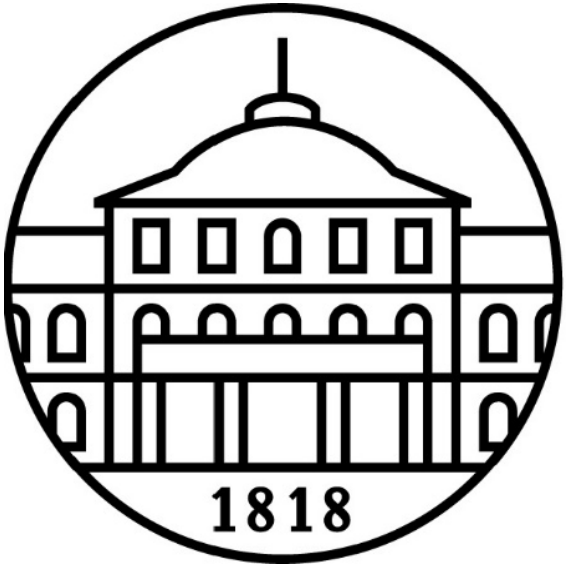
Biometrie in der Nutzpflanzenforschung

Hans-Peter Piepho

Fachgebiet Biostatistik
Universität Hohenheim
Stuttgart



Kolloquium "Statistische Methoden in der empirischen Forschung"

Analyse von Prüfglied-Umwelt Interaktionen in landwirtschaftlichen Versuchen.
Humboldt Universität Berlin, 8. Dezember 1998.

Lineare Varianzstrukturen in Feldversuchen.
Humboldt-Universität Berlin, 11. November 2008.

The use of two-way linear mixed models in multi-treatment meta-analysis.
Humboldt-Universität Berlin, 19. November 2013.

Mehrortige Sortenversuche in benachbarten Anbaugebieten – was ist die beste Ertragsschätzung pro Anbaugebiet?
Humboldt-Universität Berlin, 2. Dezember 2014.

Wolfgang Köhler

Erinnerung an einen Fachkollegen und Freund

Wolfgang Köhler

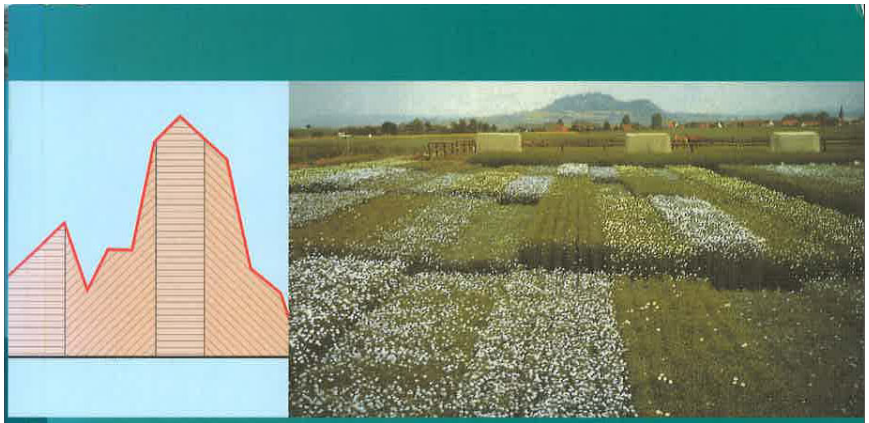
* 16. Juli 1941

† 21. Februar 2017

Am 21. Februar 2017 verstarb **Prof. Dr. rer. nat. Dr. h. c. Wolfgang Köhler** im Alter von 75 Jahren.

Nach seinem Mathematik- und Physik-Studium in Marburg und Berlin wurde er 1968 Wissenschaftlicher Mitarbeiter, Wissenschaftlicher Assistent und Assistenzprofessor an der FU Berlin, wo er 1973 promovierte und sich 1977 im Fachgebiet Biometrie und Genetik habilitierte. In seiner Berliner Zeit begründete er mit Kollegen ein Biometrisches Kolloquium, das bis heute als Berliner Kolloquium „Statistische Methoden in der empirischen Forschung" fortgeführt wird.

1980 wurde er zum Professor für *Biometrie und Populationsgenetik* an die *Justus-Liebig-Universität Gießen* berufen. Dort baute er die „Abteilung Biometrie" am FB Agrarwissenschaften auf und leitete diese bis zu seiner Pensionierung im Jahr 2006.



Köhler · Schachtel
Voleske

Biostatistik

Eine Einführung für Biologen und Agrarwissenschaftler

4. Auflage

Springer

Wolfgang Köhler
Professor für Biometrie und Populationsgenetik an der Universität Gießen

Gabriel Schachtel
Institut für Biometrie und Populationsgenetik der Universität Gießen

Peter Voleske
Biometriker in der pharmazeutischen Industrie

Biostatistik

Die Vermessung des Lebendigen

Rasch statistische Verfahren zur Auswertung experimenteller Daten lernen und handhaben – das ist mit dieser bewährten und sehr gut verständlichen Einführung in die Biometrie ganz einfach. Grundlagen und Anwendung werden gleichermaßen dargestellt, wobei biologische und agrarwissenschaftliche Beispiele besonders berücksichtigt werden. Die Autoren vermitteln dem Anwender – soweit möglich ohne Formeln und mathematische Symbolik – zunächst ein Verständnis für die hinter dem Verfahren stehende Grundidee. Übersichtliche Rechenanleitungen regen den Leser an, die Verfahren anhand der Beispieldaten wirklich nachzurechnen. Das Buch ist als Begleittext zu einer Vorlesung ebenso geeignet wie zum Selbststudium, um darauf aufbauend alleine oder mit dem Statistiker komplexe statistische Verfahren anzuwenden.

Das Werkzeug dazu finden Sie in diesem Buch.

ISBN 978-3-540-37710-8

9 783540 377108

springer.de

Biometrie in der Nutzpflanzenforschung

Die Modellierung von Daten aus Versuchen in den Nutzpflanzenwissenschaften basiert meistens auf linearen gemischten Modellen. Der Vortrag stellt einige aktuelle Beispiele solcher Anwendungen vor und diskutiert methodische Herausforderungen und Zukunftsperspektiven sowie Bezüge zu benachbarten Disziplinen in den biologischen Wissenschaften. Eine wichtige Anwendung ist Projektion der Effekte des Klimawandels basierend auf Versuchsserien.

