An empirical comparison of systematic and randomized field experimental designs

Hans-Peter Piepho

Biostatistics Unit
University of Hohenheim
Germany

1818

One key paper published 100 years ago

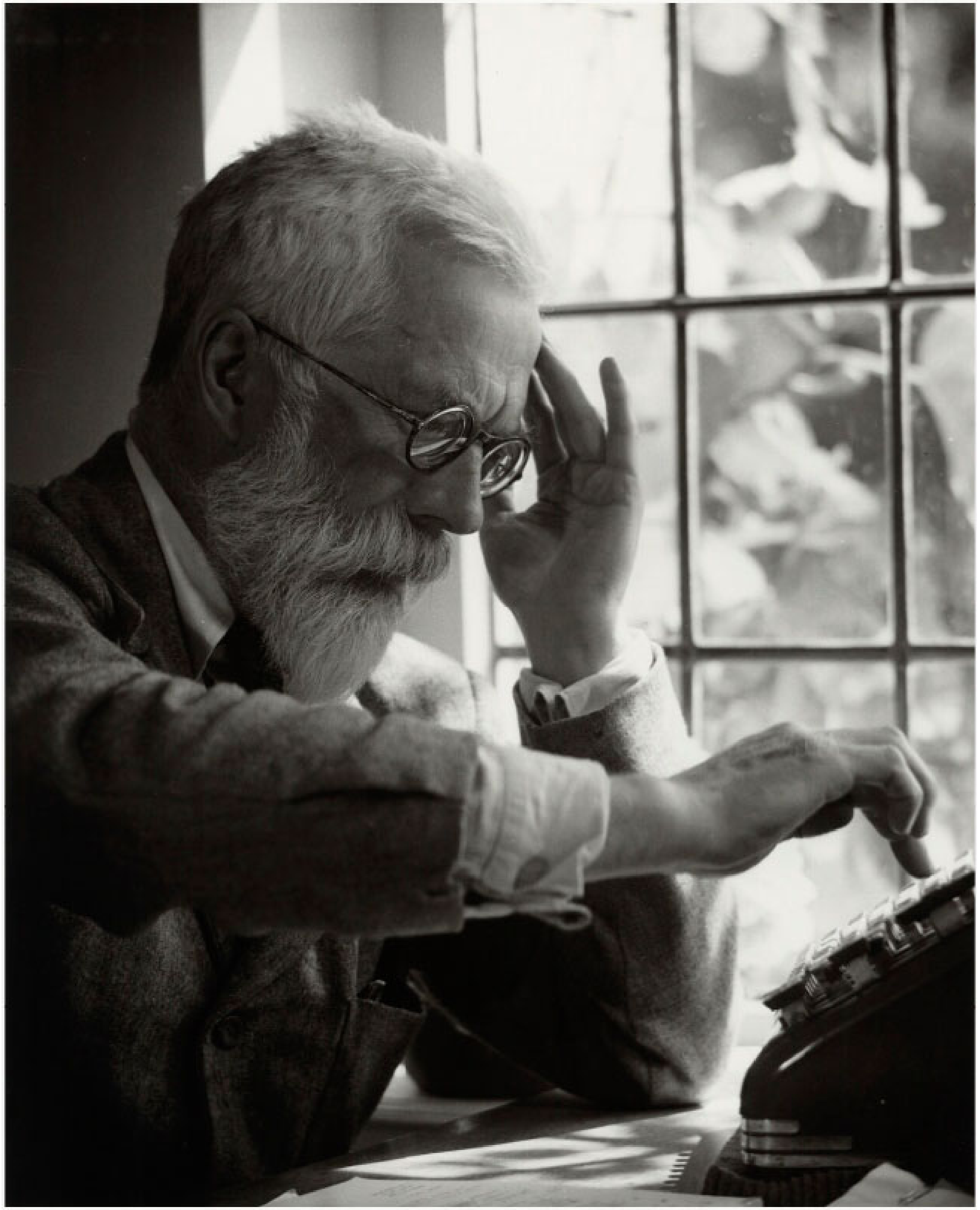
THE ARRANGEMENT OF FIELD EXPERIMENTS

R. A. Fisher, Sc.D.,

Rothamsted Experimental Station.

The Present Position.—The present position of the art of field experimentation is one of rather special interest. For more than fifteen years the attention of agriculturalists has been turned to the *errors* of field experiments. During this

[Fisher (1926) *Journal of the Ministry of Agriculture of Great Britain* **33**, 503–513.]



