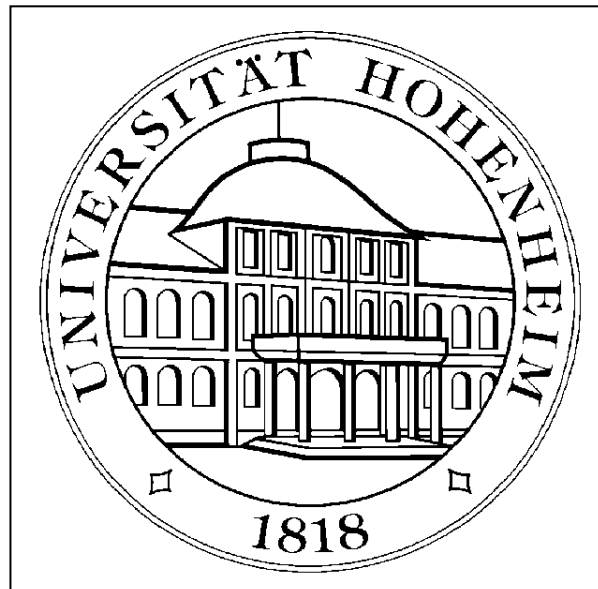


Dissecting genetic and non-genetic sources of long-term yield trend in German official cultivar trials

Hans-Peter Piepho

Biostatistics Unit
Universität Hohenheim
Germany



This is joint work with:

Friedrich Laidig, Thomas Drobek, Uwe Meyer

Bundessortenamt, Germany

⇒ Authority responsible for official variety registration trials in Germany

⇒ Bundessortenamt are currently analysing long-term data from their trials
for all major crops

Questions

- Contribution of plant breeding to yield increases (genetic trend) ?
- Contribution of improved agronomic practices to yield increases ?

Starting point:

Mackay IJ, Horwell A, Garner J, White J, McKee J, Philpott H (2011)
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. *Theoretical and Applied Genetics* **122**, 225-238.

"We find that since 1982, for the cereal crops and oil seed rape, at least **88%**
of the improvement in yield is attributable to genetic improvement, with little
evidence that changes in agronomy have improved yields."

Table of contents

1. A basic mixed model
2. Incorporating regression terms
3. Application
4. Discussion
5. Summary

1. A basic mixed model

Data

- Registration trials in Germany
- Winter wheat and spring barley
- Yield (dt ha⁻¹, where 1 dt = 10⁻¹ t) and mildew susceptibility (1-9 score)
- 1979 - 2012
- Varieties at least three years under test
- Two intensities:
 - Intensity 1 (no crop protection)
 - Intensity 2 (with crop protection)
- Split-plot design:
 - Intensity on main plots (RCBD)
 - Variety on sub-plots (completely randomized)

⇒ Variety means per trial and intensity, no weighting

1. A basic mixed model

Table 1: Basic information on the yield trial data for winter wheat and spring barley.

Crop	Trait	No. of observations	No. of genotypes	Average age of genotype	No. of locations	Percentage of GxYxL combinations
Winter wheat	Grain yield (dt/ha)	22820	286	3.51	115	2.13
	Mildew susceptibility (1-9)	17174	286	3.51	108	1.70
Spring barley	Grain yield (dt/ha)	15871	176	3.78	113	2.37
	Mildew susceptibility (1-9)	10533	176	3.78	108	1.68

1. A basic mixed model

Basic model for long-term MET data

$$y_{ijk} = \mu + G_i + L_j + Y_k + (LY)_{jk} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk} \quad (1)$$

y_{ijk} = mean yield of the i -th genotype in the j -th location and k -th year

μ = overall mean

G_i = main effect of the i -th genotype

L_j = main effect of the j -th location

Y_k = main effect of the k -th year

$(LY)_{jk}$ = jk -th location \times year interaction

$(GL)_{ij}$ = ij -th genotype \times location interaction

$(GY)_{ik}$ = ik -th genotype \times year interaction

$(GLY)_{ijk}$ = residual comprising both genotype \times location \times year interaction as well as the error of a mean

1. A basic mixed model

Same model with error and interaction separated

$$y_{ijk} = \mu + G_i + L_j + Y_k + (LY)_{jk} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk} + e_{ijk} \quad (1)$$

Here:

$(GLY)_{ijk}$ = pure genotype \times location \times year interaction

e_{ijk} = error of a mean y_{ijk}

1. A basic mixed model

Weighted two-stage analysis

(1) Compute genotype means from individual trials y_{ijk} and associated (approximate) variances $\text{var}(y_{ijk})$

(2) Fit model across years and sites, using $\text{var}(y_{ijk}) = \text{var}(e_{ijk})$ from stage (1)

1. A basic mixed model

Randomized complete block design (RCB):

$$\text{var}(y_{ijk}) = \frac{\sigma_{jk}^2}{n_{jk}}$$

Incomplete block designs:

$$\text{var}(y_{1jk}, y_{2jk}, \dots) = R_{jk}$$

Use diagonal elements of R_{jk}^{-1} as weights w_{ijk}

Approximate $\text{var}(y_{ijk}) = w_{ijk}^{-1}$ (Smith et al. 2001)

1. A basic mixed model

Mackay et al. (2011)

- Take G_i and Y_k as fixed (can't take random because of time trend)
- All other effects random (i.i.d. normal with constant variance)
- Adjusted means for G_i assess **genetic trend**
 - ⇒ Plotted against year in which variety entered trial
- Adjusted means for Y_k assess **non-genetic trend**
 - ⇒ Plotted against calendar year
- Estimate trend by linear regression based on adjusted means for G_i and Y_k
- Look at one intensity at a time

1. A basic mixed model

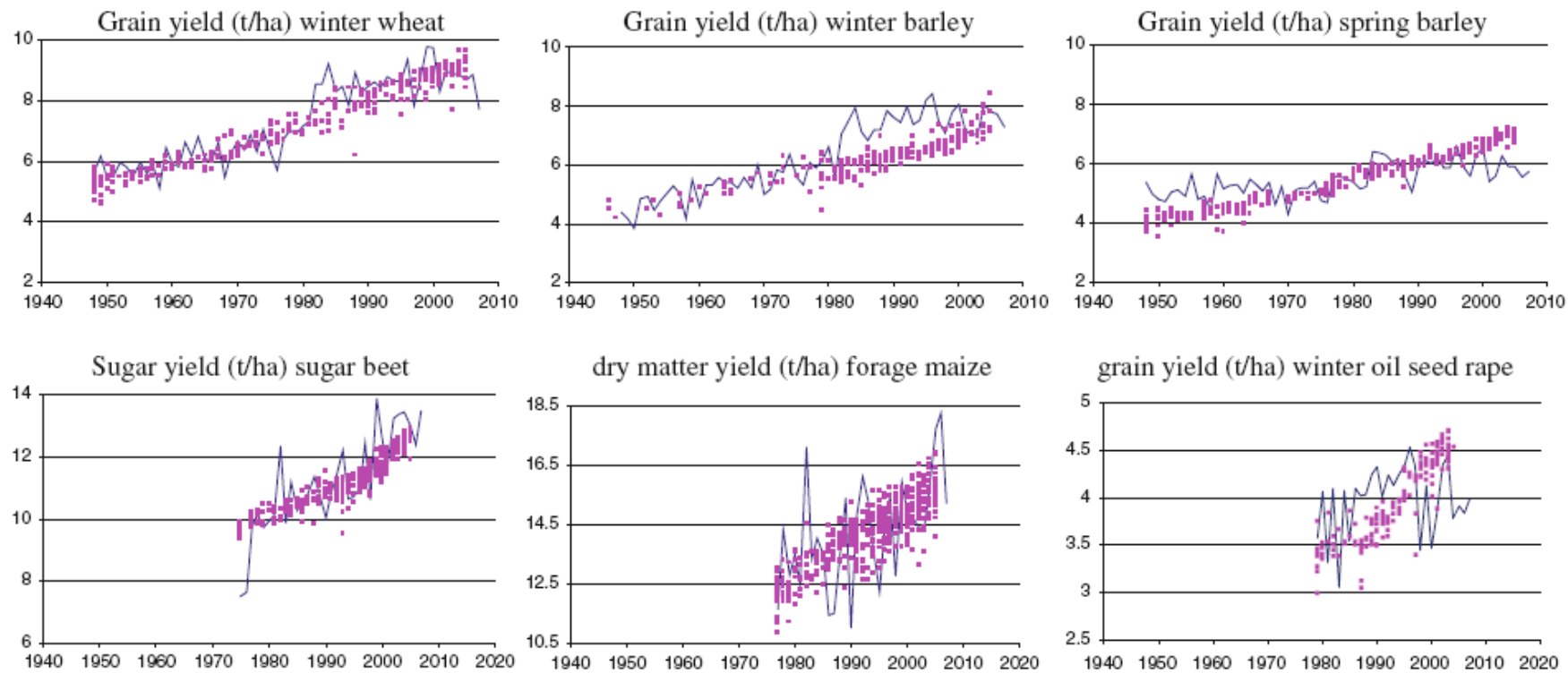


Fig. 1 Trends in variety and year effect for yield (t/ha) from 1948 to 2007. Ordinate and abscissa are on the same scale for all crops except oil seed rape. Variety and year means were estimated as described in