Picture: Volker Michel, Gülzow

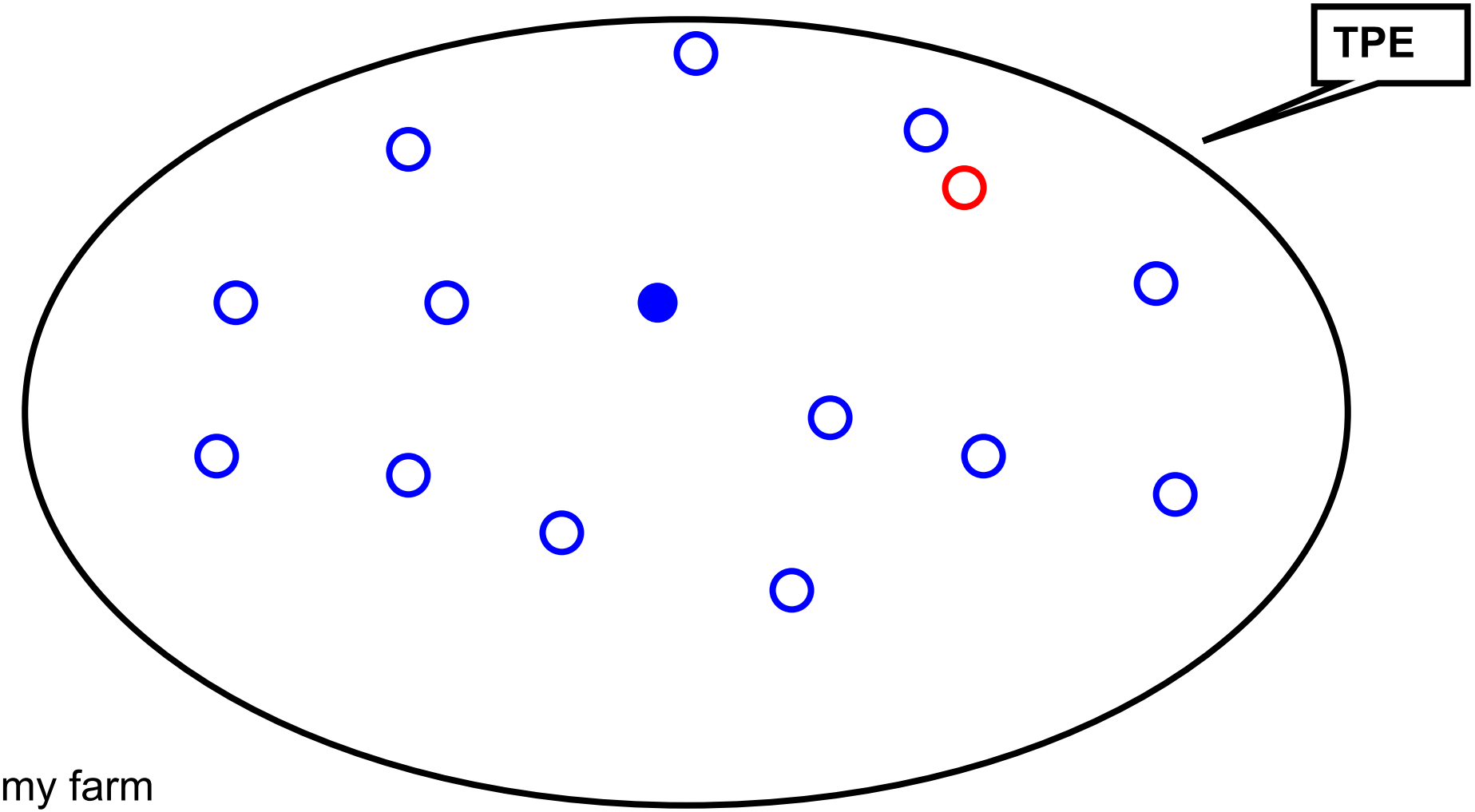
Picture: Martin Zahner, Ihinger Hof

FIELD-BASED HIGH-THROUGHPUT SYSTEM

Monitoring chlorophyll fluorescence in natural environments

High time resolution (< 1 s) assessment of photosynthetic performance at a distance

Picture: Nicolas Zendonadi, University of Bonn (trials in Arizona, USA)



my farm

locations of trial network

"mean" of target population of environments (TPE) (Piepho & Williams 2024)

1. Introduction

When and where it all began:

THE ANALYSIS OF GROUPS OF EXPERIMENTS

By F. YATES and W. G. COCHRAN

Statistical Department, Rothamsted Experimental Station, Harpenden

(With One Text-figure)

Journal of Agricultural Science (Cambridge) **28** (1938), 556-580

Yates and Cochran (1938):

"Any experimental programme which is instituted to assess the value of any particular treatment or practice or to determine the optimal amount of such treatment should therefore be so designed that it is capable of furnishing an accurate and unbiased estimate of the average response to this treatment in the various combinations of circumstances in which the treatment will subsequently be applied. The simplest and indeed the only certain way of ensuring that this condition shall be fulfilled is to choose fields on which the experiments are to be conducted by random selection from all fields which are to be covered by the subsequent recommendations."

Yates and Cochran (1938) go on to acknowledge that

"it is usually impossible to secure a set of sites selected entirely at random. An attempt can be made to see that the sites actually used are a "representative" selection, but averages of the responses from such a collection of sites cannot be accepted with the same certainty as would the averages from a random collection."

- MET are routinely analysed using linear mixed models.

- A key step in the use of such models is to decide for each factor whether it is to be regarded as fixed or random.

- Once the status of each factor as either fixed or random has been established, one may use the convention that if an effect involves at least one random factor, it should be modelled as random (Nelder 1977).

- Hence, if environment is considered as a random factor, it follows that variety $\times$ environment interactions are to be modelled as a random effect.

Fisher (1935, Section 65):

"In fact, if our concern is to ascertain not merely the best variety on the aggregate of the (...) fields actually used, but to ascertain which is the best over the whole area deemed suitable for this type of crop, within the region from which the sites of the experiment have been selected, the comparison between varieties V and interaction of varieties and places VP will be the more appropriate. For if the (…) sites have been chosen at random from this area, a significant difference in this comparison would indicate, at the level of significance used, varietal differences applicable to the whole area."

Yates and Cochran (1938) assert that

"in the ideal case where the chosen places are a strictly random selection from all possible places (…) it would appear to be legitimate (…) to compare the mean square for varieties with that for varieties × places …"

2. ANOVA models

$$\eta_{ij} = \alpha_i + E_j + (\alpha E)_{ij}$$

α = genotype (fixed)

E = environment (random)

$$\eta_{ijk} = \alpha_i + L_j + Y_k + (LY)_{jk} + (\alpha L)_{ij} + (\alpha Y)_{ik} + (\alpha LY)_{ijk}$$

L = location (random)

Y = year (random)

(Greek letters: fixed factors; Latin letters: random factors)

Interlude

(1) Series of experiments virtually always analysed in two stages:

- Analysis of individual trials $\Rightarrow$ means per genotype and environment (η_{ijk})
- Linear mixed model fitted to genotype-environment means (previous slide)

(2) In many ways analogous to network meta-analysis or analysis of clinical trials:

"Environments" $\approx$ "Studies" $\approx$ "Clinics"

(3) "Environments" and "Studies" analogous to "blocks" in incomplete block design

$\Rightarrow$ Recovery of inter-block information
$\approx$ Recovery of inter-environment information
$\approx$ Recovery of inter-study information (Madden et al. 2016)

Difference of two varieties *i* and *h* in a given location *j* and year *k*:

$$\eta_{ijk} - \eta_{hjk} = \alpha_i - \alpha_h + (\alpha L)_{ij} - (\alpha L)_{hj} + (\alpha Y)_{ik} - (\alpha Y)_{hk} + (\alpha LY)_{ijk} - (\alpha LY)_{hjk}$$

Expected value of a variety difference in the TPE:

$$E(\eta_{ijk} - \eta_{hjk}) = \alpha_i - \alpha_h$$

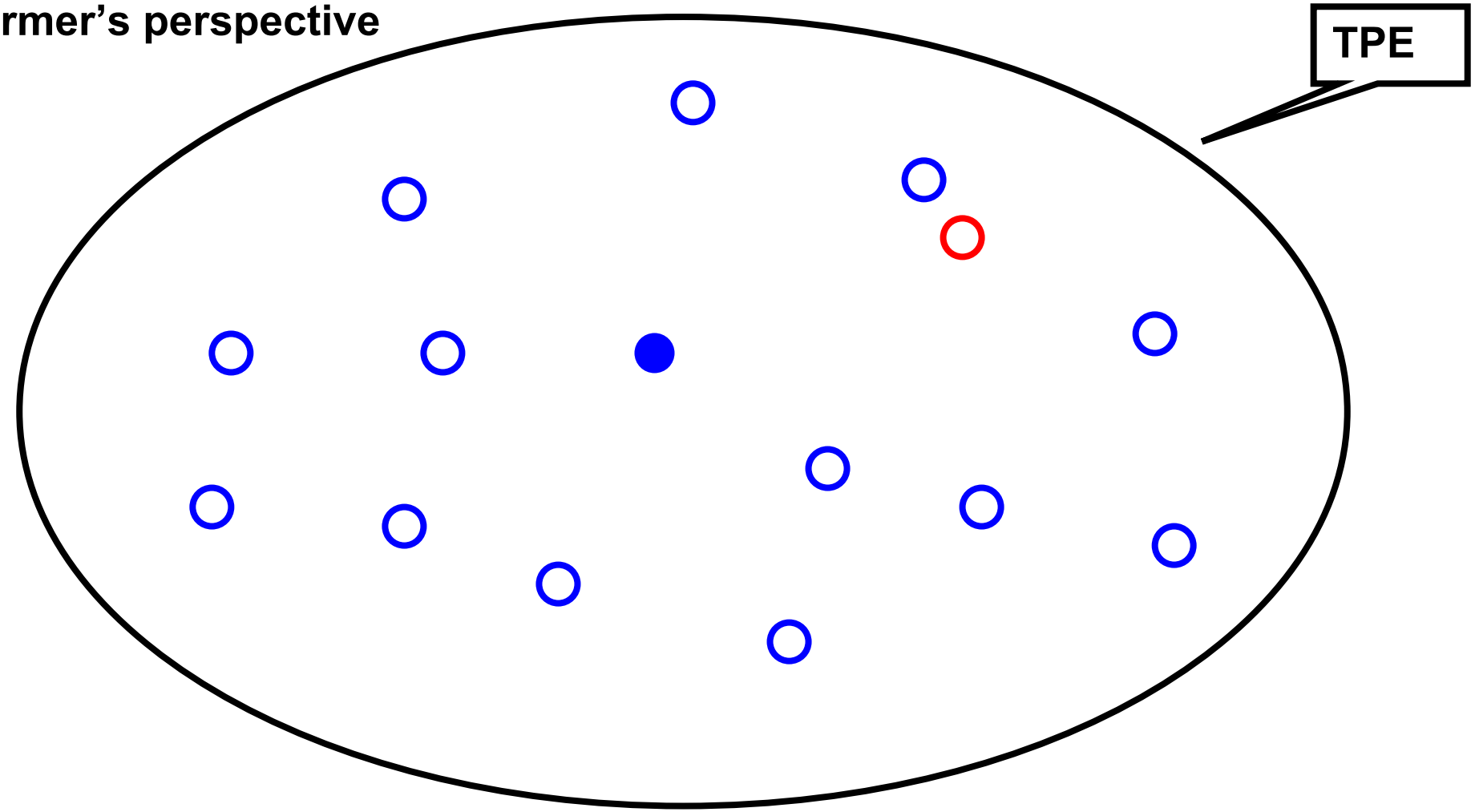
If all random effects are independently and identically distributed (i.i.d.), we find

$P(\eta_{ijk} - \eta_{hjk} > 0) > 0.5$ in a randomly selected environment if $\alpha_i - \alpha_h > 0$

(Eskridge & Mumm 1992).