R.A. Fisher’s three principles of experimental design

- Replication
- Randomization
- Blocking (local error control)

Fisher and his ‘Millionaire’ calculator

Systematic designs are common in on-farm experimentation

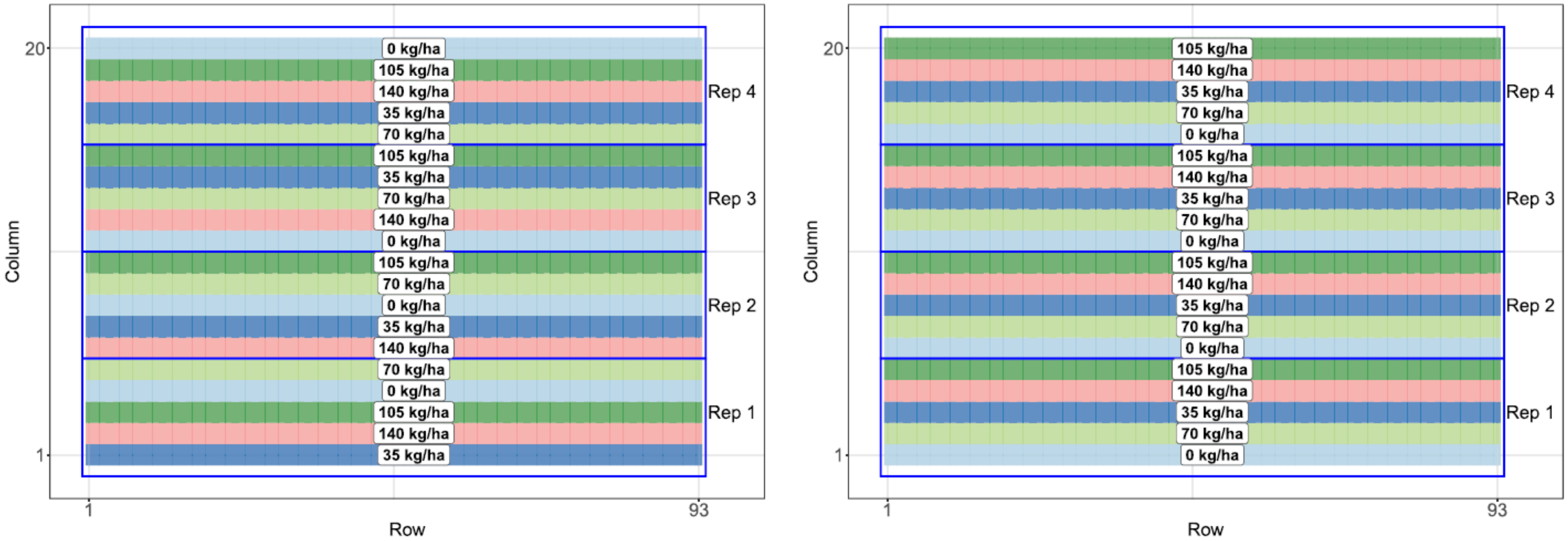
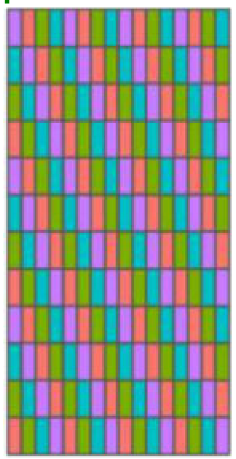


Fig. 1. The nitrogen treatments with five levels (0, 35, 70, 105 and 140 kg/ha) randomly (*left*) and systematically (*right*) allocated into strips in each replicate block.

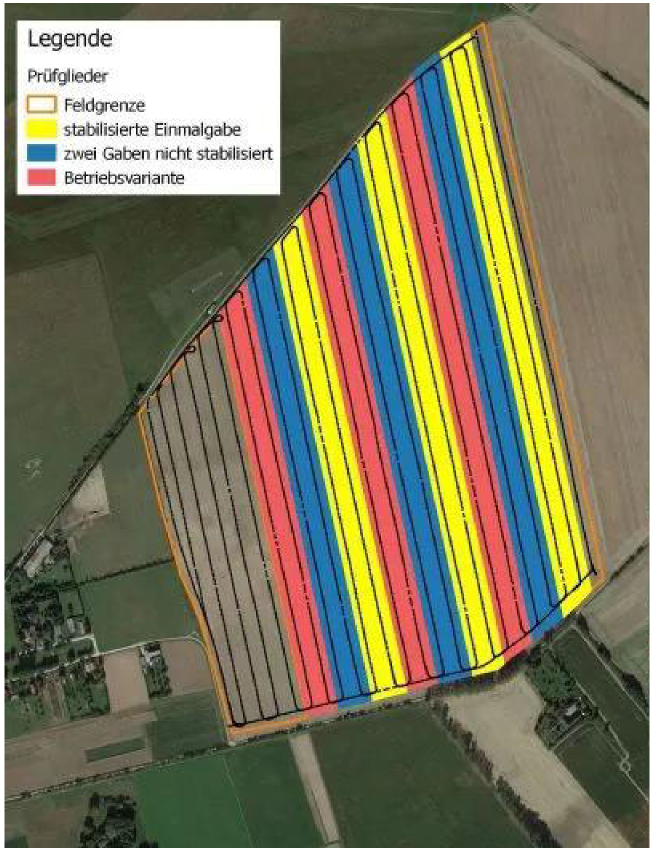
(Cao *et al*. 2024 Optimal design for on-farm strip trials—systematic or randomised?)



