UNIVERSITY OF HOHENHEIM

Modelling and evaluating breeding progress in rye for hybrid and population varieties

Friedrich Laidig
Biostatistics Unit
University of Hohenheim
Germany

Joint work with:

Hans-Peter Piepho, Hohenheim University

Dirk Rentel, Thomas Drobek and Uwe Meyer, Bundessortenamt, Germany

Alexandra Hüsken, Max-Rubner Institut, Germany

Models used in this study based on approaches described in

Piepho et al. (2014)

Laidig et al. (2016)




Questions

- Contribution of plant breeding to yield increases (genetic trend)?
- Contribution of improved agronomic practices and environmental factors to yield increases (non-genetic trend)?
- Effect of variety age on yield progress?
- Gap between trial and on-farm yields?

Overview

1. Basic mixed model
2. Genetic and non-genetic trends
3. Effect of variety age
4. Trial vs on-farm yield progress
5. Conclusions

Rye production

Cereal with
- high winter hardiness,
- high tolerance to other abiotic and biotic stress factors,
- suitable for nutrient-poor, sandy soils,
- out yield wheat and triticale under these poor conditions,
- rye bread has a high dietary value.

Country	Production (1000 t)	Acreage (1000 ha)	Grain yield (t ha^{-1})
Germany	3,174	571	5.56
Russian Federation	2,541	1,250	2.03
Poland	2,200	761	2.89
Belarus	651	241	2.70
Denmark	577	100	5.80
China	525	164	3.21
Ukraine	392	144	2.73
Canada	382	131	2.92
USA	342	168	2.04
Spain	316	157	2.01
WORLD	**12,944**	**4,403**	**2.94**

Source: FAO (2018)

Data

- Official VCU (value for cultivation and use) trials in Germany
- Regular trial period three years
- Only released hybrid and population varieties included
- Intensities:
  - Intensity 1 - no crop protection and growth regulators
  - Intensity 2 – fungicides, equal or higher nitrogen fertilization rates
- Split-plot design:
  - Intensity on main plots (RCBD)
  - Variety on sub-plots (completely randomized)
  - Hybrid and population varieties randomized together

Data (cont.)

- Investigated years 1985 – 2016
- Traits
  - Grain yield (dt ha^{-1})
  - Ear density (ears m^{-2})
  - Single ear weight (g ear^{-1})
- On-farm data

  National averages for grain yield (1985-2016) from

  harvest survey comprising all types of varieties

Basic information (VCU trials)

	Hybrid varieties	Population varieties
Total # varieties	68	23
Standards	18	10
First trial year	1982	1955
Years in trial	7	9
Observations	6500	3600
Trials	1300	
Locations	105	
% GxYxL combinations	2.45	2.56

Basic model for long-term MET data

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

y_{ijk} = mean yield of the i-th genotype in the j-th location and k-th year
μ = general mean
G_i = main effect of the i-th genotype (variety)
L_j = main effect of the j-th location,
Y_k = main effect of the k-th year
$(LY)_{jk}$ = jk-th location × year interaction
$(GL)_{ij}$ = ij-th genotype × location interaction
$(GY)_{ik}$ = ik-th genotype × year interaction
$(GLY)_{ijk}$ = residual comprising both genotype × location × year interaction and error of the mean

⇨ Separate analysis for hybrid and population varieties

Graphical representation

Take G_i and Y_k as fixed (can't take random because of time trend) and the other effects as random i. i. d. normal with constant variance

Adjusted means for G_i assess genetic trend
$\Rightarrow$ Plotted against year in which variety entered trial (first trial year)

Adjusted means for Y_k assess non-genetic trend
$\Rightarrow$ Plotted against harvest year
(Two-step approach; *Mackay et al., 2011; Rijk et al. 2012)*