$\Rightarrow$ rationale for recommending the variety with the largest mean across environments for the whole TPE $\Rightarrow$ **fine from breeder's perspective**

Farmer's perspective



my farm

locations of trial network

"mean" of target population of environments (TPE)

In the best-case scenario, the farmer's location is identical to one of the trial locations:

$$E\left(\eta_{ijk} - \eta_{hjk} \middle| j\right) = \alpha_i - \alpha_h + (\alpha L)_{ij} - (\alpha L)_{hj}$$

$$P\left(\eta_{ijk} - \eta_{hjk} > 0 \middle| j\right) > 0.5 \quad \text{if} \quad \alpha_i - \alpha_h + (\alpha L)_{ij} - (\alpha L)_{hj} > 0$$

$\Rightarrow$ a farmer only needs to identify the variety with the largest long-term mean at his or her location.

But: farmer's location $\neq$ trial locations (even the "closest" one)

The key question then is how to best estimate this mean

Two important general conclusions

(1) For a farmer the estimated variety mean in the TPE is likely to provide a more reliable basis for variety recommendations than the variety mean at the trial location considered to be most similar to the farmer's location

(2) Can judge the relative merits of TPE means versus location means via the Mean Squared Error of a Difference (MSED) because environments (locations, years) are modelled as random

(Piepho & Williams 2024)

3. Random genotypic effects as an option

Two good reasons to model genotypes as random:

(1) Want to do genomic prediction (GBLUP)

(Bernardo 1994; Meuwissen et al. 2001)

(2) Want to borrow strength across zones

(Atlin et al. 2000)

Bottom line:

- Random genotypes is an option
- Random environments is a must

Random-effects model

$$\eta_{ijk} = \mu + a_i + L_j + Y_k + (LY)_{jk} + (aL)_{ij} + (aY)_{ik} + (aLY)_{ijk}$$

a = random genotype

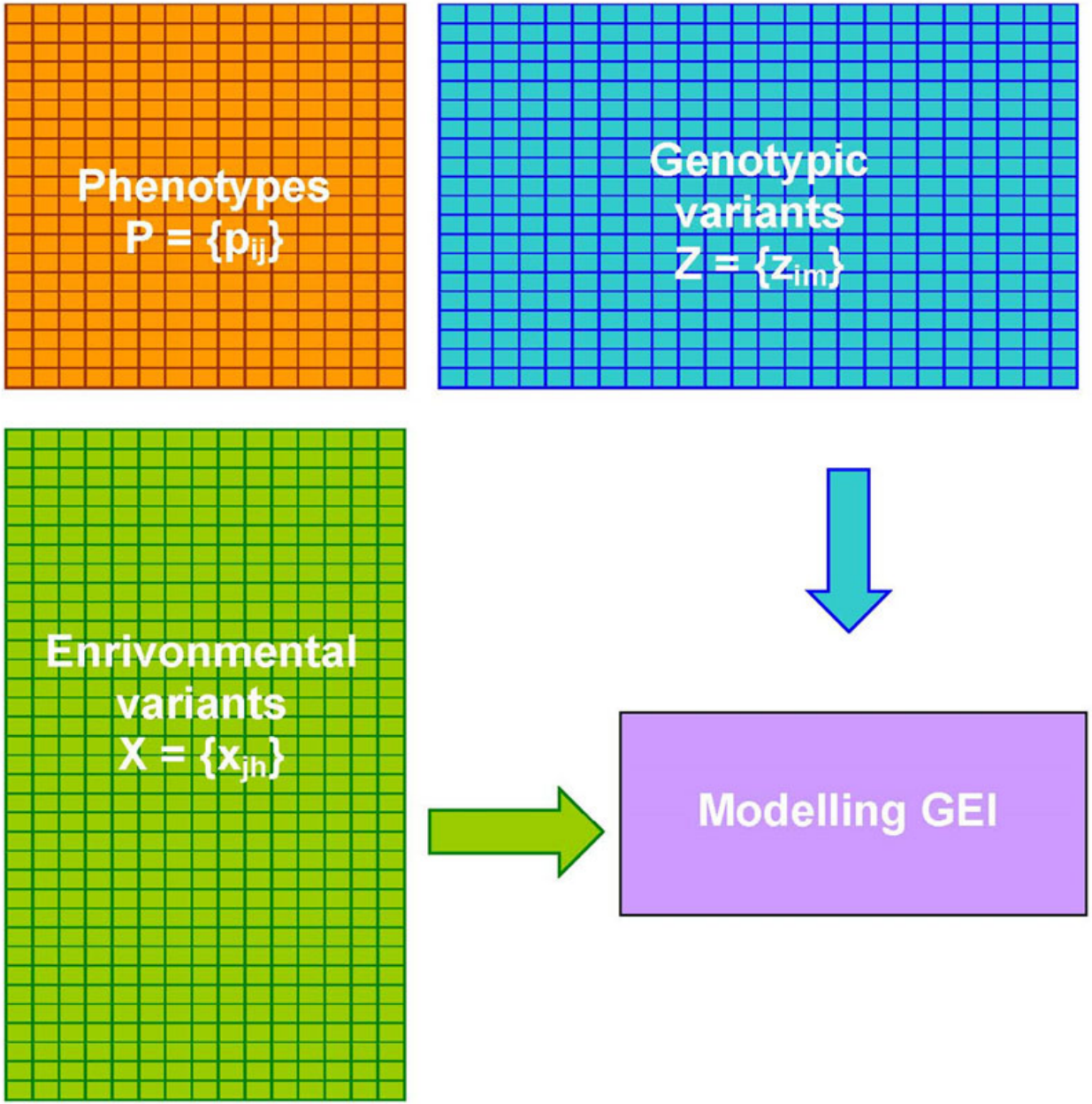
L = random location

Y = random year

Long-term mean of the *i*-th genotype at the *j*-th location:

$$\eta_{ij} = \mu + a_i + L_j + (aL)_{ij}$$

4. Improving predictions using genotypic and environmental covariates



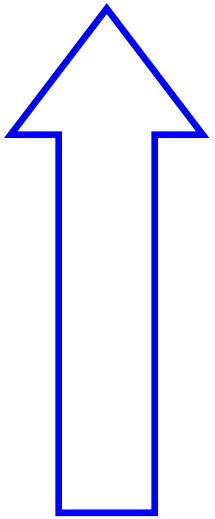
(Piepho 2022)

Two types of models using environmental covariates (EC)

(1) Stratification of environments into environmental types (ET)
(Bustos-Korts et al. 2022)