“Systematic treatment assignments are more favourable than random treatment assignment.“

(Alesso *et al*. 2021 Design of on-farm precision experiments to estimate site-specific crop responses)

Systematic designs are common in on-farm experimentation



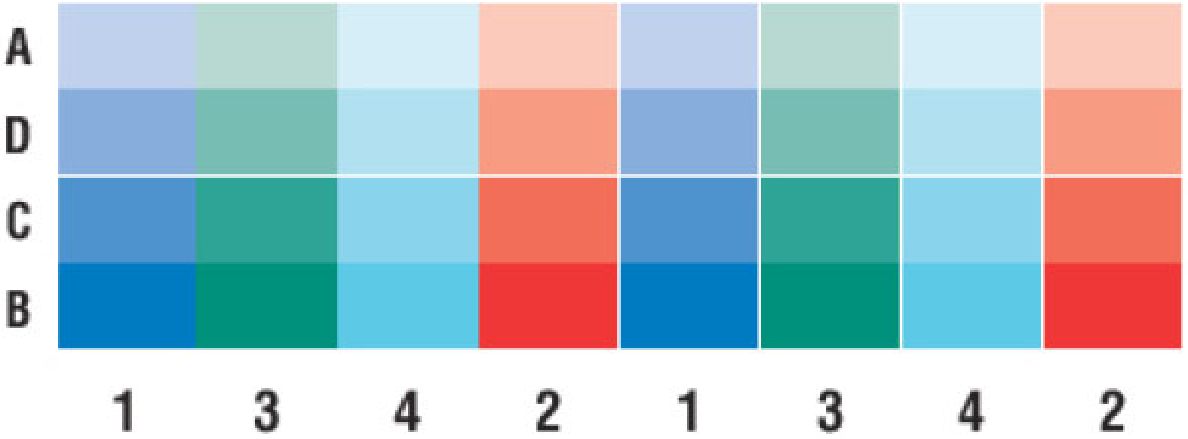


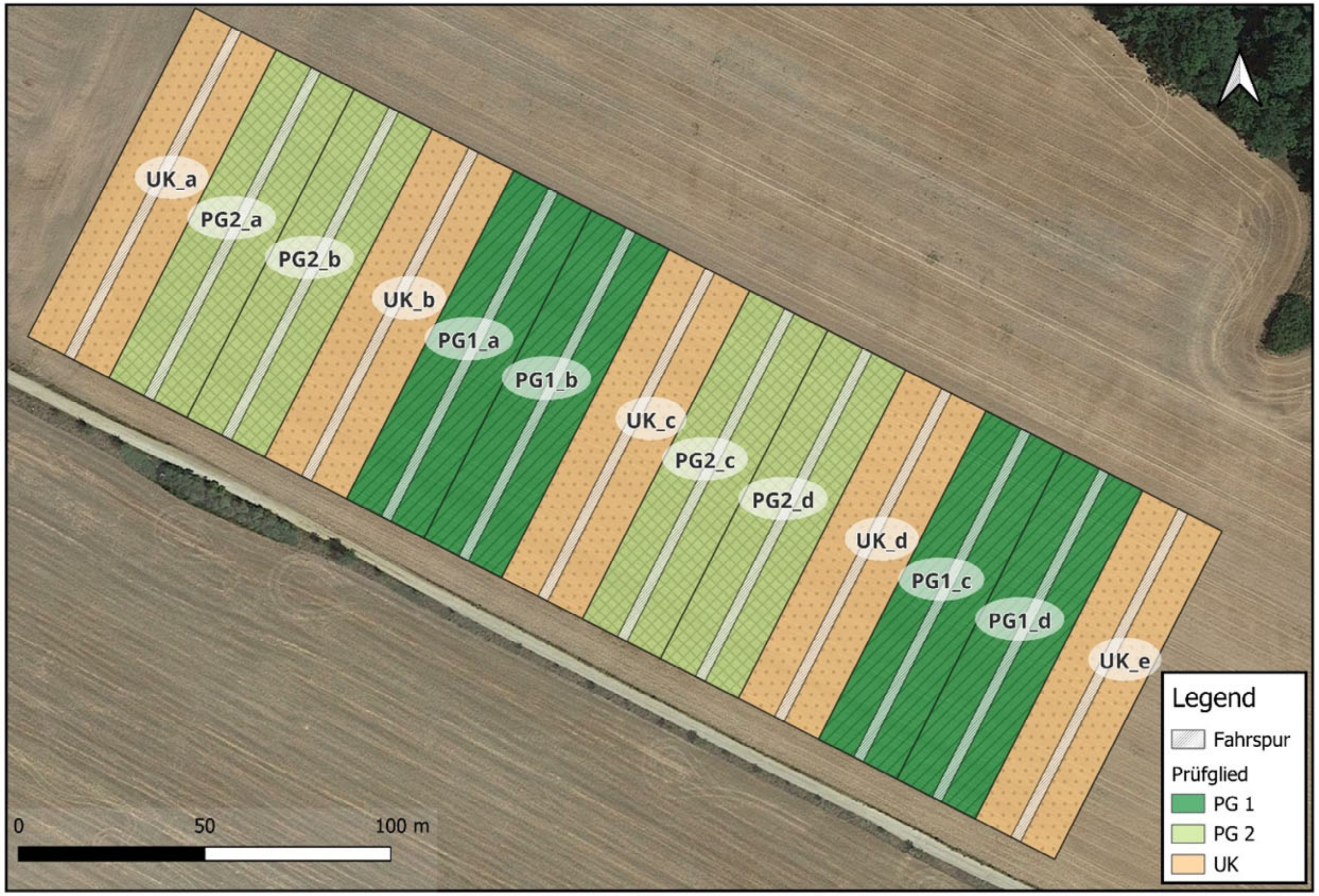