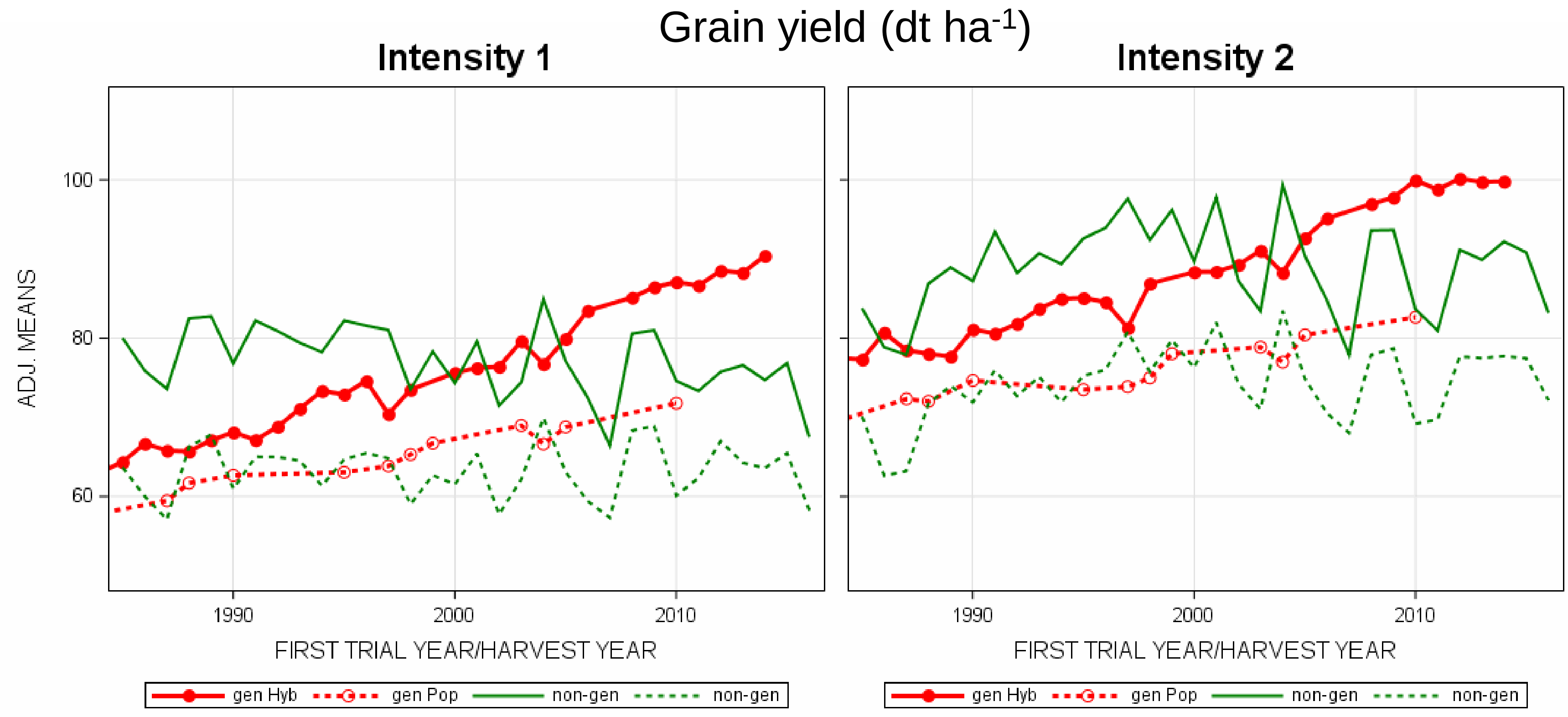
Alternative representation of genetic trend

$G_i = C_p + H_i'$, where C_p is a categorical effect for groups of varieties with the same first trial year r_i. H_i' is the random deviation from group mean.

Adjusted means for C_p assesses genetic trend
$\Rightarrow$ Plotted against first trial year r_i

2. Genetic and non-genetic trends

Graphical representation (cont.)