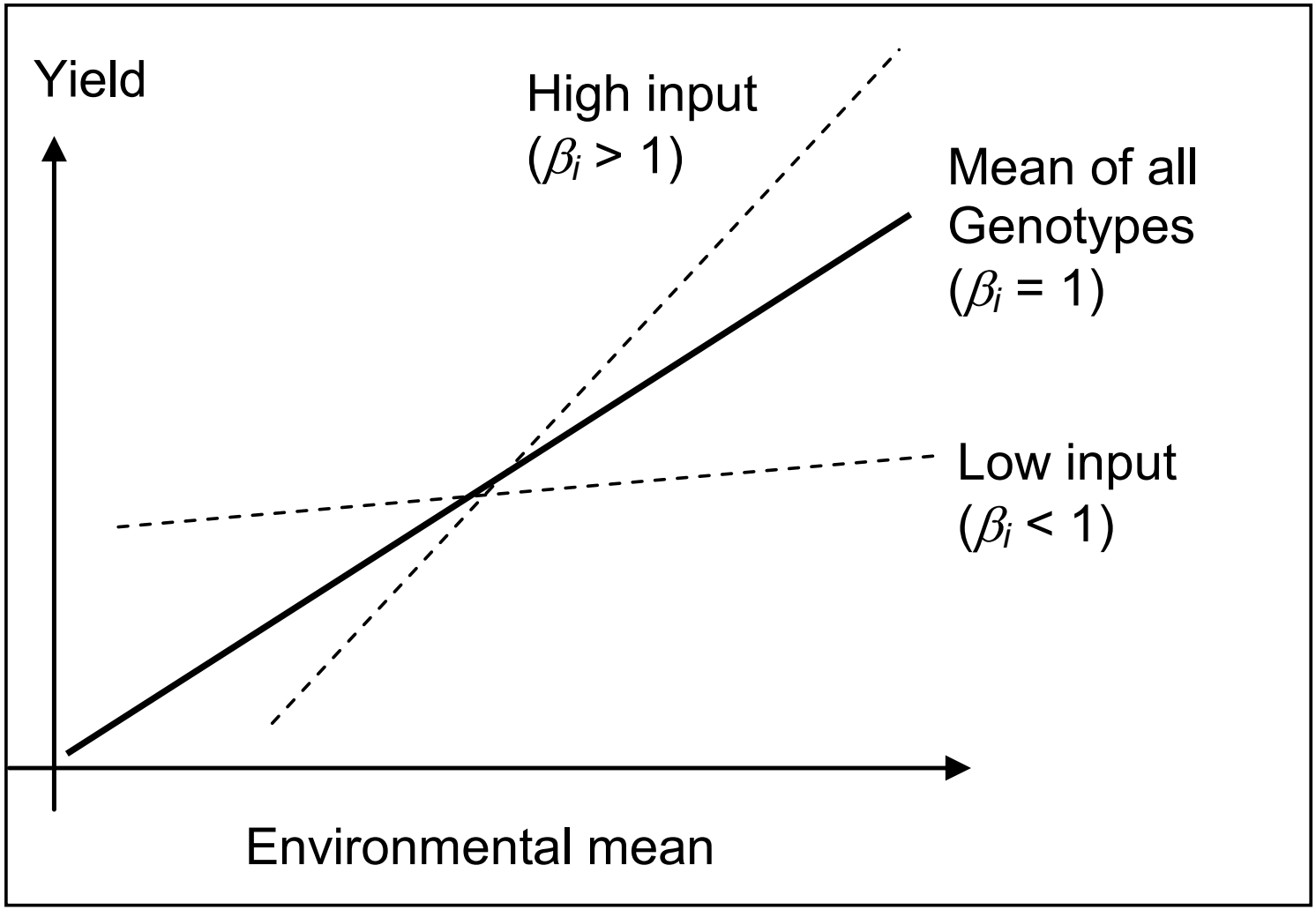
(2) Regression models with EC:

- Factorial regression (Denis 1980)
- Extended Finlay-Wilkinson regression
(Piepho & Blancon 2023)
- Environmental kinship (Jarcquin et al. 2014)



Increasing
Complexity

Finlay-Wilkinson regression



(Yates & Cochran 1938; Finlay & Wilkinson 1963)

Factorial regression

$$\eta_{ij} = \alpha_i + \gamma_{i1} x_{j1} + \gamma_{i2} x_{j2} + \ldots + \gamma_{ip} x_{jp}$$

where

η_{ij} = expected response for the *i*-th genotype (*i* = 1, …, *I*)
in the *j*-th environment (*j* = 1, …, *J*)

α_i = intercept for the *i*-th genotype

γ_{ik} = slope for the *k*-th environmental covariate (EC) for the *i*-th genotype

x_{jk} = value of the *k*-th covariate (*k* = 1, …, *p*) for the *j*-th environment

(Denis, 1988; Piepho & Blancon, 2023)

Now take genotypes as random

$$\alpha_i = \mu_\alpha + a_i$$

$$\gamma_{ik} = \mu_{\gamma k} + c_{ik}$$

$$\eta_{ij} = \left(\mu_\alpha + a_i\right) + \left(\mu_{\gamma 1} + c_{i1}\right)x_{j1} + \left(\mu_{\gamma 2} + c_{i2}\right)x_{j2} + \ldots + \left(\mu_{\gamma p} + c_{ip}\right)x_{jp} \quad ,$$

where

$$\begin{pmatrix} a_i \\ c_i \end{pmatrix} \sim N\left[\begin{pmatrix} 0 \\ 0_p \end{pmatrix}, \Sigma\right]$$

$$c_i = \left(c_{i1}, \ldots, c_{ip}\right)^T$$

Three different choices for Σ

(i) *Unstructured model for* Σ

$\Rightarrow$ random coefficients regression
$\Rightarrow$ translational invariance (Longford, 1993; Buntaran et al., 2021)
$\Rightarrow$ We may denote this approach in our context as *random FR*

(ii) *The kinship approach*

$$\Sigma = \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & I_p \sigma_\gamma^2 \end{pmatrix}$$

(Jarquin et al., 2014)

$\Rightarrow$ not translationally invariant
$\Rightarrow$ more parsimonious than *random FR*
$\Rightarrow$ regularization argument, implies ridge regression (Ruppert et al., 2003, p.66)

(iii) *Reduced rank regression* (RRR)

$$\Sigma = \Lambda\Lambda^T$$

where Λ is a $J \times q$ matrix of factor loadings for q factors and J environments

$\Rightarrow$ translationally invariant
$\Rightarrow$ but can be good approximation to unstructured model

(Tolhurst et al., 2022; Piepho & Blancon 2023)

Using regression models for predictions into new environments

The values of EC may not be known for a new environment

$\Rightarrow$ need a distributional model for the EC:

$$x_{jk} = \mu_x + L_{x(j)} + Y_{x(k)} + (LY)_{x(jk)}$$

where

x_{jk}	= value of the EC in the *j*-th location in the *k*-th year
μ_x	= intercept
$L_{x(j)} \sim N(0, \sigma^2_{x(L)})$	= random main effect for the *j*-th location
$Y_{x(k)} \sim N(0, \sigma^2_{x(Y)})$	= random main effect for the *k*-th year
$(LY)_{x(jk)} \sim N(0, \sigma^2_{x(LY)})$	= random location-year interaction for the *j*-th location and *k*-th year

A single regression term

$\gamma_i x$

γ_i = slope for *i*-th genotype
x = value of EC

Need plug-in value = expected value ξ for x:

$\gamma_i \xi$

Uncertainty associated with the value of $x \Rightarrow$ variance υ_x

Here: initially assume that both γ_i and ξ are known

$\Rightarrow$ contribution to the prediction variance for the response is $\gamma_i^2 \upsilon_x$

Estimating the prediction variance

The overall prediction variance associated with $\gamma_i x$ has two components:

(i) Both γ_i and $\xi = E(x)$ in the product $\gamma_i \xi$ need to be estimated

(ii) Uncertainty about the value x takes in a new environment: $\upsilon_x = \gamma_i^2 \sigma_x^2$

(i) Variance due to the estimation of $\gamma_i \xi$

$$\mathrm{var}\left(\hat{\gamma}_i \hat{\xi}\right) = \gamma_i^2 \,\mathrm{var}\left(\hat{\xi}\right) + \xi^2 \,\mathrm{var}\left(\hat{\gamma}_i\right) + \mathrm{var}\left(\hat{\gamma}_i\right)\mathrm{var}\left(\hat{\xi}\right)$$ (Goodman, 1960)

The naïve plug-in estimator of γ_i^2 is biased

$$\mathrm{var}(\hat{\gamma}_i) = E(\hat{\gamma}_i^2) - [E(\hat{\gamma}_i)]^2 = E(\hat{\gamma}_i^2) - \gamma_i^2$$

$\Rightarrow$

$$E(\hat{\gamma}_i^2) = \gamma_i^2 + \mathrm{var}(\hat{\gamma}_i)$$

$\Rightarrow$

Estimate γ_i^2 by $\widetilde{\gamma}_i^2 = \hat{\gamma}_i^2 - \mathrm{var}(\hat{\gamma}_i)$ and ξ^2 by $\widetilde{\xi}^2 = \hat{\xi}^2 - \mathrm{var}(\hat{\xi})$

$\Rightarrow$

$$est.\mathrm{var}(\hat{\gamma}_i\hat{\xi}) = \hat{\gamma}_i^2\,\mathrm{var}(\hat{\xi}) + \hat{\xi}^2\,\mathrm{var}(\hat{\gamma}_i) - \mathrm{var}(\hat{\gamma}_i)\mathrm{var}(\hat{\xi})$$

(ii) Estimation of $\upsilon_x = \gamma_i^2 \sigma_x^2$

The naïve plug-in estimator $\hat{\upsilon}_x = \hat{\gamma}_i^2 \hat{\sigma}_x^2$ is biased

Bias may be reduced by using the estimator

$$\tilde{\upsilon}_x = \left[\hat{\gamma}_i^2 - \mathrm{var}\left(\hat{\gamma}_i\right)\right]\hat{\sigma}_x^2$$