Figure 6: Replicated two way strip design for an experiment with two variables, each with four levels (denoted 1, 2, 3, 4 for one variable and A, B, C, D for the other). Note that the experiment has 16 (4 x 4) treatments, replicated twice.

(GRDC 2021 Designing your own on-farm experiments)

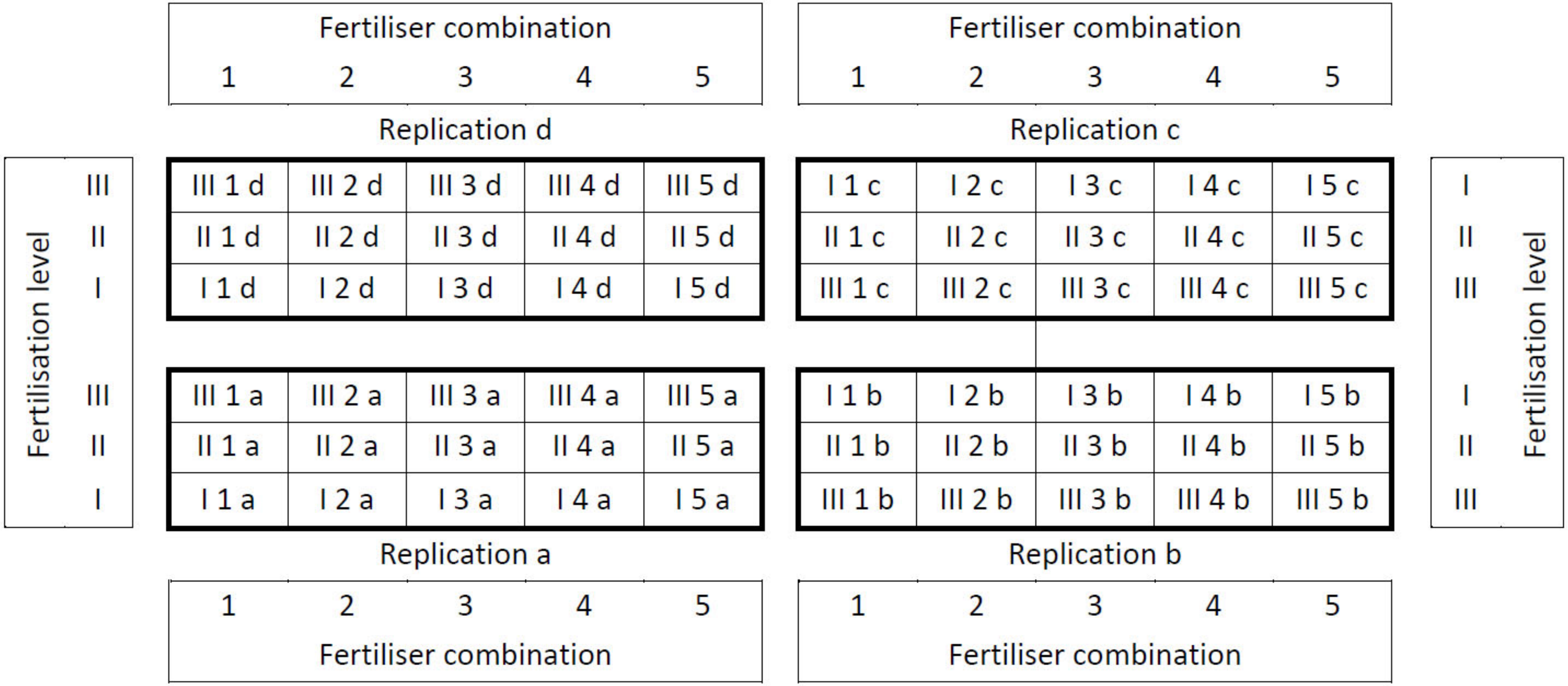
https://grdc.com.au/resources-and-publications/all-publications/publications/2021/designing-your-own-on-farm-experiments