“[Materials and methods](#)” section. Variety effects (*squares*) are plotted against the year in which the variety first entered the trial. Year means are plotted as a *line*

(Mackay et al., 2011)

— year means
■ variety means

1. A basic mixed model

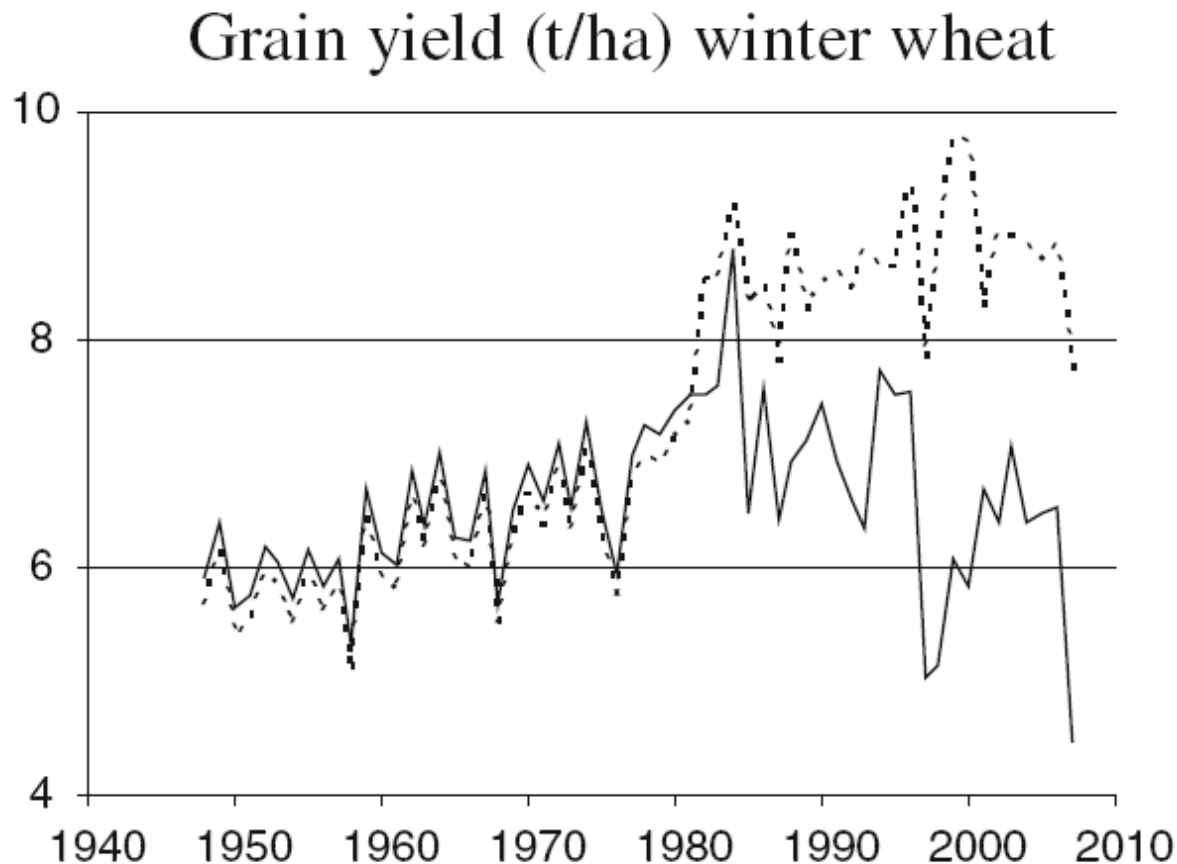


Fig. 3: Trends in year effects in UK variety trials (Mackay et al. 2011)

— — — Dotted line: analysis with treated trials only (from 1982 onwards).

———— Solid line: analysis with untreated trials only (from 1982 onwards).

2. Incorporating regression terms

Genetic trend

$$G_i = \beta r_i + H_i \quad (2)$$

β = fixed regression coefficient for genetic trend

r_i = year of first trial for i -th variety

$$H_i \sim N(0, \sigma_H^2)$$

Non-genetic trend

$$Y_k = \gamma t_k + Z_k \quad (3)$$

γ = fixed regression coefficient for agronomic trend

t_k = calendar year

$$Z_k \sim N(0, \sigma_Z^2)$$

2. Incorporating regression terms

The fixed part of the model so far

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k \quad (4)$$

η_{ik} = expected response of the i -th genotype in the k -th year

2. Incorporating regression terms

Effect of disease resistance breakdown

- It has been noted that disease resistance decays with time for a number of years
- This is expected to have an effect on time trends for yield etc.

Extension of the fixed part of the model

- Use **age at testing** as another covariate in the model

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k + \delta a_{ik} \quad (5)$$

a_{ik} = age of i -th genotype in k -th year of testing

δ = fixed regression coefficient for the age covariate a_{ik}

2. Incorporating regression terms

Effect of disease resistance breakdown cont'd

Age at testing for the i -th genotype in the k -th year:

$$a_{ik} = t_k - r_i \tag{6}$$

Extended regression model:

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k + \delta(t_k - r_i) \tag{7}$$

2. Incorporating regression terms

Problem: Model over-parameterized (multi-collinearity)

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k + \delta(t_k - r_i) \quad (7)$$

Can be rearranged:

$$\eta_{ik} = \mu + \tilde{\beta} r_i + \tilde{\gamma} t_k, \text{ where} \quad (8)$$

$$\tilde{\beta} = \beta - \delta \text{ and} \quad (9)$$

$$\tilde{\gamma} = \gamma + \delta \quad (10)$$

⇒ Regression on r_i and t_k may be biased due to disease resistance breakdown!

⇒ Can't separate out trend due to resistance breakdown

2. Incorporating regression terms

Apparent trends

Apparent genetic trend: $\tilde{\beta} = \beta - \delta$

Apparent non-genetic trend: $\tilde{\gamma} = \gamma + \delta$

Yield: $\delta < 0$

$\Rightarrow \tilde{\beta} > \beta \Rightarrow$ genetic trend over-estimated

$\Rightarrow \tilde{\gamma} < \gamma \Rightarrow$ non-genetic trend under-estimated

2. Incorporating regression terms

Estimation of an effect of resistance breakdown (δ)

\Rightarrow Compare the two intensities

$$\text{Intensity 1: } \eta_{ik1} = \mu_1 + \tilde{\beta}_1 r_i + \tilde{\gamma}_1 t_k \quad (11)$$

$$\text{Intensity 2: } \eta_{ik2} = \mu_2 + \tilde{\beta}_2 r_i + \tilde{\gamma}_2 t_k \quad (12)$$

$$\tilde{\beta}_m = \beta - \delta_m \quad (m = 1, 2) \text{ and} \quad (13)$$

$$\tilde{\gamma}_m = \gamma + \delta_m \quad (m = 1, 2) \quad (14)$$

2. Incorporating regression terms

Difference of response for both intensities

$$\begin{aligned}\eta_{ik2} - \eta_{ik1} &= (\mu_2 - \mu_1) + \tilde{\beta}_2 r_i + \tilde{\gamma}_2 t_k - (\tilde{\beta}_1 r_i + \tilde{\gamma}_1 t_k) \\ &= (\mu_2 - \mu_1) + \bar{\beta} r_i + \bar{\gamma} t_k\end{aligned}\tag{15}$$

where

$$\bar{\beta} = \tilde{\beta}_2 - \tilde{\beta}_1 \text{ and}\tag{16}$$

$$\bar{\gamma} = \tilde{\gamma}_2 - \tilde{\gamma}_1\tag{17}$$

When relations (13) and (14) hold: $\bar{\beta} = \delta_1 - \delta_2$ and $\bar{\gamma} = \delta_2 - \delta_1$

Lack-of-fit test: $H_0 : \bar{\beta} = -\bar{\gamma}$

2. Incorporating regression terms

Difference of response for both intensities (cont'd)

$$\begin{aligned}\eta_{ik2} - \eta_{ik1} &= (\mu_2 - \mu_1) + (\delta_1 - \delta_2)r_i - (\delta_1 - \delta_2)t_k \\ &= (\mu_2 - \mu_1) - (\delta_1 - \delta_2)a_{ik}\end{aligned}\tag{18}$$

If we assume that $\delta_2 = 0$, then regression on a_{ik} estimates $-\delta_1$

But: $H_0: \delta_2 = 0$ is an untestable assumption!