The **overall prediction variance** is obtained by adding the variances in **(i)** and **(ii)**

Conjecture: the main problem is υ_x

Extension to multiple EC and inclusion of intercept

This is straightforward but will be omitted here

Jointly consider intercept and all slopes:

$$\gamma_i \rightarrow \gamma_i'$$
$$\xi \rightarrow \xi'$$
$$x \rightarrow x'$$

Overall prediction variance

$$\upsilon = \mathrm{var}\left(\hat{\gamma}_i'^T \hat{\xi}'\right) + \upsilon_{x'} + \upsilon_R$$

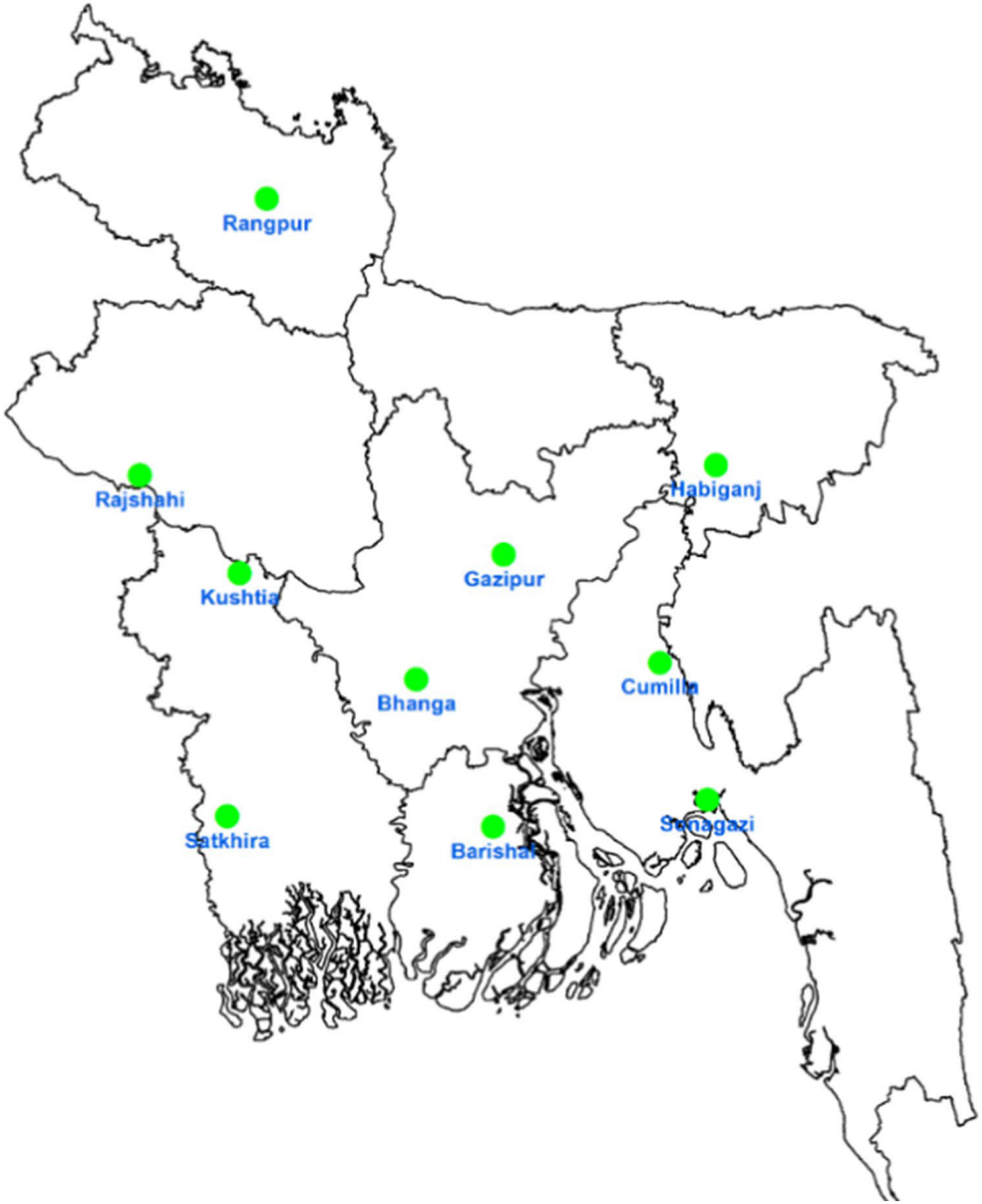
υ_R = variance of deviations from regression

Example: Rice data from Bangladesh (BRRI)

(Maksym Hrachov)

- Stability trials
- Long-term data (1970 to 2020)
- 10 Locations
- 41 rice varieties in the winter season, 45 rice varieties in the monsoon season
- RCBD with three replications at each trial
- Leave-one-environment-out cross-validation &
  Leave-one-year-out cross-validation
- 47 weather and soil EC extracted from public databases (agERA5 and FAO)

(Hrachov et al. 2025)



Locations of Stability Trials with rice in Bangladesh

Table 6. Leave-one-environment-out (LOEO) and leave-one-year-and-location-out (LYLO) cross-validation means (summer rice). PCC – Pearson's correlation coefficient, MSEPD – mean squared error of predicted difference, MSPE – mean squared prediction error, Baseline – model without genotype-covariate interactions, Kinship – model with an environmental relationship matrix, RRR1 and RRR2 – reduced rank regression of rank one and two with observed covariates, RFR – random factorial regression with observed covariates, FW1-US and FW2-US – random factorial regression with one and two synthetic covariates.

Type	Model	Mean PCC		Mean MSEPD		Mean MSPE	
		LOEO	LYLO	LOEO	LYLO	LOEO	LYLO
With the main EC effect	Baseline	0.618	0.589	0.753	0.807	1.03	1.07
	Kinship	0.608	0.591	0.769	0.800	1.03	1.06
	RRR1	0.614	0.582	0.774	0.834	1.04	1.09
	RRR2	0.603	0.580	0.789	0.837	1.05	1.09
	RFR	0.601	0.577	0.791	0.831	1.05	1.07
	FW1-US	0.634	0.592	0.729	0.800	0.864	1.01
	FW2-US	0.638	0.593	0.708	0.800	0.857	0.988
Without the main EC effect	Baseline	0.618	0.589	0.753	0.807	0.868	1.02
	Kinship	0.609	0.585	0.768	0.811	0.875	1.02
	RRR1	0.614	0.583	0.774	0.833	0.882	1.04
	RRR2	0.603	0.580	0.789	0.836	0.895	1.05
	RFR	0.601	0.578	0.790	0.831	0.903	1.04
	FW1-US	0.634	0.592	0.729	0.800	0.864	1.03
	FW2-US	0.638	0.593	0.708	0.800	0.857	1.03

Hitchhiker's Guide to the Galaxy

Ford Prefect & Arthur Dent, stranded on prehistoric Earth and trying to get out:

Arthur Dent and Ford Prefect know exactly where they don't want to be. They don't want to be stranded on prehistoric Earth with a load of unwanted telephone sanitizers and advertising executives who have been thrown off their home planet of Golgafrincham, a world which has subsequently been wiped out by a particularly virulent disease contracted from an unexpectedly dirty telephone. Unfortunately that is precisely where they are. But fortunately they have found a way of coping with their predicament: they are drunk.

. . . . .

ARTHUR: Have you got an answer?

FORD: *No*, but I've got a different name for the problem!