(Büchse 2025 Example of a commercial on-farm fertilizer experiment)

Großhettstedt



(Perner 2023/24 OFE with two insecticide treatments PG1 & 2, UK = control)



Crop rotation: Sugar beet – Winter wheat – Summer barley

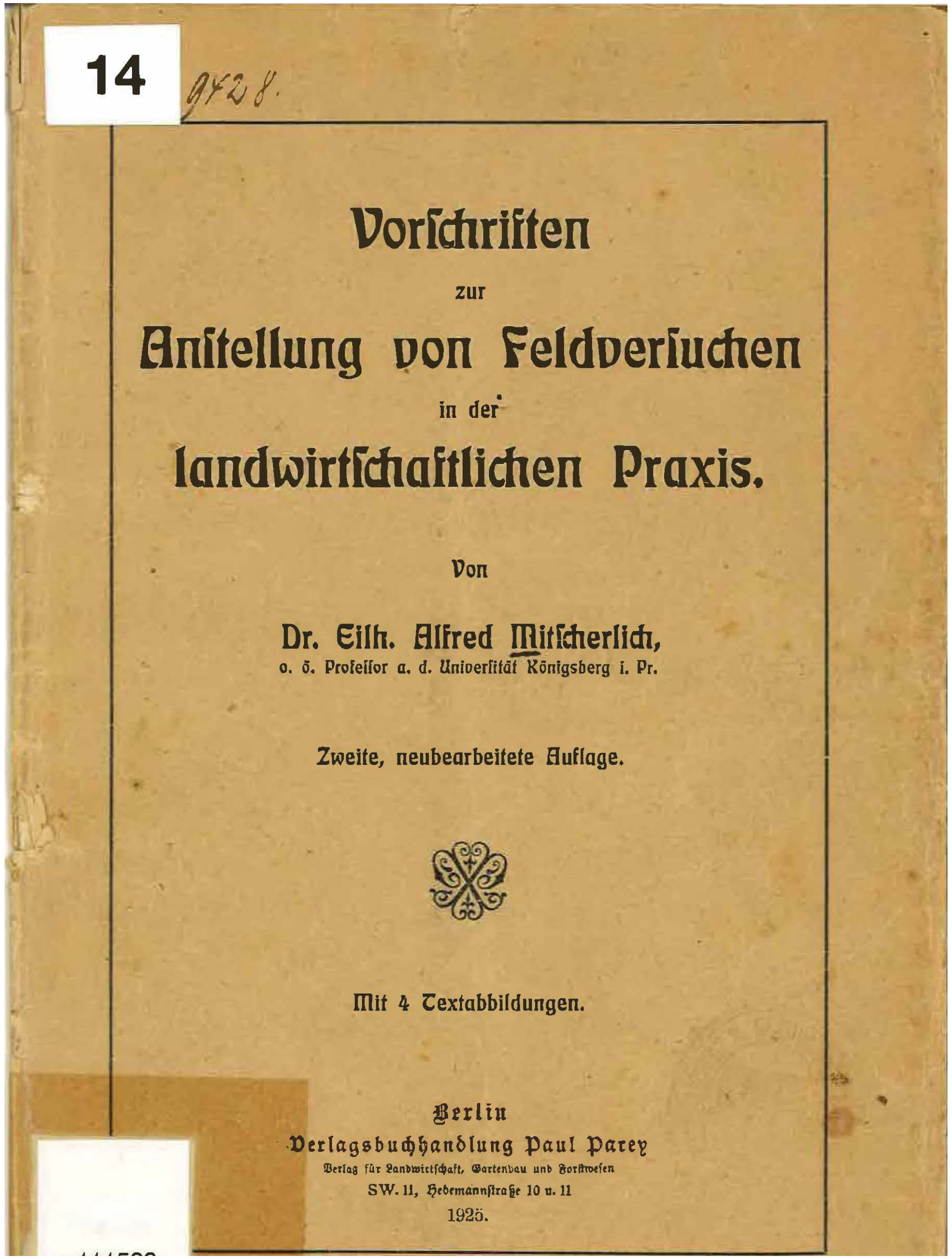
Nutrient depletion experiment in Giessen (1956–2016) (Macholdt et al. 2019)