yield levels ⇒ I2 > I1, hybrids > population varietes

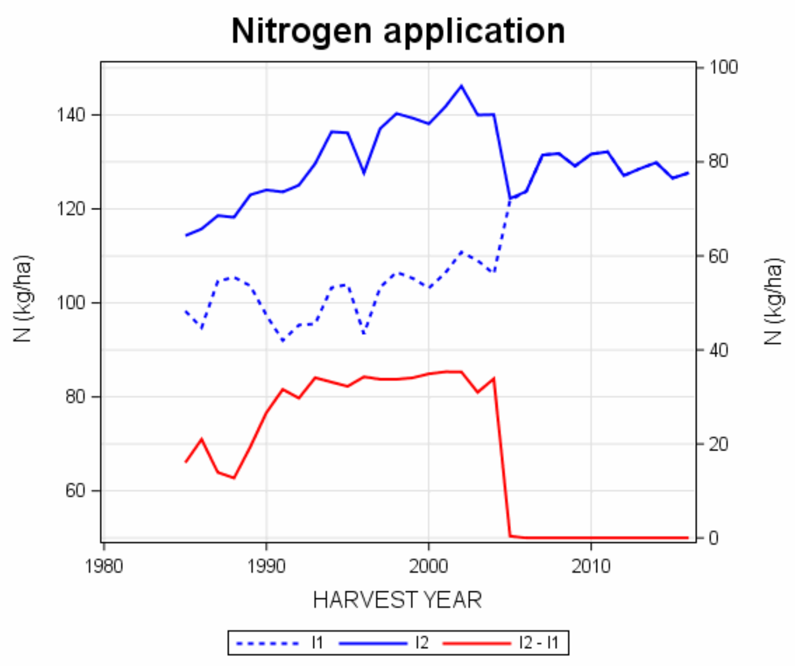
genetic trends ⇒ about linear increasing

non-genetic trends ⇒ parallel year-to-year variation

⇒ different slopes: period 1985 – 2004 and 2005 – 2016

2. Genetic and non-genetic trends

Reason for change of non-genetic trend?



Change of nitrogen application rate

- Period *p*=1 1985 – 2004; N-rate I2 ≠ I1
- Period *p*=2 2005 – 2016; N-rate I2 = I1

⇒ Focus on disease susceptibility of varieties,

2. Genetic and non-genetic trends

Genetic trend

$G_i = \beta r_i + H_i \qquad H_i \sim \mathrm{N}(0, \sigma_H^2)$

β fixed regression coefficient for genetic trend
r_i first year in trial of *i*-th variety

Non-genetic trend (two regression lines)

$Y_{pk} = \mu_p + \gamma_p\, t_{pk} + Z_k$, where $p = 1, 2 \quad Z_k \sim \mathrm{N}(0, \sigma_Z^2)$

γ_p fixed regression coefficient for non-genetic trend of period p

If t_k is the harvest year, $t_{1k} = \begin{cases} t_k & if \quad t_k \leq 2004 \\ 0 & if \quad t_k > 2004 \end{cases}$ and

$$t_{2k} = \begin{cases} 0 & if \quad t_k < 2005 \\ t_k & if \quad t_k \geq 2005 \end{cases}$$

$E(y_{pijk}) = \eta_{pik} = \mu_p + \beta r_i + \gamma_p\, t_{pk}$ fixed (regression) part of model

2. Genetic and non-genetic trends

Results

I2 < I1 (genetic) — I2 > I1 (non-genetic)

Traits	Int.	Type	Trends: genetic (r_i) Slope β	&	non-genetic (t_{pk}) Slope γ_1	Slope γ_2	§
Yield (dt ha^{-1})	2	Hyb	0.77***	ns	0.55**	0.28ns	*
	2	Pop	0.24***	*	0.64***	0.41ns	**
	1	Hyb	0.87***	ns	-0.14ns	0.00ns	ns
	1	Pop	0.30***	ns	0.06ns	0.17ns	ns
Ear density (ears m^{-2})	2	Hyb	2.49***	ns	2.10ns	0.91ns	ns
	2	Pop	0.87*	ns	0.87ns	-0.14ns	ns
	1	Hyb	3.02***	*	0.36ns	1.59ns	ns
	1	Pop	0.65ns	ns	-0.06ns	1.36ns	ns
Ear weight (g ear^{-1})	2	Hyb	0.007***	ns	0.007ns	0.005ns	ns
	2	Pop	0.002*	ns	0.010*	0.009ns	*
	1	Hyb	0.009***	**	-0.001ns	-0.003ns	ns
	1	Pop	0.005***	ns	0.003ns	0.001ns	ns

Change of agronomic conditions 1985 - 2016
VCU trials

Sowing rate	-43***	kernels m^{-2} (15%)
Sowing date	3*	days earlier
Harvesting date	1ns	days earlier
Daily air temperature	+0.9***	°C

& Test of deviation from <u>linear</u> genetic trend
§ Test of deviation from <u>unique</u> non-gen trend

Significance levels
- * 5%
- ** 1%
- *** 0.1%

3. Effect of variety age

Extended regression model

- Disease susceptibility may increase with time for a number of years
- Inefficient maintenance breeding may lower performance, etc.
- This is expected to have an effect on time trends for yield etc.
- Use age at testing as another covariate in the model

⇨ $a_{ik} = t_k - r_i$ age at testing for the *i*-th variety in the *k*-th harvest year

$\eta_{pik} = \mu_p + \beta r_i + \gamma_p t_{pk} + \delta a_{ik}$ extended regression model

δ fixed regression coefficient for age covariate a_{ik} and of negative value for yield

Extended regression model (cont.)