5. Breeding progress and climate change

$$\eta_{ijk} = \alpha_i + L_j + Y_k + (LY)_{jk} + (\alpha L)_{ij} + (\alpha Y)_{ik} + (\alpha LY)_{ijk}$$

α = genotype L = location Y = year

- Take α_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 α_i assess genetic trend

  $\Rightarrow$ Plotted against year in which variety entered trial
- *Adjusted means* for Y_k assess non-genetic trend

  $\Rightarrow$ Plotted against calendar year
- Estimate trend by linear regression based on adjusted means for α_i and Y_k

(Mackay et al. 2011)

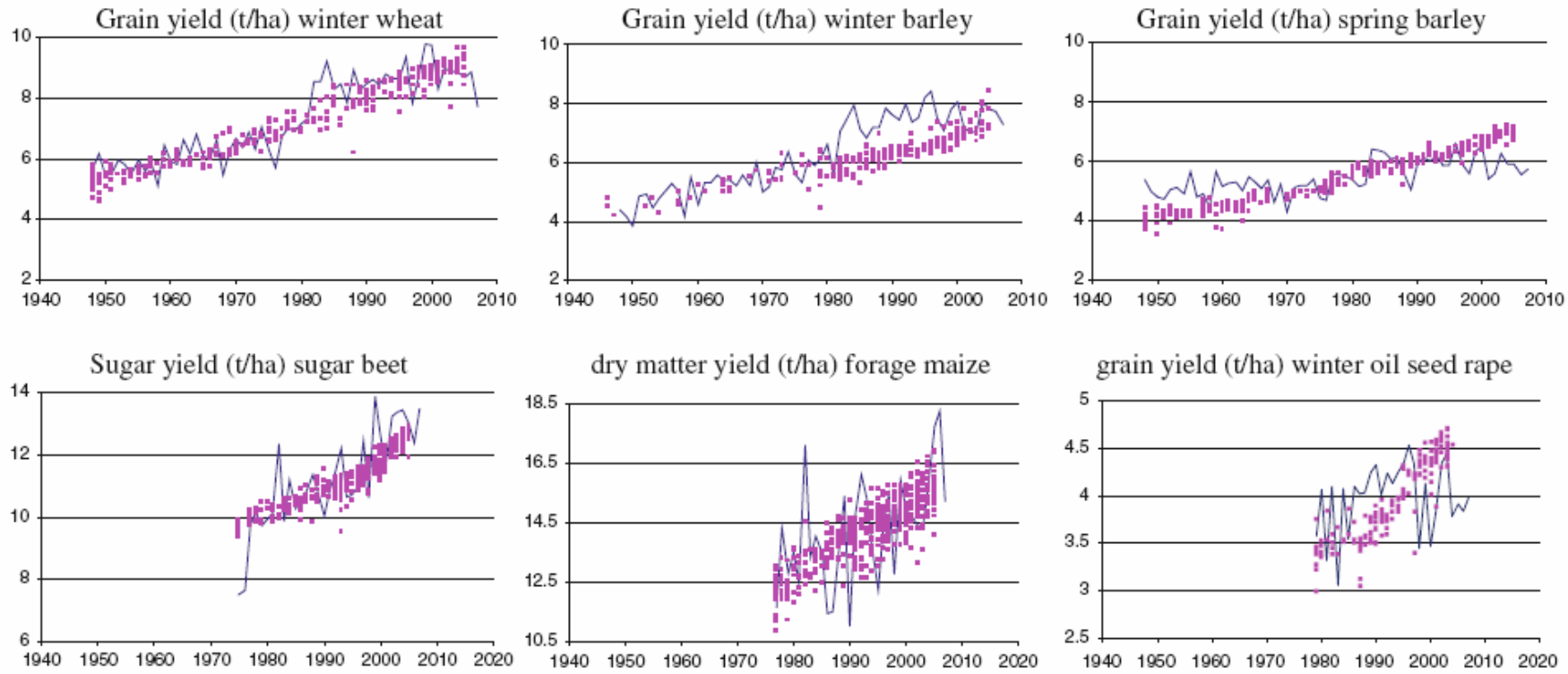


Fig. 1 Trends in variety and year effect for yield (t/ha) from 1948 to 2007. Ordinate and abscissa are on the same scale for all crops except oil seed rape. Variety and year means were estimated as described in "Materials and methods" section. Variety effects (*squares*) are plotted against the year in which the variety first entered the trial. Year means are plotted as a *line*

(Mackay et al., 2011)

— year means

■ variety means

Genetic trend

$$\alpha_i = \beta r_i + H_i$$

β = fixed regression coefficient for genetic trend

r_i = year of first trial for *i*-th variety

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

(Friedrich Laidig)

Non-genetic trend

$$Y_k = \gamma t_k + Z_k$$

γ = fixed regression coefficient for agronomic trend

t_k = calendar year

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

(Laidig et al. 2014)

Graphical displays

Overall trend: frontier line

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

$\Rightarrow$ realized trend if best new variety is planted each year

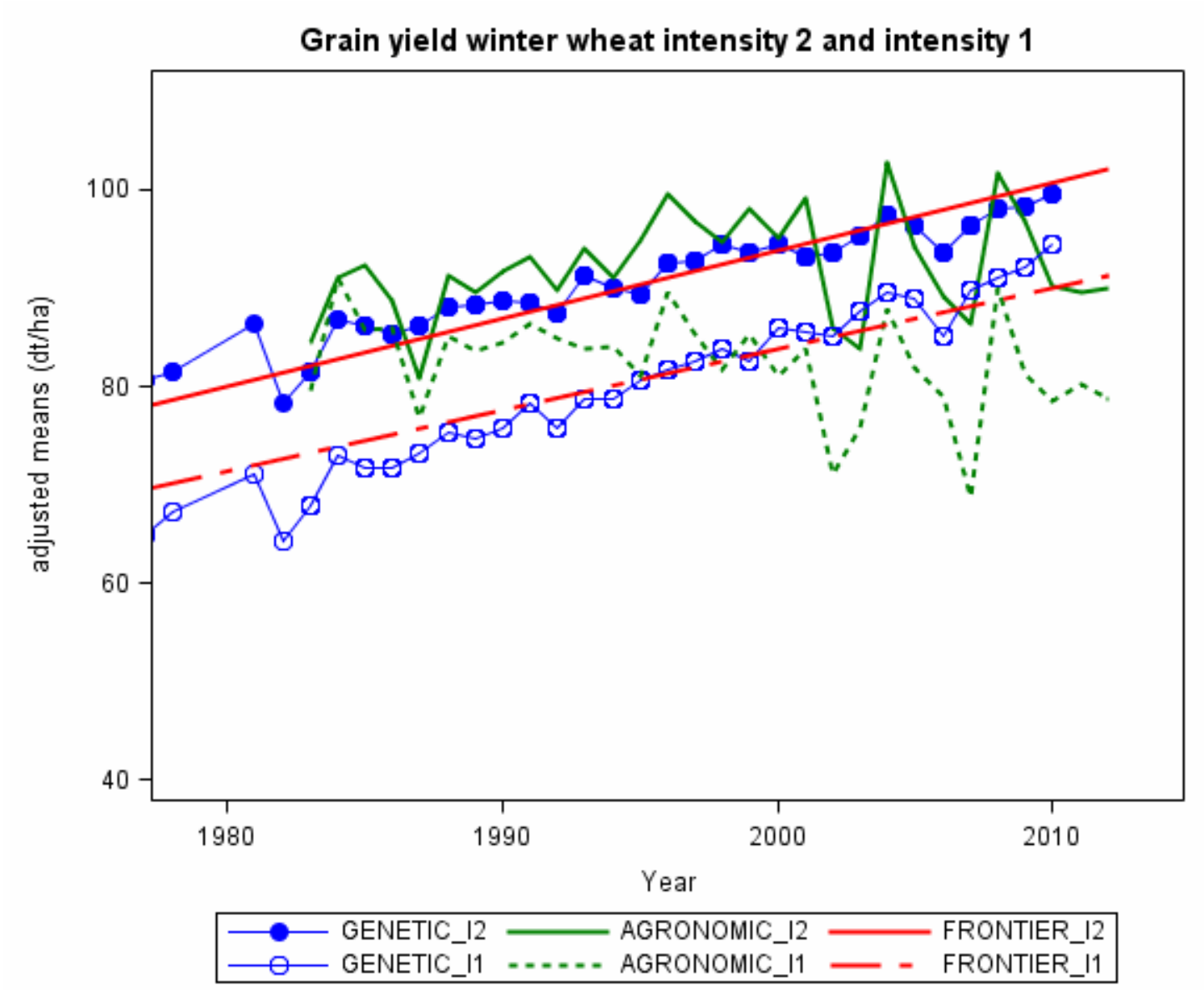


Figure:

GENETIC:
variety group means

AGRONOMIC:
year means

FRONTIER:
frontier line

I1: Intensity 1
I2: Intensity 2

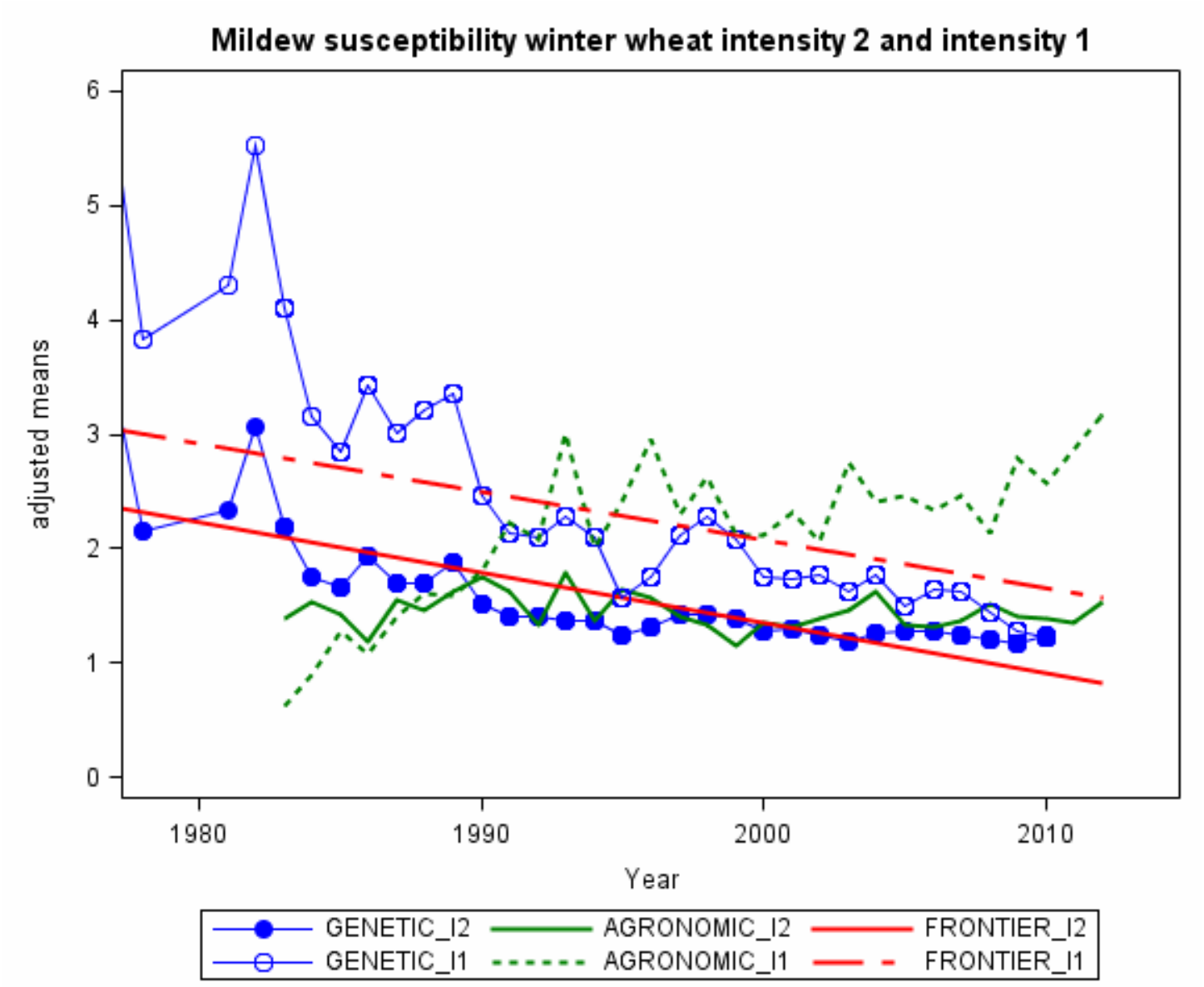

Mildew susceptibility winter wheat intensity 2 and intensity 1
adjusted means
0
1
2
3
4
5
6
1980
1990
2000
2010
Year
GENETIC_I2
AGRONOMIC_I2
FRONTIER_I2
GENETIC_I1
AGRONOMIC_I1
FRONTIER_I1

Table: Estimates of regression coefficients for winter wheat

Trait	Intensity	Slope estimates for regression on			
		Year of first trial (r_i)		Calendar year (t_k)	
		Estimate	s.e.	Estimate	s.e.
Yield (dt/ha)	1	+0.82	0.04	-0.20	0.11
	2	+0.53	0.04	+0.16	0.12
Mildew (1-9)	1	-0.095	0.006	+0.052	0.0092
	2	-0.042	0.003	-0.002	0.0048

Contents lists available at ScienceDirect

Journal of Cleaner Production

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

Breeding progress reduces carbon footprints of wheat and rye

L. Riedesel [a], F. Laidig [b], S. Hadasch [b], D. Rentel [c], B. Hackauf [d], H.-P. Piepho [b], T. Feike [a,*]

[a] Institute for Strategies and Technology Assessment, Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Stahnsdorfer Damm 81, 14532 Kleinmachnow, Germany
[b] Institute of Crop Science, Biostatistics Unit, University of Hohenheim, Fruwirthstrasse 23, 70599 Stuttgart, Germany
[c] Bundessortenamt, Osterfelddamm 80, 30627 Hannover, Germany
[d] Institute for Breeding Research on Agricultural Crops, Julius Kühn Institute (JKI) – Federal Research Centre for Cultivated Plants, Rudolf-Schick-Platz 3a, 18190 Groß Lüsewitz, Germany

$$\eta_{ijk} = \alpha_i + L_j + Y_k + (LY)_{jk} + (\alpha L)_{ij} + (\alpha Y)_{ik} + (\alpha LY)_{ijk}$$

+ genetic and non-genetic trends

+ nitrogen fertilizer rate

(Ludwig Riedesel)

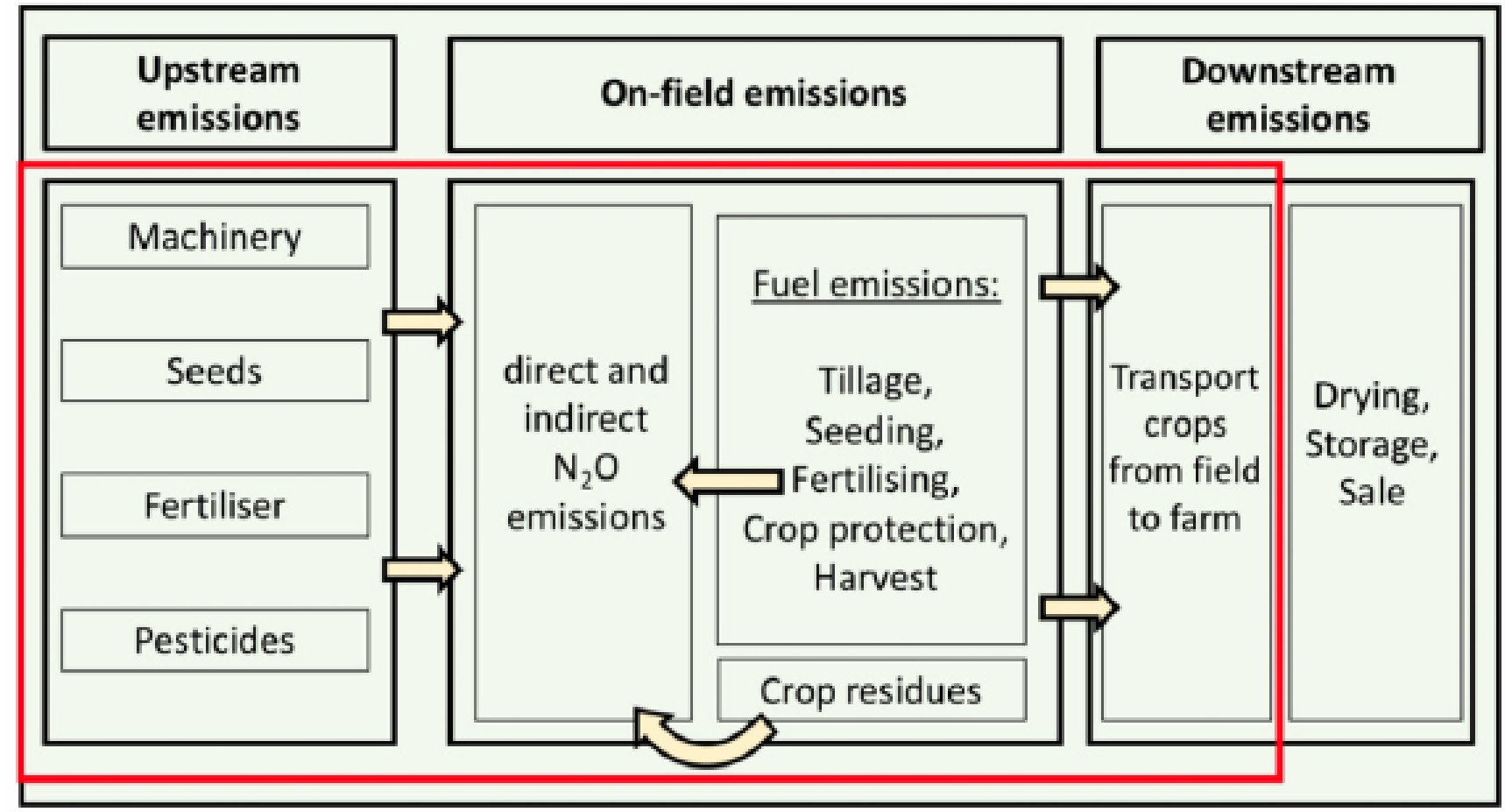


Fig. 2. System boundaries of the partial life cycle assessment of wheat and rye production. All items within the red frame are considered in this study.