Quelle: Deutsche Fotothek

Eilhard Alfred Mitscherlich

German agronomist and soil scientist

* 29 August 1874 in Berlin
† 3 February 1956 in Paulinenaue



Instructions

for the

conduct of field experiments

in

agricultural practice

First edition: 1919
(Fisher starts work at Rothamsted)

Second edition: 1925
(Fisher's SMRW first published)

Will man, um einen mehr feldmäßigen Bestand zu erzielen, diese Schutzstreifen vermeiden, so empfiehlt es sich vielleicht, 7 Drillreihen in

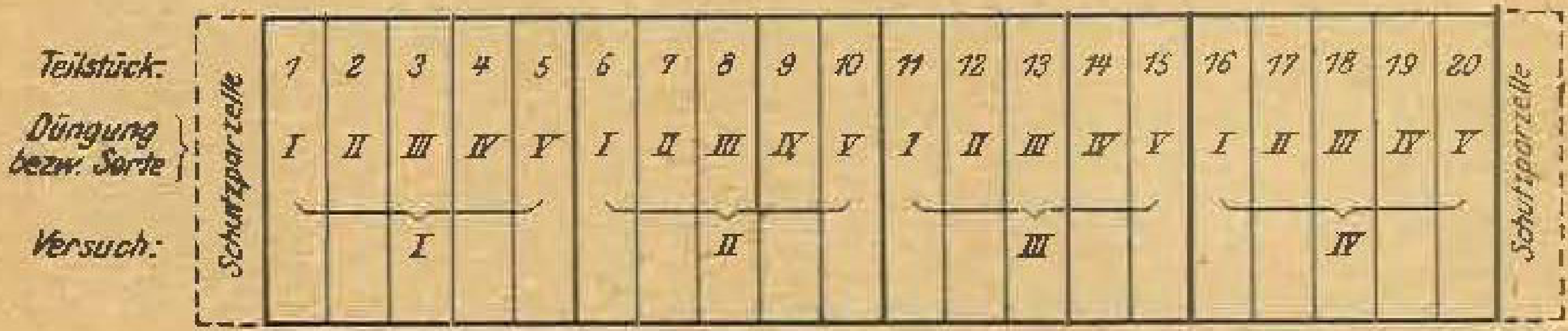


Abb. 1.

20 cm Entfernung anzulegen und bei 5 Sorten diese Streifen 7 × 20 × 5 = 7 m lang zu nehmen. Bei der Ernte müßten aber dann, um die ungleichartige Beschattung durch die Nachbarteilstücke auszuschalten, je die beiden äußeren Drillreihen nicht mit berücksichtigt werden, so daß dann bei einem Versuche nur eine Fläche von 7 × 1,00 m = 7 qm abgeerntet wird.