Even so: If we find that $(\delta_1 - \delta_2) \neq 0$, we may conclude that there is indeed an ageing effect, although we cannot quantify that effect separately for the two intensities.

3. Application

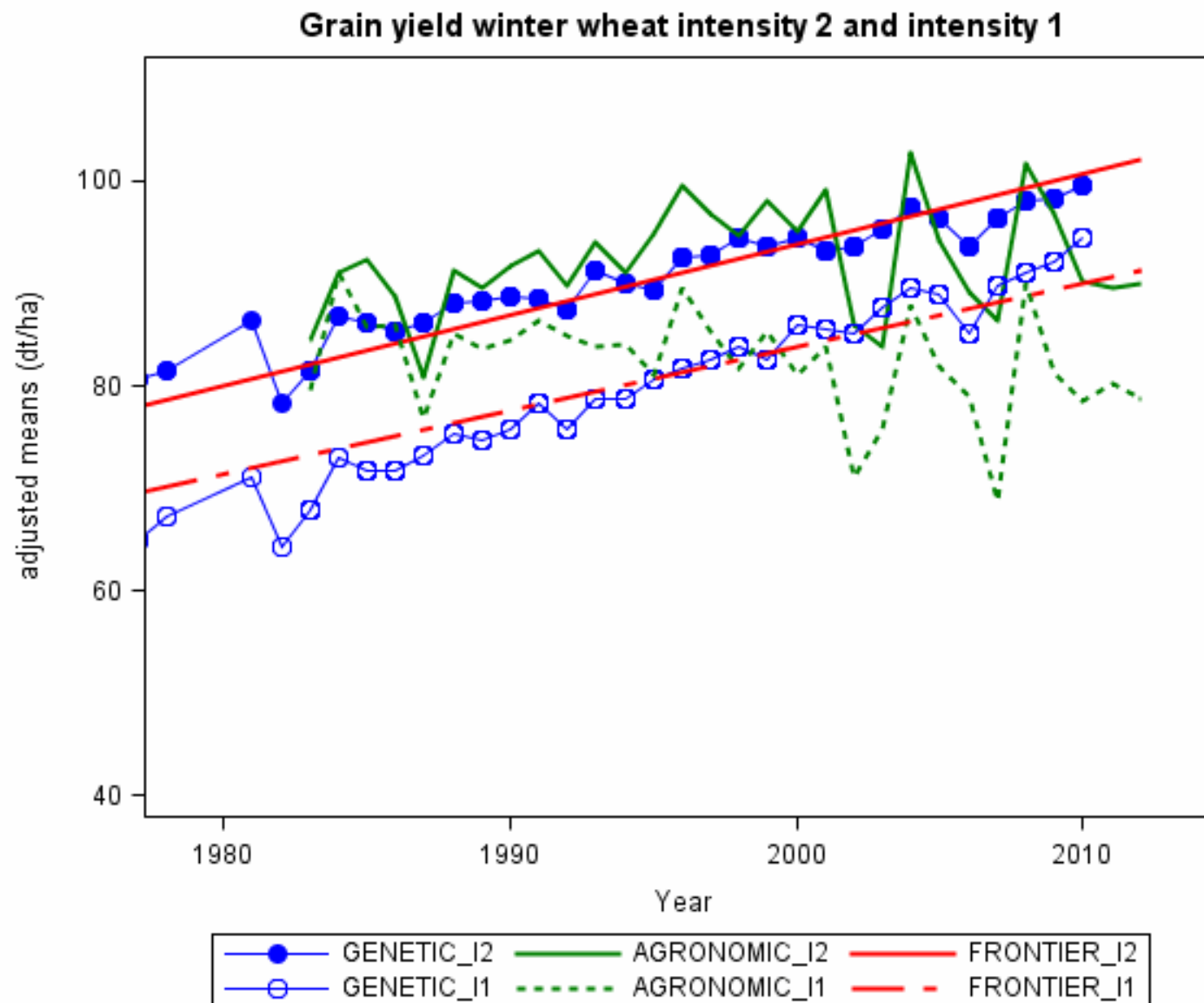


Figure 2:

GENETIC:
variety group means

AGRONOMIC:
year means

FRONTIER:
frontier line

I1: Intensity 1
I2: Intensity 2

3. Application

Table 4: Estimates of regression coefficients for winter wheat

Trait	Int.	Slope estimates for regression on					
		Yr of first trial (r_i)		Calendar year (t_k)		Age at test (a_{ik})	
		Estimate (dt/ha)	s.e.	Estimate (dt/ha)	s.e.	Estimate (dt/ha)	s.e.
Yield	1	+0.82	0.04	-0.20	0.11	---	---
	2	+0.53	0.04	+0.16	0.12	---	---
	2-1	-0.29 [§] ($\delta_1 - \delta_2$)	0.02	+0.36 [§] $-(\delta_1 - \delta_2)$	0.07	0.29 ($\delta_1 - \delta_2$)	0.02
Mildew	1	-0.095	0.006	+0.052	0.0092	---	---
	2	-0.042	0.003	-0.002	0.0048	---	---
	2-1	+0.053 [§] ($\delta_1 - \delta_2$)	0.005	-0.055 [§] $-(\delta_1 - \delta_2)$	0.008	-0.054 ($\delta_1 - \delta_2$)	0.004

§: Test of $H_0 = \bar{\beta} = -\bar{\gamma}$ not significant at 5% level (Table 3).

3. Application

Comparison of old and new varieties in joint trial

Ahlemeyer and Friedt (2010):

Genetic trend during 1996-2007: $\beta = 0.34 \text{ dt ha}^{-1} \text{ yr}^{-1}$ (Intensity 2)

Our long-term analysis (Intensity 2):

Apparent genetic trend: $\tilde{\beta}_2 = \beta - \delta_2 = 0.53 \text{ dt ha}^{-1} \text{ yr}^{-1}$

From regression of differences on a_{ik} : $(\delta_1 - \delta_2) = -0.29 \text{ dt ha}^{-1} \text{ yr}^{-1}$

3. Application

Comparison of old and new varieties in joint trial

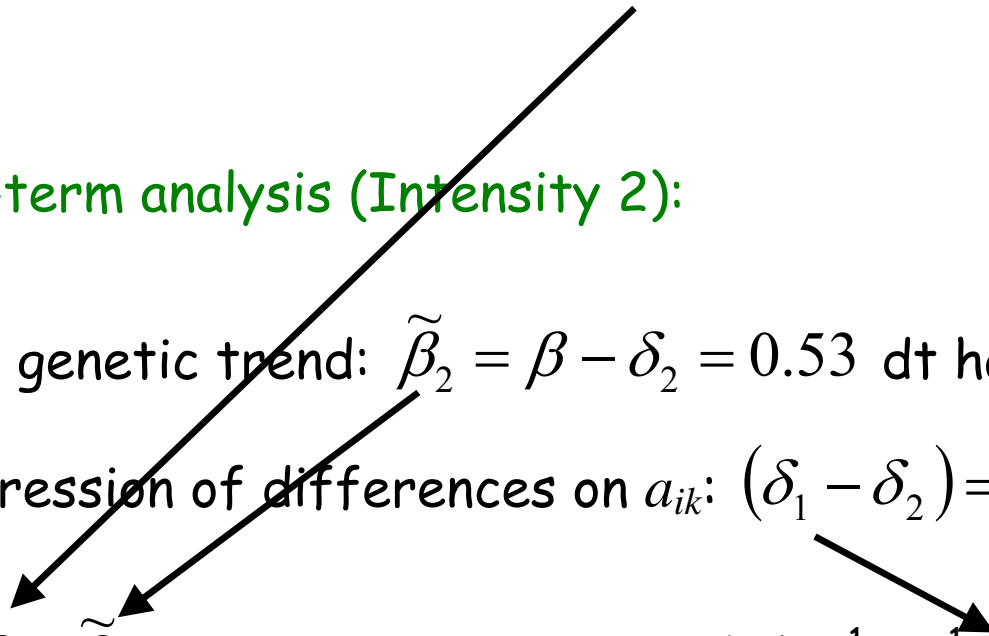
Ahlemeyer and Friedt (2010):