Problem: Model over-parameterized (multi-collinearity)

$$\eta_{pik} = \mu_p + \beta r_i + \gamma_p\, t_{pk} + \delta(t_k - r_i)$$

Can be reparametrized as:

$\eta_{pik} = \mu_p + \tilde{\beta} r_i + \tilde{\gamma}_p\, t_{pk}$, where
$\tilde{\beta} = \beta - \delta$ and
$\tilde{\gamma}_p = \gamma_p + \delta$

⇒ Regression on r_i and t_k may be biased due to age effects!
⇒ Can't separate out trend due to age effect

True genetic trend: $\beta = \tilde{\beta} + \delta$
True non-genetic trend: $\gamma_p = \tilde{\gamma}_p - \delta$

If age effect negative: $\delta < 0$:

$\beta < \tilde{\beta}$ ⇒ genetic trend over-estimated
$\gamma_p > \tilde{\gamma}_p$ ⇒ non-genetic trend under-estimated

3. Effect of variety age

How to estimate age effect δ ?

⇒ Compare the two intensities

Intensity 1: $\eta_{pik1} = \mu_{p1} + \tilde{\beta}_1 r_i + \tilde{\gamma}_{p1}\, t_{pk}$, where $p = 1, 2$

Intensity 2: $\eta_{pik2} = \mu_{p2} + \tilde{\beta}_2 r_i + \tilde{\gamma}_{p2}\, t_{pk}$

Intensity 1		Intensity 2
$\tilde{\beta}_1 = \beta - \delta_1$	and	$\tilde{\beta}_2 = \beta - \delta_2$
$\tilde{\gamma}_{p1} = \gamma_p + \delta_1$	and	$\tilde{\gamma}_{p2} = \gamma_p + \delta_2$

Assumptions:
(1) genetic trends identical in I1 and I2
(2) non-genetic trends identical in I1 and I2

Difference of response for both intensities under assumptions (1), (2)

$$\eta_{pik2} - \eta_{pik1} = (\mu_{p2} - \mu_{p1}) + (\delta_1 - \delta_2) r_i - (\delta_1 - \delta_2) t_{pk}$$
$$= (\mu_{p2} - \mu_{p1}) - (\delta_1 - \delta_2) a_{ik}$$

⇒ regression on a_{ik} estimates $\bar{\delta} = -(\delta_1 - \delta_2)$

How to estimate age effect δ (cont.)?

⇒ Compare the two intensities

Intensity 1: $\eta_{pik1} = \mu_{p1} + \tilde{\beta}_1 r_i + \tilde{\gamma}_{p1}\, t_{pk}$, where $p = 1, 2$

Intensity 2: $\eta_{pik2} = \mu_{p2} + \tilde{\beta}_2 r_i + \tilde{\gamma}_{p2}\, t_{pk}$

Intensity 1		Intensity 2
$\tilde{\beta}_1 = \beta - \delta_1$	and	$\tilde{\beta}_2 = \beta - \delta_2$
$\tilde{\gamma}_{p1} = \gamma_p + \delta_1$	and	$\tilde{\gamma}_{p2} = \gamma_p + \delta_2$
$\tilde{\gamma}_{p1} = \gamma_{p1} + \delta_1$	and	$\tilde{\gamma}_{p2} = \gamma_{p2} + \delta_2$

Assumptions:

(1) genetic trends indentical

(2) non-genetic trends identical

(3) non-genetic trends not identical in I1 and I2

Difference of response for both intensities under assumptions (1), (3)

$$\eta_{pik2} - \eta_{pik1} = (\mu_{p2} - \mu_{p1}) + (\delta_1 - \delta_2) r_i + \left[\left(\gamma_{p2} - \gamma_{p1} \right) - (\delta_1 - \delta_2) \right] t_{pk}$$
$$= (\mu_{p2} - \mu_{p1}) + \left(\gamma_{p2} - \gamma_{p1} \right) t_{pk} - (\delta_1 - \delta_2) a_{ik}$$