(Mitscherlich 1919, 1925, p.10)

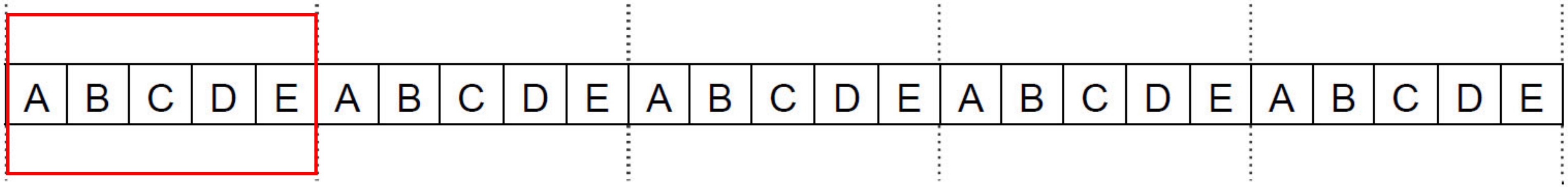
MITSCHERLICH-Design

A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E

Figure: A systematic design with five treatments (A-E) and five replicates

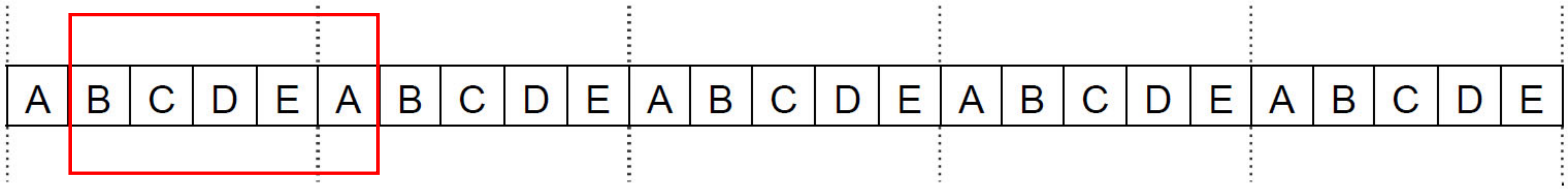
(von Lochow & Schuster 1961)

MITSCHERLICH-Design



A “sliding block”

MITSCHERLICH-Design



A “sliding block”

⇔ Nearest neighbour analysis (Papadakis 1937; Wilkinson et al. 1983)

A geostatistical model for MITSCHERLICH'S slide method

$$y_k = \mu + \sum_{j \in M_k} b_j + \tau_{i(k)} + e_k$$

b_j = Effect of j-th sliding block

M_k = Set of all indices j for sliding blocks overlapping k-th plot

$i(k)$ = Index of treatment allocated to k-th plot $(i = 1, \ldots, v)$

$\tau_{i(k)}$ = Treatment effect

e_k = Plot error

(Piepho & Vo-Thanh 2020)

Variances and covariances

- Complete blocks with v treatments
- Every plot is overlapped by v blocks

$$b_j \sim N\left(0, \sigma_b^2\right)$$
$$e_k \sim N\left(0, \sigma_e^2\right)$$

$$\operatorname{var}\left(y_k\right) = v\sigma_b^2 + \sigma_e^2$$

$$\operatorname{cov}\left(y_k, y_{k'}\right) = h_{k,k'}\sigma_b^2 \quad , \quad \left(k \neq k'\right)$$

where $h_{k,k'} = \max\left(0, v - \left|k - k'\right|\right)$

$\Rightarrow$ Linear Variance (LV) model (similar to but not the same as Williams 1986)