Genetic trend during 1996-2007: $\beta = 0.34 \text{ dt ha}^{-1} \text{ yr}^{-1}$ (Intensity 2)

Our long-term analysis (Intensity 2):

Apparent genetic trend: $\tilde{\beta}_2 = \beta - \delta_2 = 0.53 \text{ dt ha}^{-1} \text{ yr}^{-1}$

From regression of differences on a_{ik} : $(\delta_1 - \delta_2) = -0.29 \text{ dt ha}^{-1} \text{ yr}^{-1}$



The diagram consists of three arrows. The first arrow starts from the β term in the first equation and points to the β term in the final equation. The second arrow starts from the $\tilde{\beta}_2$ term in the second equation and points to the $\tilde{\beta}_2$ term in the final equation. The third arrow starts from the $(\delta_1 - \delta_2)$ term in the third equation and points to the δ_1 term in the final equation.

$$\Rightarrow \delta_2 = \beta - \tilde{\beta}_2 = 0.34 - 0.53 = -0.19 \text{ dt ha}^{-1} \text{ yr}^{-1} \Rightarrow \delta_1 = -0.48 \text{ dt ha}^{-1} \text{ yr}^{-1}$$

3. Application

Non-genetic trend

$$\tilde{\gamma}_m = \gamma + \delta_m \quad (m = 1, 2)$$

Estimator 1:

$$\tilde{\gamma}_1 = \gamma + \delta_1 \Rightarrow \gamma = \tilde{\gamma}_1 - \delta_1 = -0.20 + 0.48 = 0.28$$

Estimator 2:

$$\tilde{\gamma}_2 = \gamma + \delta_2 \Rightarrow \gamma = \tilde{\gamma}_2 - \delta_2 = 0.16 + 0.19 = 0.35$$

3. Application

Summary for wheat yield (trends in $\text{dt ha}^{-1} \text{yr}^{-1}$)

(1) From long-term data:

Apparent genetic trend $\tilde{\beta}_1 = 0.82, \tilde{\beta}_2 = 0.53$

Apparent non-genetic trend $\tilde{\gamma}_1 = -0.20, \tilde{\gamma}_2 = 0.16$

(2) From single-year MET with old and new varieties (Ahlemeyer & Friedt):

True genetic trend $\beta = 0.34$

(3) Combining the information from both sources

True non-genetic trend $\gamma = 0.28 \quad / \quad \gamma = 0.35$

Effect of resistance decay $\delta_1 = -0.48, \delta_2 = -0.19$

4. Conclusion

- Basic 3-way linear mixed model can be used to dissect genetic and non-genetic sources of long-term trend
- Incorporating regression terms simplifies trend analysis
- Effects of disease breakdown cause an upward bias for genetic trend and downward bias for non-genetic trend estimates
- Can partly separate effects of disease breakdown by comparing two agronomic intensities (with and without crop protection), but can't make further progress without introducing untestable assumptions
- Full separation is possible by testing old and new varieties concomitantly
- Based on results for wheat, genetic and non-genetic trends seem to split fairly equally

5. References

- Ahlemeyer J, Friedt W (2010) Progress in winter wheat yield in Germany - What's the share of the genetic gain? pp. 19-23. In: 61. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs.
- Mackay IJ, Horwell A, Garner J, White J, McKee J, Philpott H (2011) 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. *Theoretical and Applied Genetics* 122:225-238.
- Laidig F, Piepho HP, Drobek T, Meyer U (2014) Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. *Theoretical and Applied Genetics* 127, 2599-2617.
- Piepho HP, Laidig F, Drobek T, Meyer U (2014) Dissecting genetic and non-genetic sources of long-term yield trend in German official cultivar trials. *Theoretical and Applied Genetics* 127:1009-1018.

Thanks!

Graphical displays

Overall trend: **frontier line**

$$\eta_k = \mu + (\beta + \gamma)t_k \quad (22)$$

⇒ realized trend if best new variety is planted each year

Genotype-specific regression lines:

$$\eta_{ik} = \mu + \beta r_i + \gamma t_k \quad (4)$$

$$\eta_{ik} = \mu + \beta r_i + H_i + \gamma t_k \quad (23)$$

(fixing age variable r_i at value of i -th genotype)

3. Application

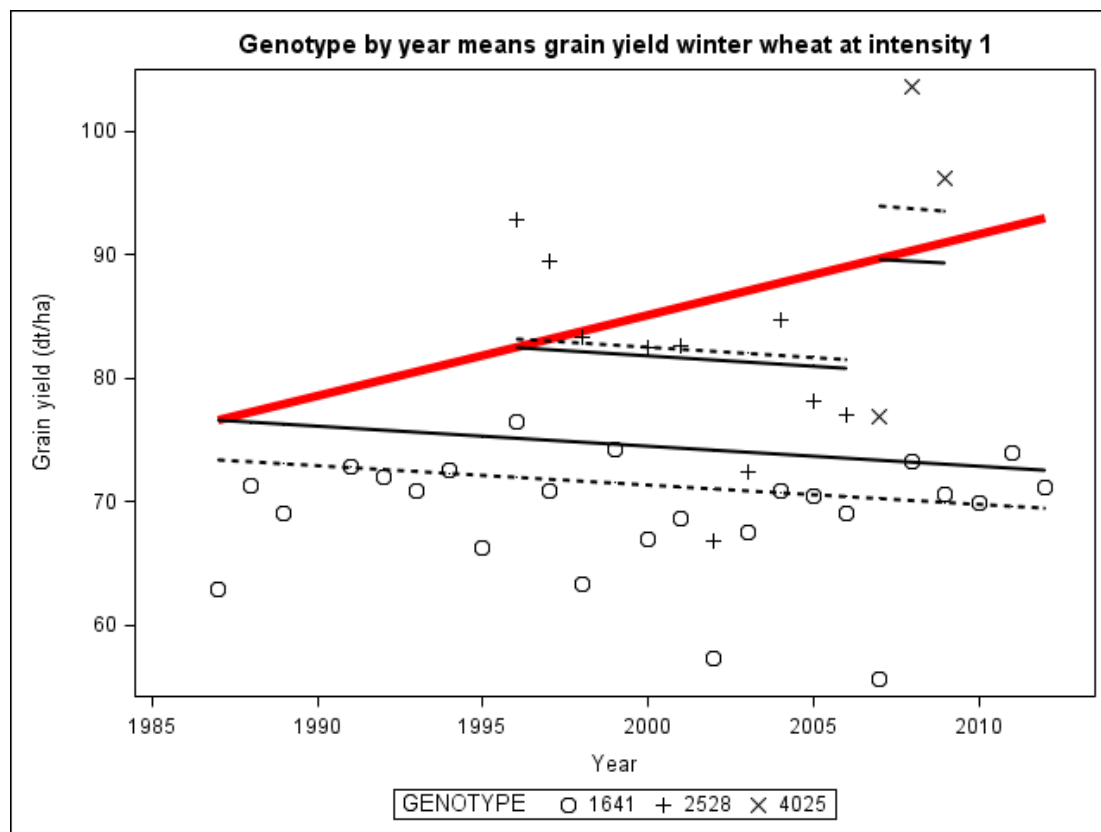


Figure 1: Plot for winter wheat grain yield (dt/ha).