⇨ joint regression on t_{pk} and a_{ik} estimates $\bar{\delta} = -(\delta_1 - \delta_2)$

3. Effect of variety age

Graphical representation

Consider basic model of differences of I2 – I1

$$y_{p1ijk2} - y_{p1jk1} =$$
$$= (\mu_{p2} - \mu_{p1}) + G_i + L_j + Y_{pk} + (LY)_{jk} + (GL)_{ij} + (GY)_{ik} + (GLY)_{ijk}$$

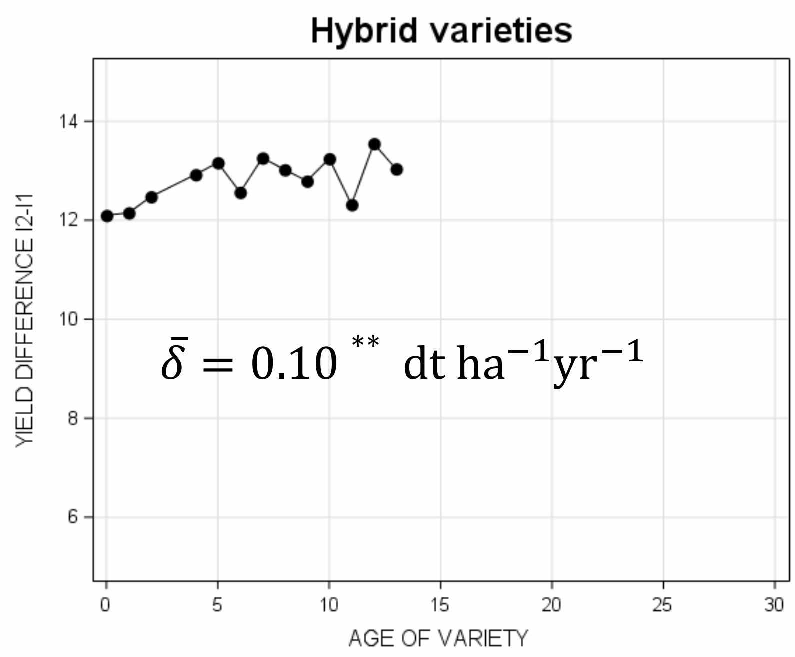
Representation of age effects:
Take Y_{pk} be as fixed effect and
$(GY)_{ik} = D_q + (ZH)_{ik}$ where D_q is a categorical fixed effect for the q-th age class, and the other effects are random.

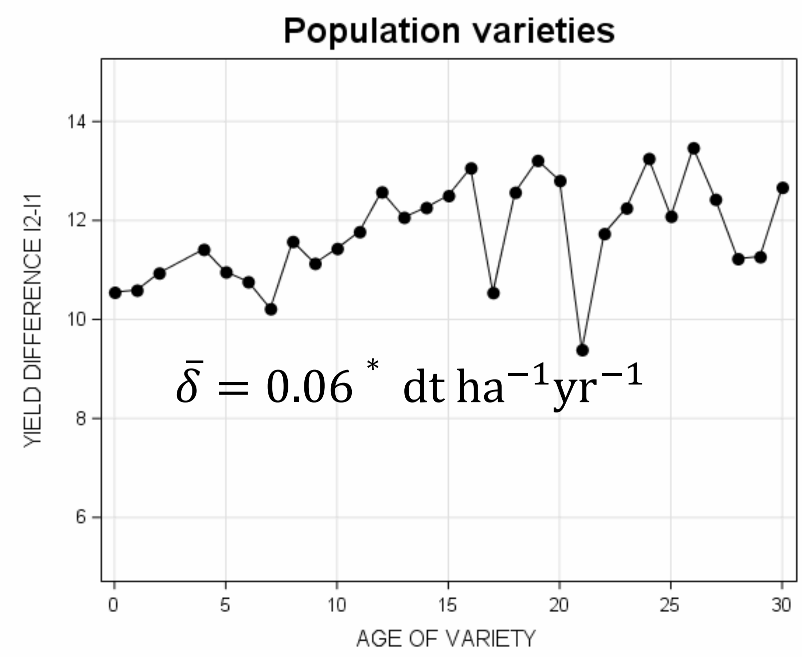
Jointly estimate adjusted means for Y_{pk} and D_q
Adjusted age means for D_q visualizes age trend $\bar{\delta} \Rightarrow$ Plotted against age

3. Effects of variety age

Graphical representation (cont.)