(Riedesel et al. 2022)

Life cycle assessment (LCA)

$$\text{Total CFP } \left(\text{kg } CO_2\text{e kg}^{-1}\right) = \frac{Total\ GHGL}{Yield} \left(\frac{kg\ CO_2e\ ha^{-1}}{MT\ ha^{-1}}\right)$$

CFP = Carbon footprint

GHGL = Greenhouse gas load

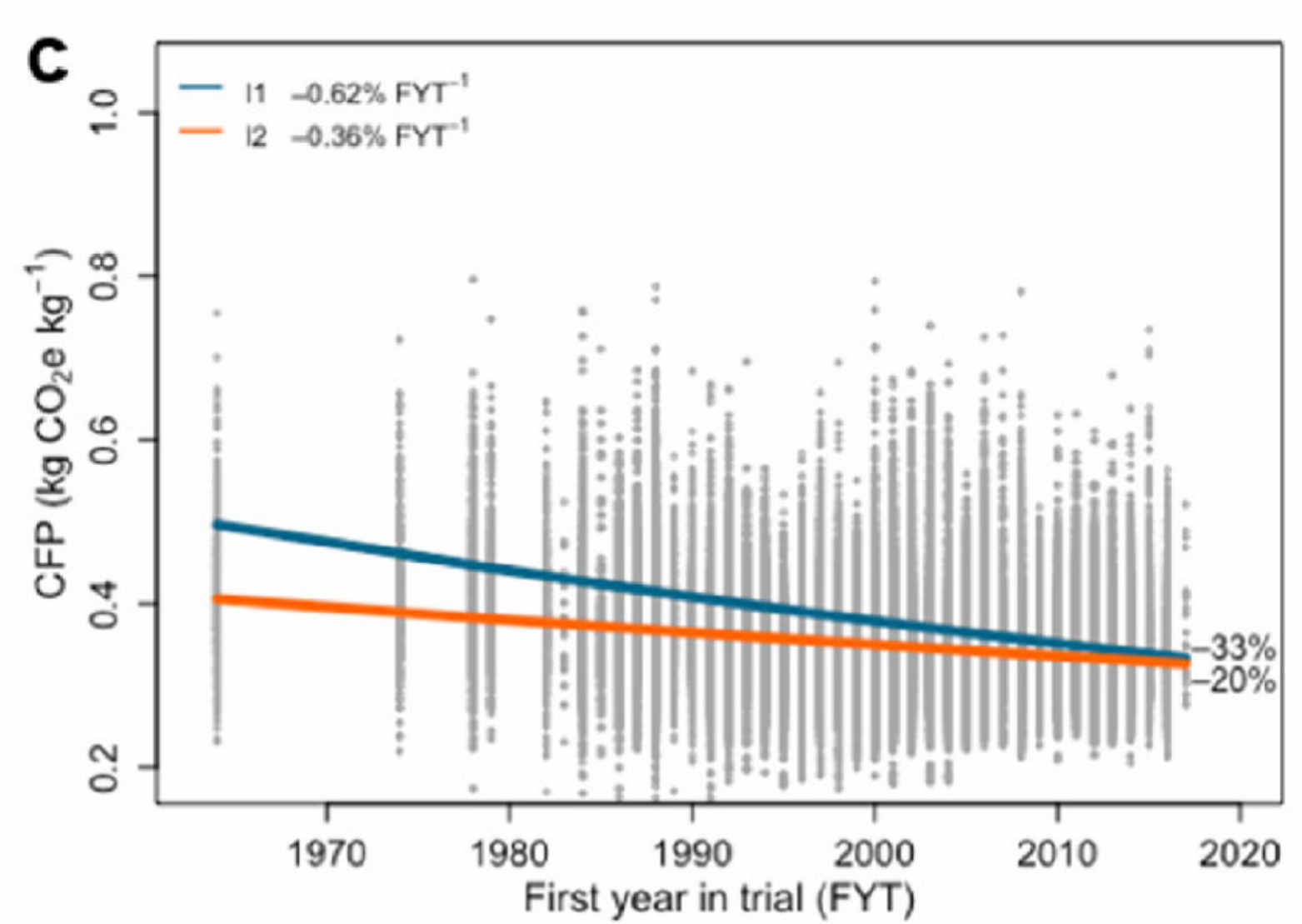


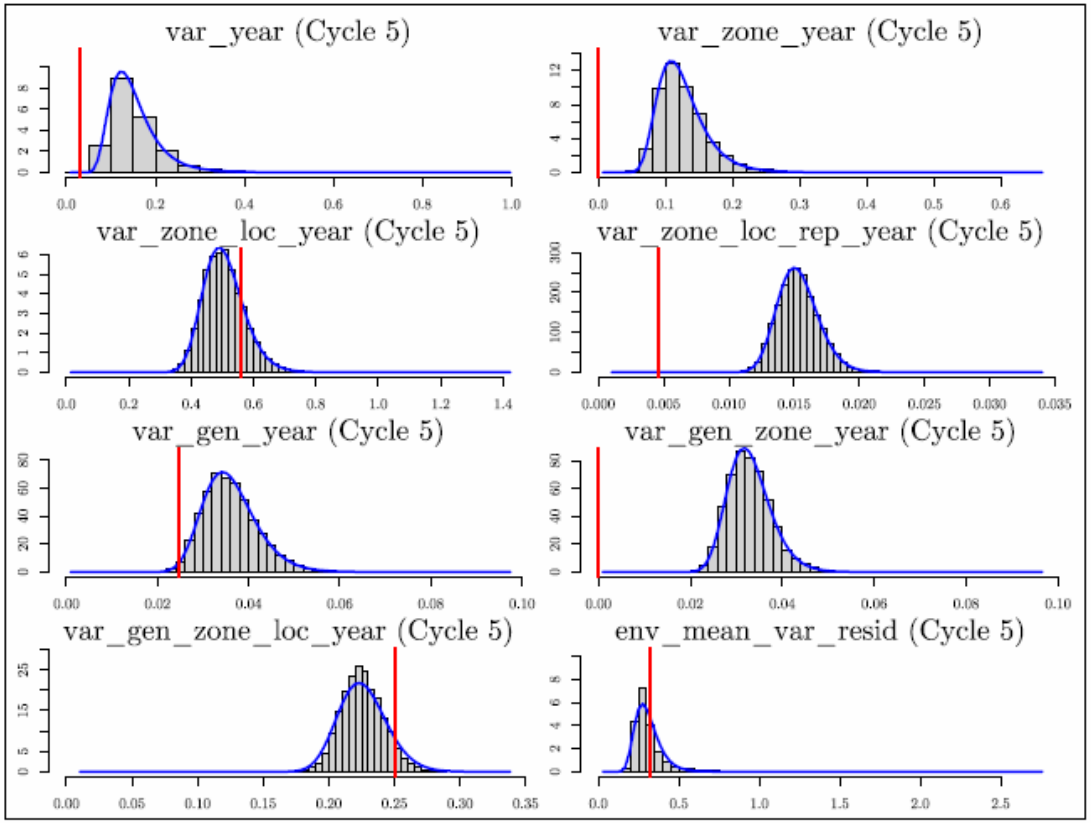
Figure: Breeding progress for carbon footprint (CFP) winter wheat

(Riedesel et al. 2022)

Der Preisträger des Zukunftspreises der GPW 2025, Dr. Ludwig Riedesel (Mitte), mit Prof. Michael Frei (links) und Prof. Johannes Isselstein (rechts).

6. Bayesian Updating

- Variety testing and plant breeding programs have decades of long-term data
- This is plenty of information to objectively inform priors for variance parameters
- Field trials organized by cycles ⇒ can update priors by cycles



(Stephan Bark 2025, Master thesis)

References

Bark, S. (2025): Optimal design for multi-environment trials in plant breeding: A mixed model and Bayesian framework. Master thesis, LMU München.

Hrachov, M., Piepho, H.P., Rahman, N.F., Malik, W. (2025): Regression approaches for modelling genotype-environment interaction and making predictions into a target population of environments. https://doi.org/10.48550/arXiv.2507.18125

Laidig, F., Piepho, H.P., 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, H.P., Blancon, J. (2023): Extending Finlay-Wilkinson regression with environmental covariates. *Plant Breeding* **142**, 621-631. https://doi.org/10.1111/pbr.13130

Piepho, H.P., Williams, E.R. (2024): Factor-analytic variance-covariance structures for prediction into a target population of environments. *Biometrical Journal* **66**, e202400008. https://doi.org/10.1002/bimj.202400008

Riedesel, L., Laidig, F., Hadasch, S., Rentel, D., Hackauf, B., Piepho, H.P., Feike, T. (2022): Breeding progress reduces carbon footprints of wheat and rye. *Journal of Cleaner Production* **377**, 134326.