- Solid red line: eq. (22), **frontier line:** $\eta_k = \mu + (\beta + \gamma)t_k$
- Solid black line (shown for 3 genotypes): $\eta_{ik} = \mu + \beta r_i + \gamma t_k$
- - - Dashed black line (shown for 3 genotypes): $\eta_{ik} = \mu + \beta r_i + H_i + \gamma t_k$

3. Application

Two predictions from extended model

$$\eta_{ik1} = \mu_1 + \tilde{\beta}_1 r_i + \tilde{\gamma}_1 t_k \quad (11)$$

$$\eta_{ik2} = \mu_2 + \tilde{\beta}_2 r_i + \tilde{\gamma}_2 t_k \quad (12)$$

(1) Genetic and non-genetic trends should differ between intensities

$$\begin{aligned} \tilde{\beta}_m &= \beta - \delta_m \quad (m = 1, 2) \text{ and} \\ \tilde{\gamma}_m &= \gamma + \delta_m \quad (m = 1, 2) \end{aligned}$$

(2) Frontier lines should have about equal slopes for both intensities

$$\tilde{\beta}_1 + \tilde{\gamma}_1 = \tilde{\beta}_2 + \tilde{\gamma}_2 = \beta + \gamma$$

\Rightarrow unbiased from disease resistance breakdown

3. Application

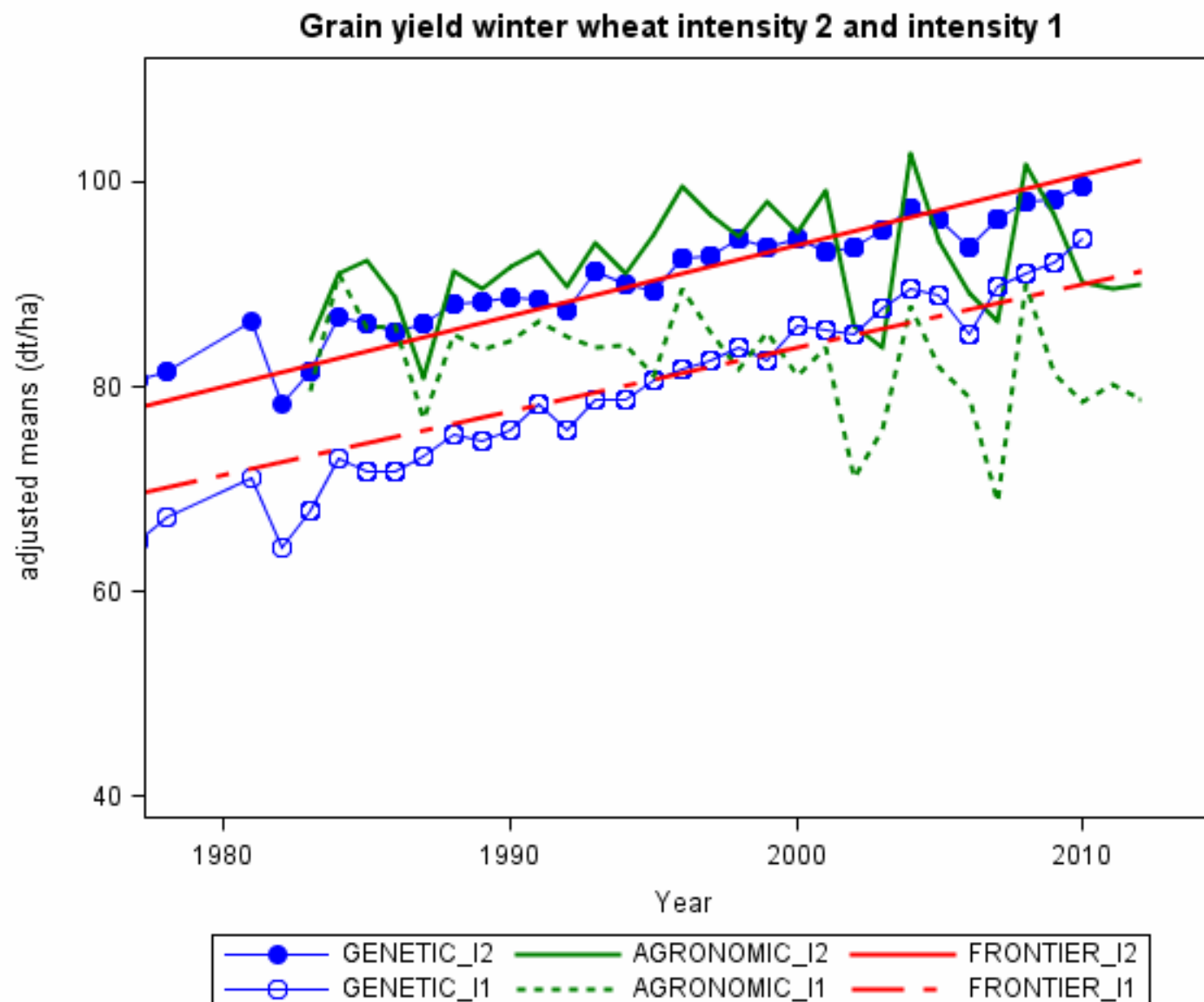


Figure 2:

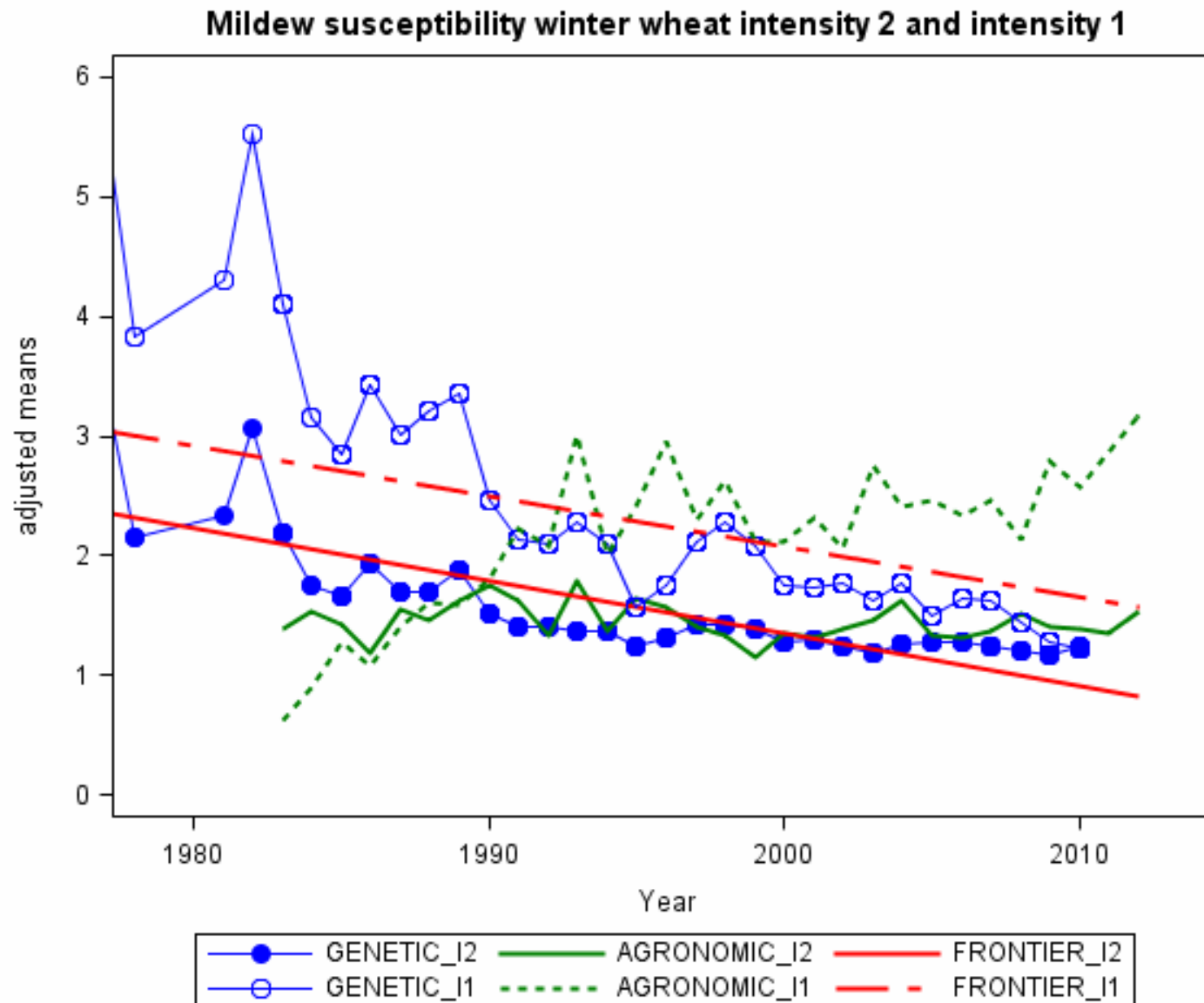
GENETIC:
variety group means

AGRONOMIC:
year means

FRONTIER:
frontier line

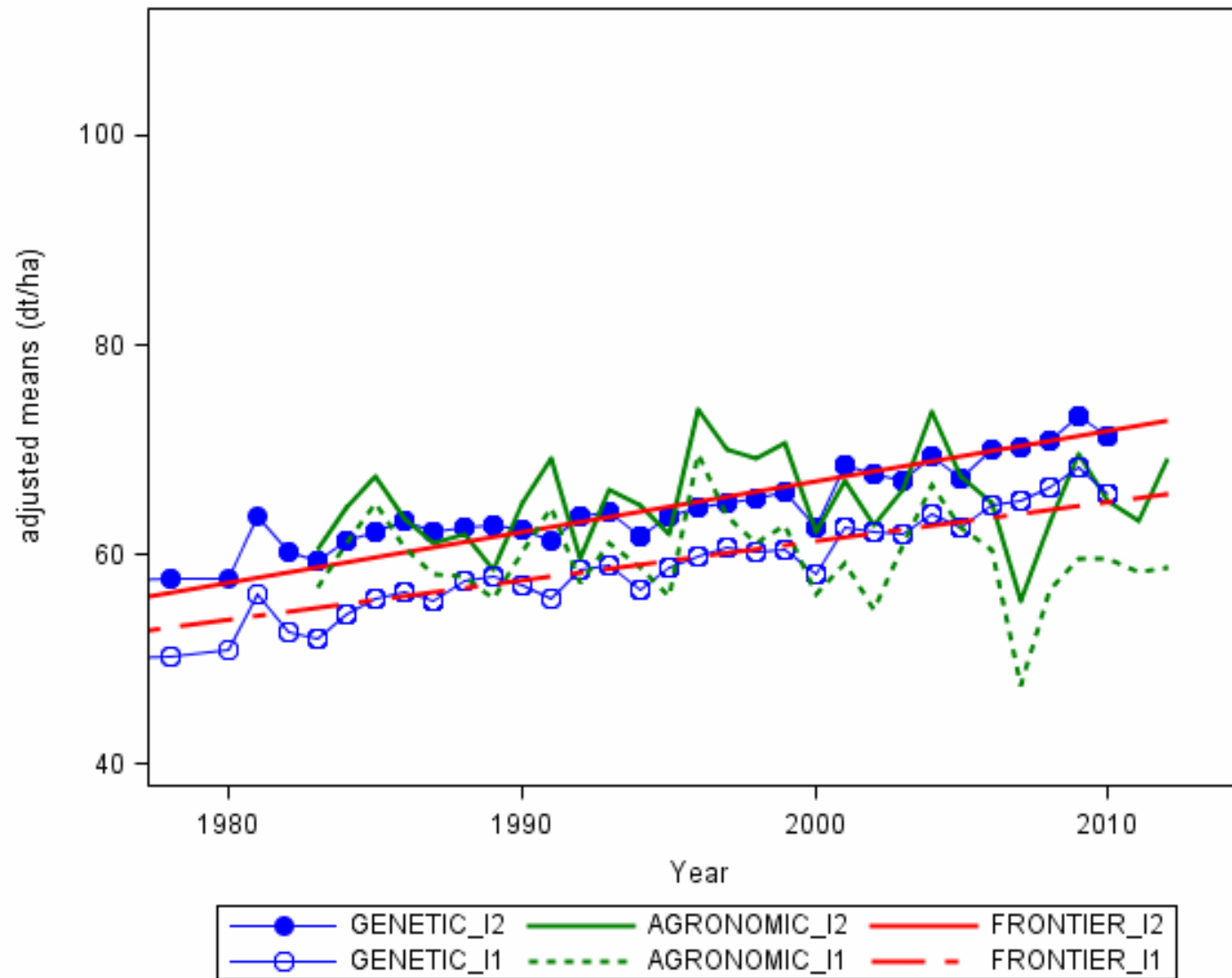
I1: Intensity 1
I2: Intensity 2

3. Application

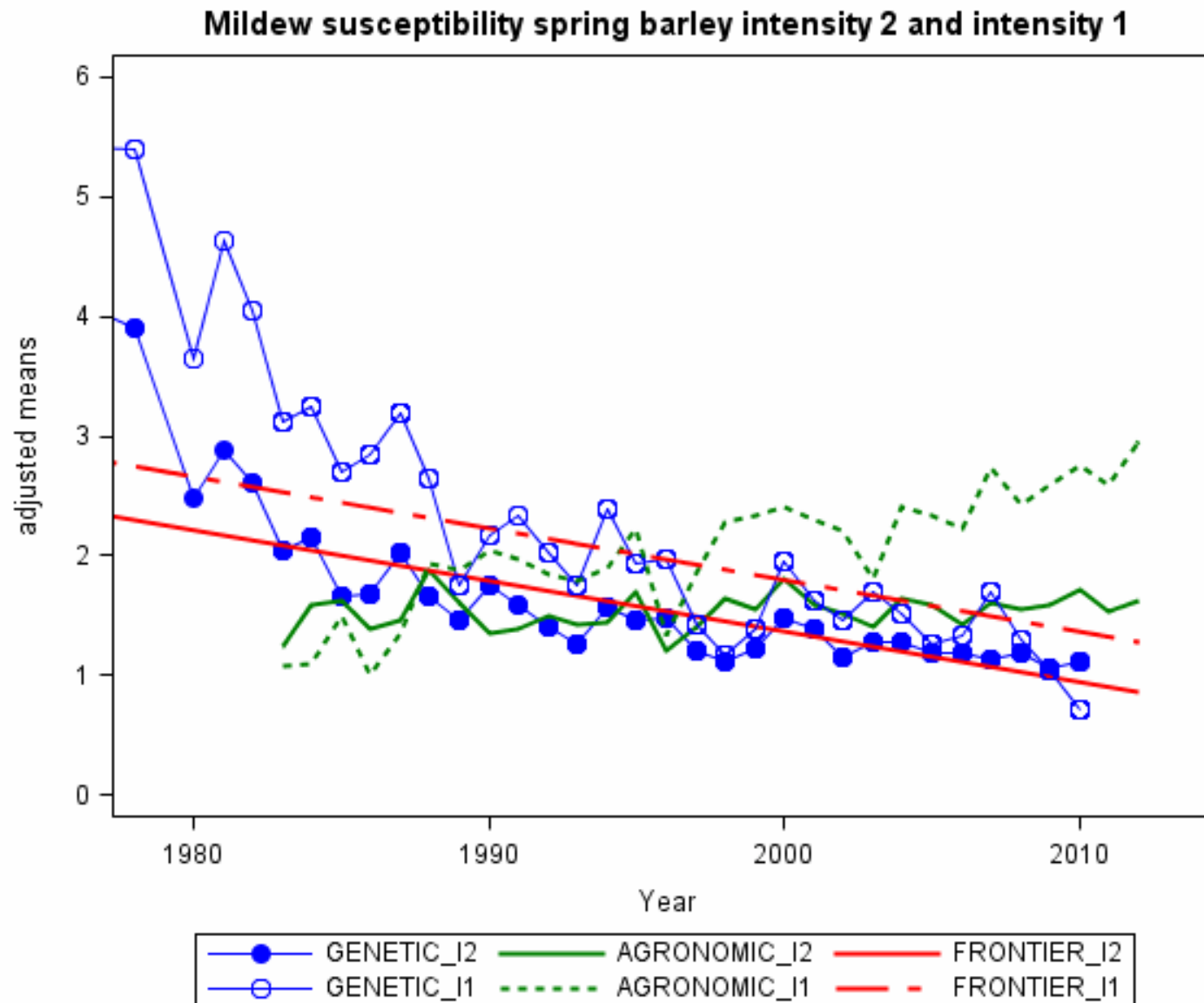


3. Application

Grain spring barley intensity 2 and intensity 1



3. Application



3. Application

Two more predictions from extended model

$$\eta_{ik2} - \eta_{ik1} = (\mu_2 - \mu_1) + \bar{\beta}r_i + \bar{\gamma}t_k \quad (15)$$

(3) Difference of intensities should have zero slope for frontier line

$$\bar{\beta} = \delta_1 - \delta_2 \text{ and } \bar{\gamma} = \delta_2 - \delta_1 \Rightarrow \bar{\beta} + \bar{\gamma} = 0$$