Grain yield: age trends for I2 – I1: $\bar{\delta} = -(\delta_1 - \delta_2)$





If $\delta_2 = 0$ then grain yield for I1

⇒ hybrid variety decreasing by 0.10 dt ha^{-1} yr^{-1}

⇒ population variety decreasing by 0.06 dt ha^{-1} yr^{-1}

3. Effects of variety age

Results(cont.)

Age effect $\bar{\delta} = -(\delta_1 - \delta_2)$, $\delta_2 = 0$ not realistic, no full disease control

⇨ assume $\delta_2 = \frac{1}{3}\delta_1$

Yield	Int.		Hybrids	Population verities
Age effect	I2-I1	$\bar{\delta}$	0.10	0.06
	I1	δ_1	-0.15	-0.09
	I2	δ_2	-0.05	-0.03
True genetic trend	I1	$\beta_1 = \tilde{\beta}_1 + \delta_1$	0.87-0.15=0.72	0.30-0.09=0.21
	I2	$\beta_2 = \tilde{\beta}_2 + \delta_2$	0.77-0.05=0.72	0.24-0.03=0.21
True non-genetic trend	I1, *p*=1	$\gamma_{p1} = \tilde{\gamma}_{p1} - \delta_1$	-0.14-(-0.15) =0.29	0.06-(-0.09) =0.15
	I1, *p*=2		0.00-(-0.15) =0.15	0.17-(-0.09) =0.26
	I2, *p*=1	$\gamma_{p2} = \tilde{\gamma}_{p2} - \delta_2$	0.55-(-0.05) =0.60	0.64-(-0.03) =0.67
	I2, *p*=2		0.28-(-0.05) =0.33	0.41-(-0.03) =0.44

Reduced basic model for long-term MET data

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

y_{ijk} = mean yield of the i-th genotype in the j-th location and k-th year
μ = overall mean
L_j = main effect of the j-th location
Y_k = main effect of the k-th year, (confounded with genotype effect), fixed
$(LY)_{jk}$ = jk-th location × year interaction
$(GL)_{ij}$ = ij-th genotype × location interaction
$(GY)_{ik}$ = ik-th genotype × year interaction
$(GLY)_{ijk}$ = residual comprising both genotype × location × year interaction and error of the mean

4. Trial vs on-farm progress

Overall trend

$Y_{pk} = \mu_p + \varphi_p \, t_{pk} + U_k$, where $p = 1,2$; $\quad U_k \sim N(0, \sigma_U^2)$

When t_k is the harvest year, $t_{1k} = \begin{cases} t_k & if \quad t_k \leq 2004 \\ 0 & if \quad t_k > 2004 \end{cases}$ and

$$t_{2k} = \begin{cases} 0 & if \quad t_k < 2005 \\ t_k & if \quad t_k \geq 2005 \end{cases}$$

φ_p fixed regression coefficient for overall trend of period p

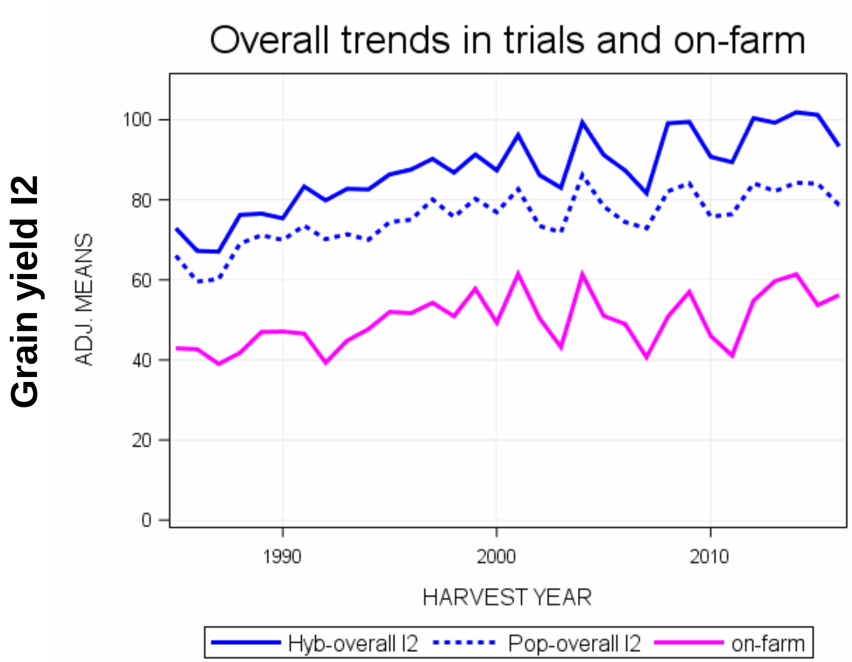
$E(y_{pijk}) = \eta_{pik} = \mu_p + \varphi_p \, t_{pk}$ fixed part of reduced model

Graphical representation of overall trend

Adjusted means for Y_k visualizes overall trend
⇒ Plotted against harvest year

4. Trial vs on-farm progress

Yield gap



Questions
1. Why plateauing of on-farm yields after year 2000?
2. Why large yield gap?

Results

		Grain yield (dt/ha)									
		Trend estimates					Prediction			gap	
		overall (t_{pk})					overall			trial - on-farm	
Type	Int.	slope φ_1		slope φ_2		§	1985	2016	diff.	1985	2016
Hyb	2	1.24	***	1.03	*	ns	71.1	100.2	29.1	27.3	44.5
Pop	2	0.90	***	0.60	ns	*	64.5	83.1	18.6	20.6	27.4
on-farm	-	0.38	**	-			43.8	55.7	11.9	-	-