(4) Non-genetic and genetic trend should have opposing signs and equal absolute value

$$\bar{\beta} = -\bar{\gamma}$$

3. Application

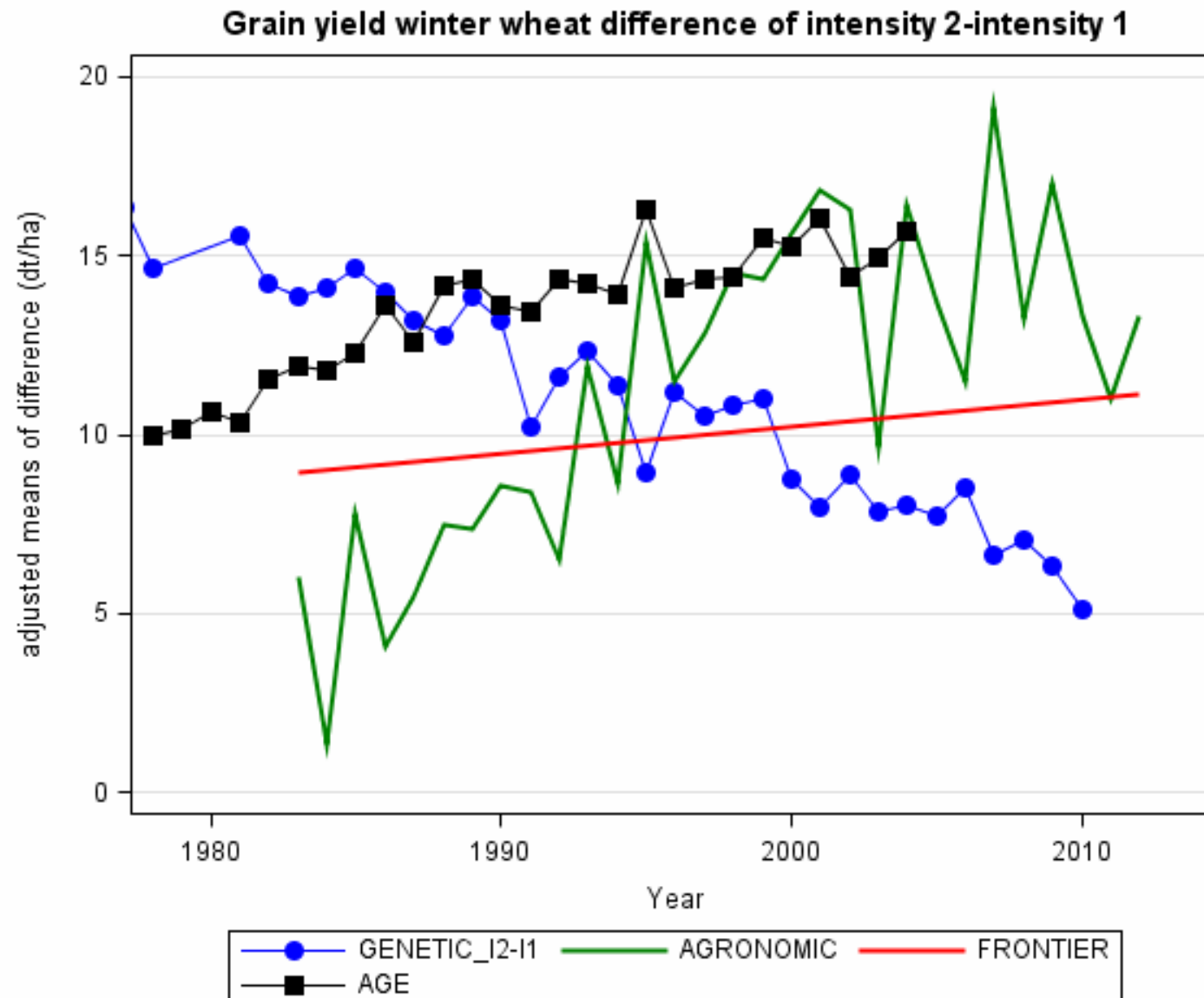


Figure 3:

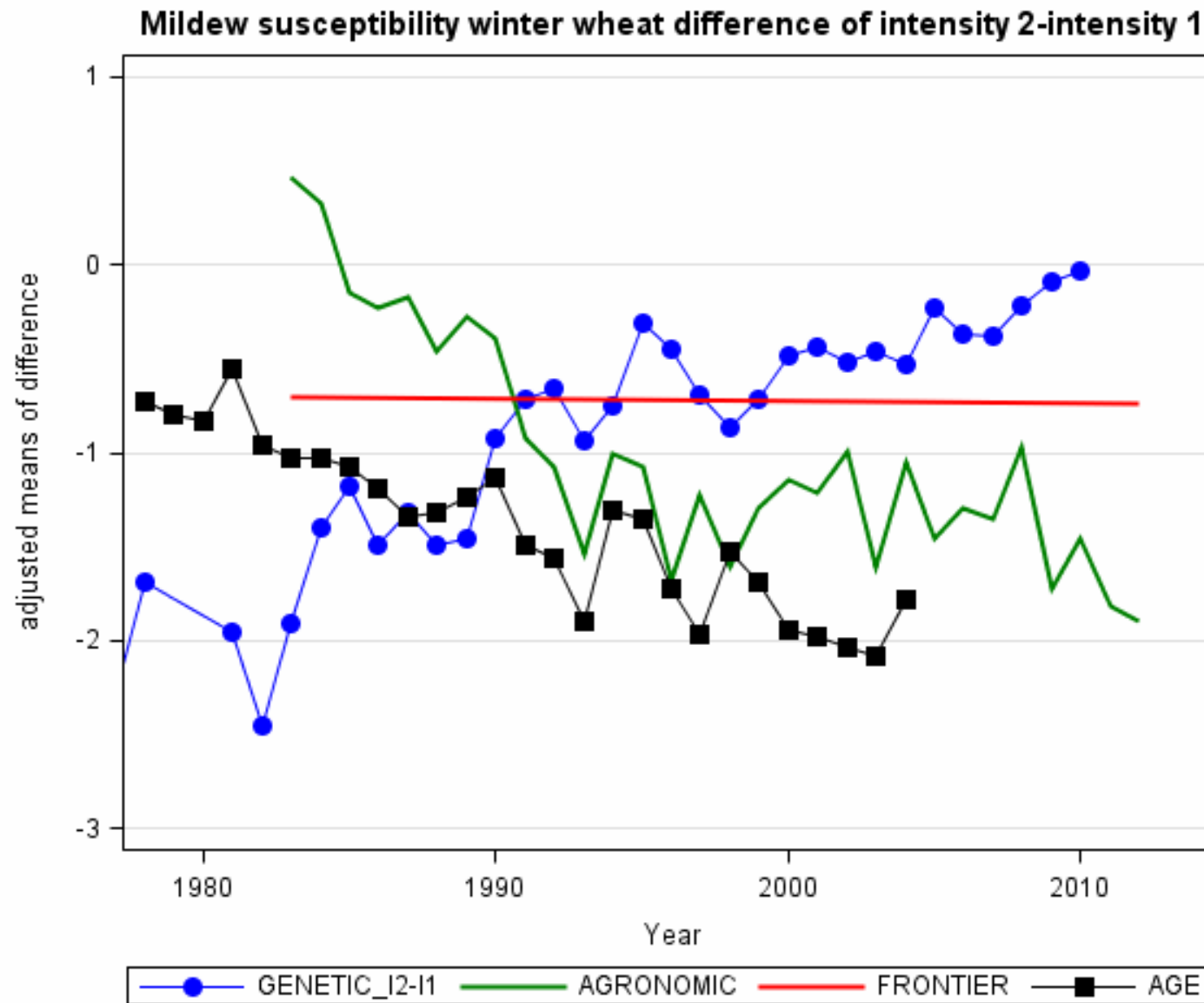
GENETIC:
variety group means

AGRONOMIC:
year means

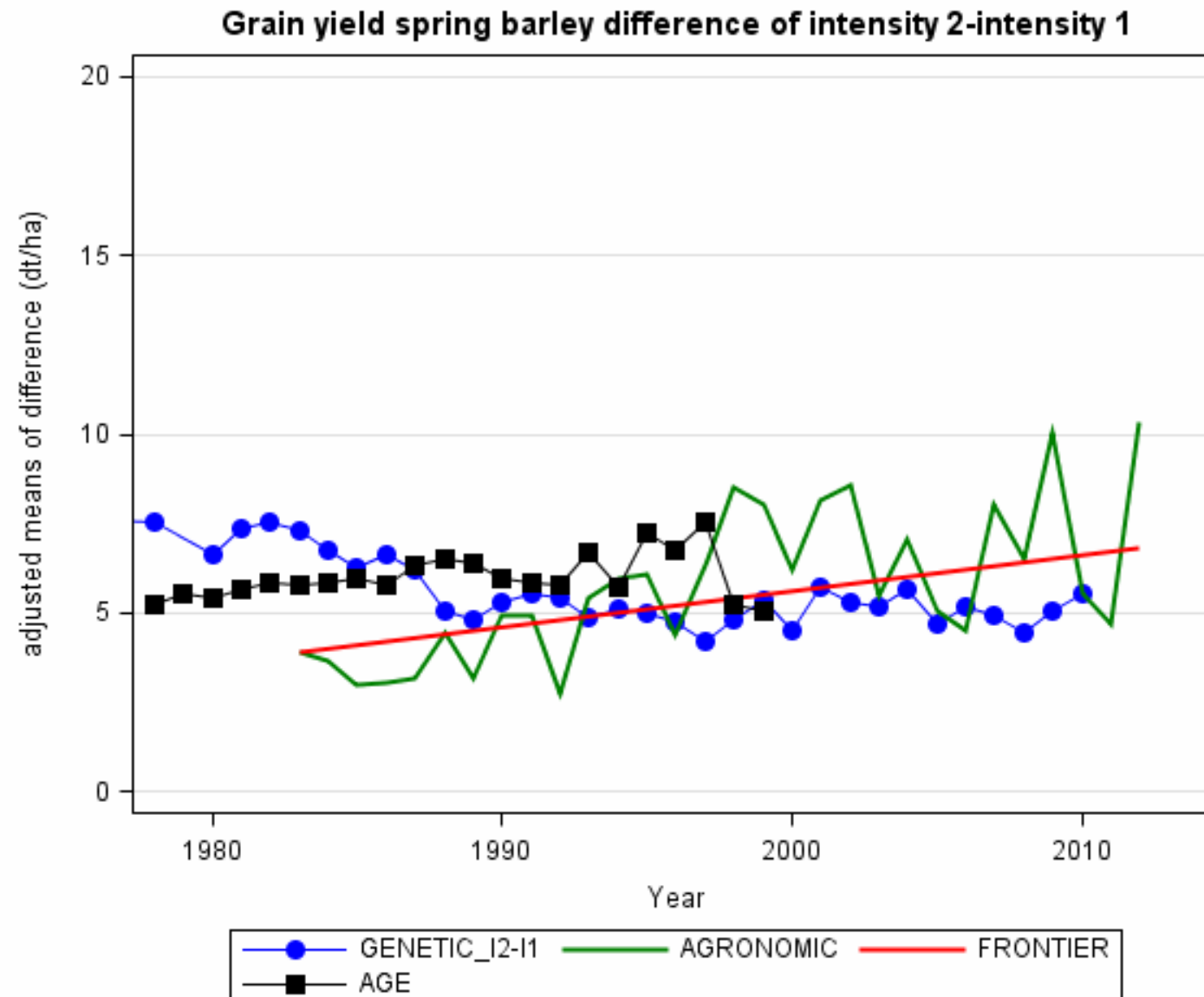
FRONTIER:
frontier line

Difference I2 - I1

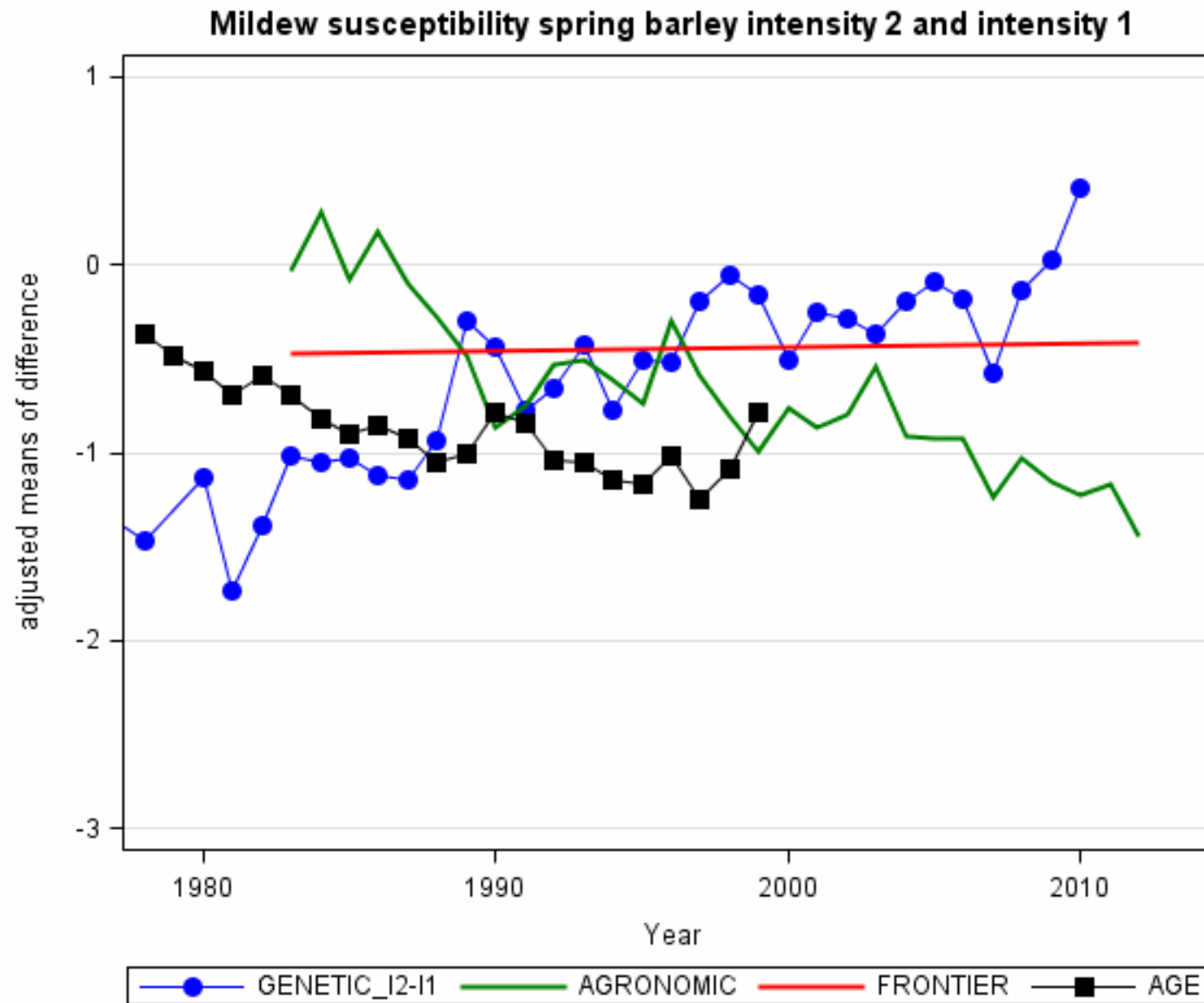
3. Application



3. Application



3. Application



3. Application

Silage maize

Total biomass (dt/ha):

$$\tilde{\beta} = 6.29$$

$$\tilde{\gamma} = -4.13$$

Dry matter yield (dt/ha):

$$\tilde{\beta} = 1.92$$

$$\tilde{\gamma} = -0.65$$

(Only one intensity)

⇒ Is this real, or are there any further biases in the agronomic trend?

3. Application

One potential problem

- In early years silage maize may have been tested mainly in favourable locations
- In later years, more marginal locations may have been added to the system
- This may have caused a downward bias in agronomic trend

Small simulation to explore this

	location									
Year	1	2	3	4	5	6	7	8	9	10
1										
2										
3										
4										
5										

3. Application

Simulation model

$$y_{jk} = L_j + Y_k + e_{jk} \quad (j = 1, \dots, 34; k = 1, \dots, 30)$$

$$L_j = M_j + \lambda(j - 35/2)$$

$$M_j \sim N(0,1)$$

λ : Trend for 'early' versus 'late' locations

$$Y_k \sim N(0,1)$$

$$e_{ij} \sim N(0,1)$$

3. Application

Analysis model

$$y_{jk} = \mu_k + L_j + Y_k + e_{jk}$$

Note: $\gamma = 0$ for simulation model!

	Estimate of γ	
λ	L_j fixed	L_j random
0.0	0.067 (se=0.065)	0.022 (se=0.029)
-0.5	0.067 (se=0.065)	-0.500 (se=0.029)

3. Application

Silage maize: real long-term data

Total biomass (dt/ha):

L_j random

$$\tilde{\beta} = 6.29 \text{ (s.e.} = 0.50)$$

$$\tilde{\gamma} = -4.13 \text{ (s.e.} = 1.22)$$

L_j fixed

$$\tilde{\beta} = 6.29 \text{ (s.e.} = 0.50)$$

$$\tilde{\gamma} = -4.44 \text{ (s.e.} = 1.25)$$

⇒ Trend in L_j does not seem to be a problem for silage maize

⇒ Comparing models with fixed and random L_j provided a good check for bias