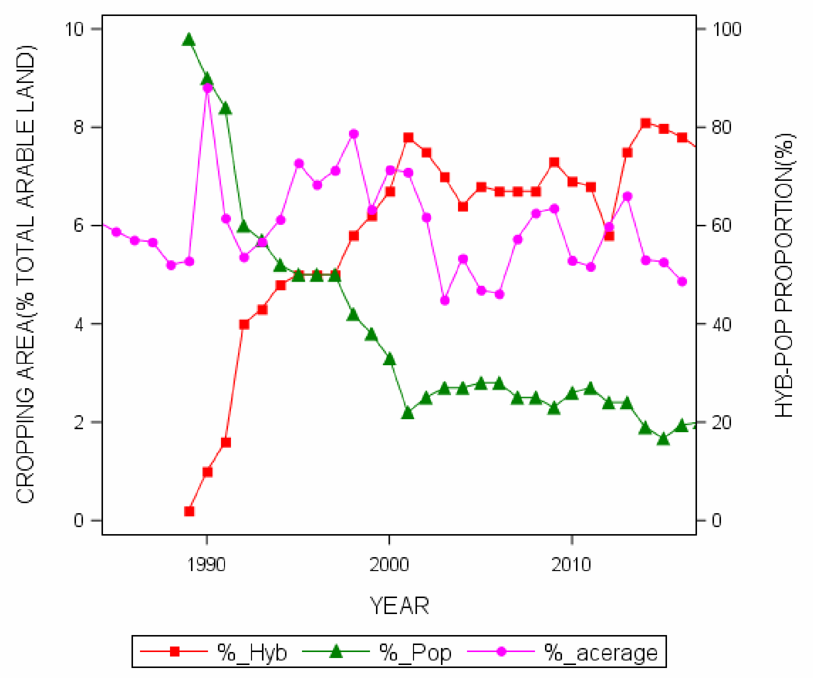
Denmark 58 dt ha^{-1}

§ Test of deviation from <u>unique</u> overall trend

Significance levels
* 5%
** 1%
*** 0.1%

4. Trial vs on-farm progress

1. Why plateauing of on-farm yields after year 2000 ?



⇨ Transition to higher yielding hybrid varieties completed after year 2000

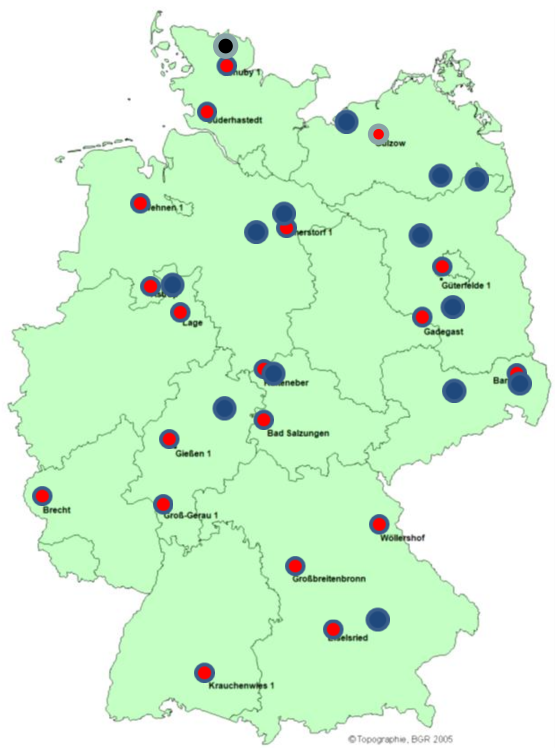
4. Trial vs on-farm progress

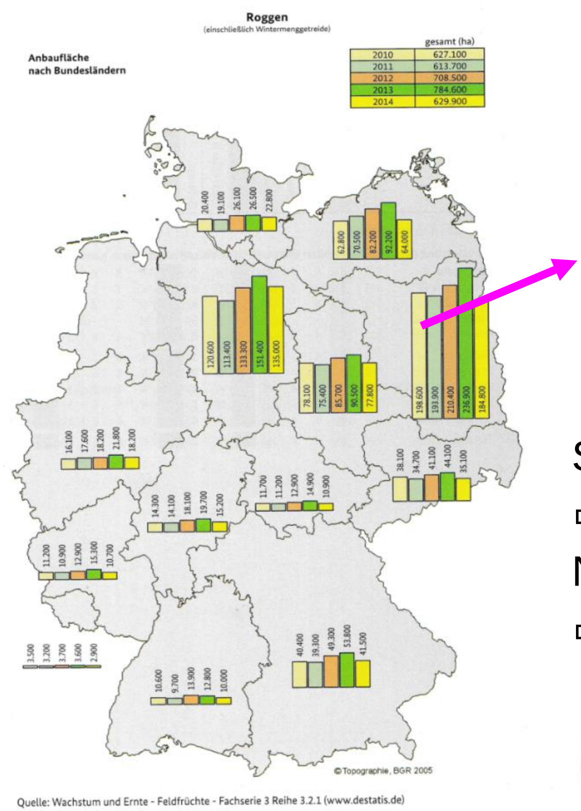
2. Why large yield gap between trials and on-farm yield ?

VCU trial sites 2016

On-farm cropping area by federal states 2016

Soil quality (Ackerzahl) 45 - 50





30 % of national acerage
Pop 39% (national average 20%)

Site conditions (soil fertility)

⇨ trials > on-farm

Not full yield potential on-farm

⇨Economic yield optimum

● VCU 1: 1st testing year

● VCU 2: 2nd and 3rd testing year

5. Conclusions

- Genetic trends for hybrids ∼ three times of population varieties
- Change in N rate in 2005 ⇨ two non-genetic regression functions
- Genetic and non-genetic trends ⇨ biased due to age effect
- On-farm yield level $\sim \frac{1}{2}$ of hybrids in trials
- Widening yield gap between on-farm and trials
- Gaps caused mainly by better growing conditions at trial sites and by economical factors

References

Laidig F, Piepho HP Rentel D, Drobek T, Meyer U, Huesken A (2017) Breeding progress, variation, and correlation of grain and quality traits in winter rye hybrid and population varieties and national on-farm progress in Germany over 26 years. Theor Appl Genet 130:981–998

Mackay IJ, Horwell A, Garner J, White J, McKee J, Philpott H (2011) Reanalysis 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. Theor Appl Genet 122:225–238

Piepho HP, Laidig F, Drobek T, Meyer U (2014) Dissecting genetic and non-genetic sources of long-term yield in German official variety trials. Theor Appl Genet 127:1009-1018

Rijk B, van Ittersum M, Withagen J (2013) Genetic progress in Dutch crop yields. Field Crops Research 149:262–268