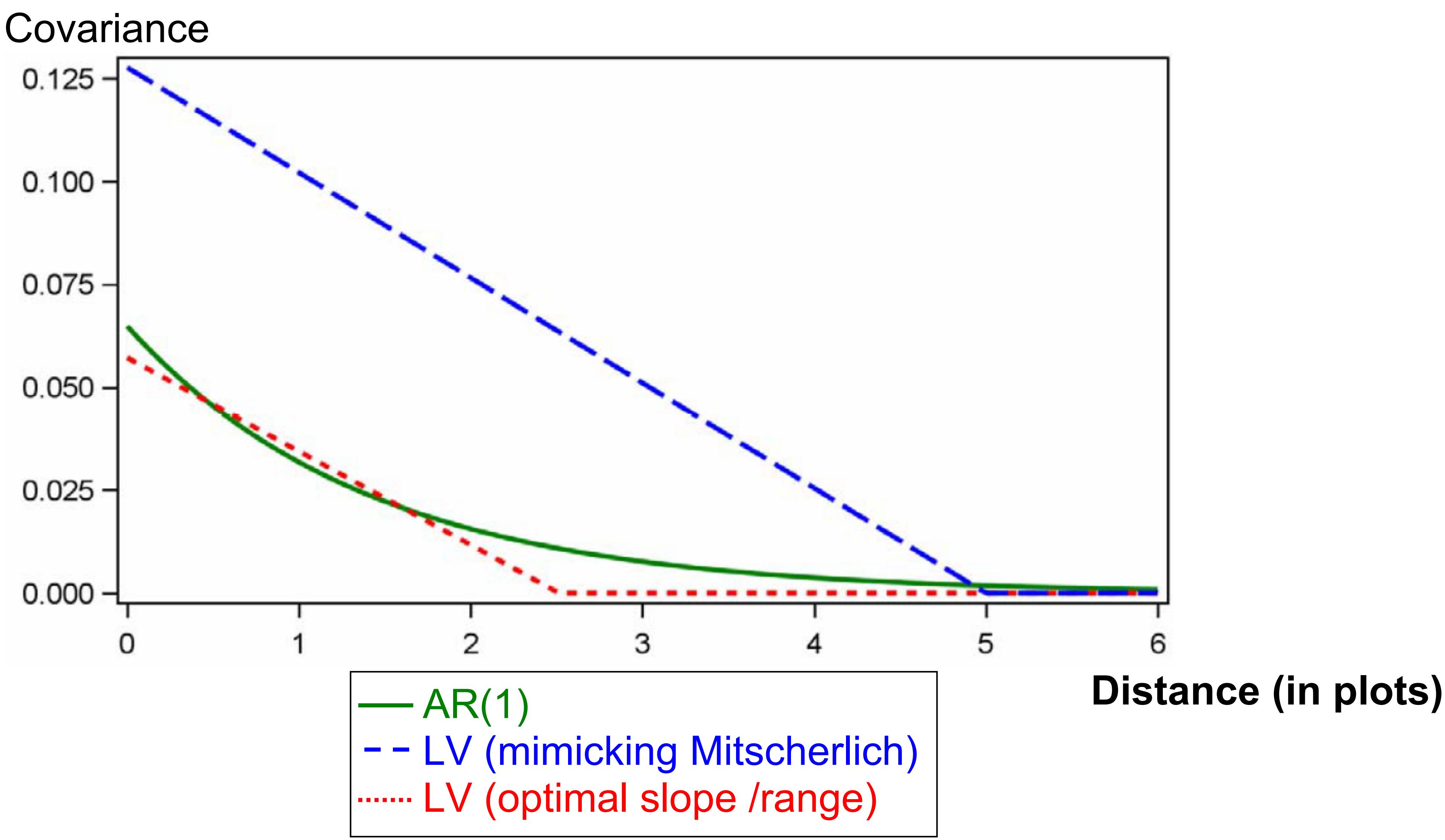


Figure: Fitted spatial models (data from von Lochow & Schuster 1961)

$\Rightarrow$ MITSCHERLICH design is A-optimal for LV (Piepho & Vo-Thanh 2020)

Design generated for 2D extension of LV model for v = 16 treatments

6	4	15	7	6	9	15	12	13	14	15	12
10	14	3	16	10	14	3	16	10	1	3	16
13	9	2	8	13	4	11	7	5	9	2	8
5	1	11	12	5	1	2	8	6	4	11	7
6	4	15	7	6	9	15	12	13	14	15	12
3	14	16	10	3	14	16	10	3	1	16	10
13	9	2	8	13	4	11	7	5	9	2	8
5	1	11	12	5	1	2	8	6	4	11	7

(Piepho & Vo-Thanh 2020)

Design generated for 2D extension of LV model for v = 16 treatments

6	4	15	7	6	9	15	12	13	14	15	12
10	14	3	16	10	14	3	16	10	1	3	16
13	9	2	8	13	4	11	7	5	9	2	8
5	1	11	12	5	1	2	8	6	4	11	7
6	4	15	7	6	9	15	12	13	14	15	12
3	14	16	10	3	14	16	10	3	1	16	10
13	9	2	8	13	4	11	7	5	9	2	8
5	1	11	12	5	1	2	8	6	4	11	7

(Piepho & Vo-Thanh 2020)

Fisher strongly advised against systematic designs

To sum up: systematic arrangements should be avoided, since with these it is usually possible to estimate the experimental error in several different ways, giving widely different results, each way depending on some one set of assumptions as to the distribution of natural fertility, which may or may not be justified.

(Fisher 1925; 1973, Statistical Methods for Research Workers, p. 271).

Designs evaluated using wheat uniformity trial data of Mercer & Hall (1911)

CRD Completely randomized design

RCBD Randomized complete block design

SYS A B C D E | E D C B A | A B C D E | E D C B A
(Fisher 1925)

MIT A B C D E | A B C D E | A B C D E | A B C D E
(Mitscherlich 1919)

COV RCBD with positional covariate within blocks
(Fisher 1925; Piepho *et al*. 2026)

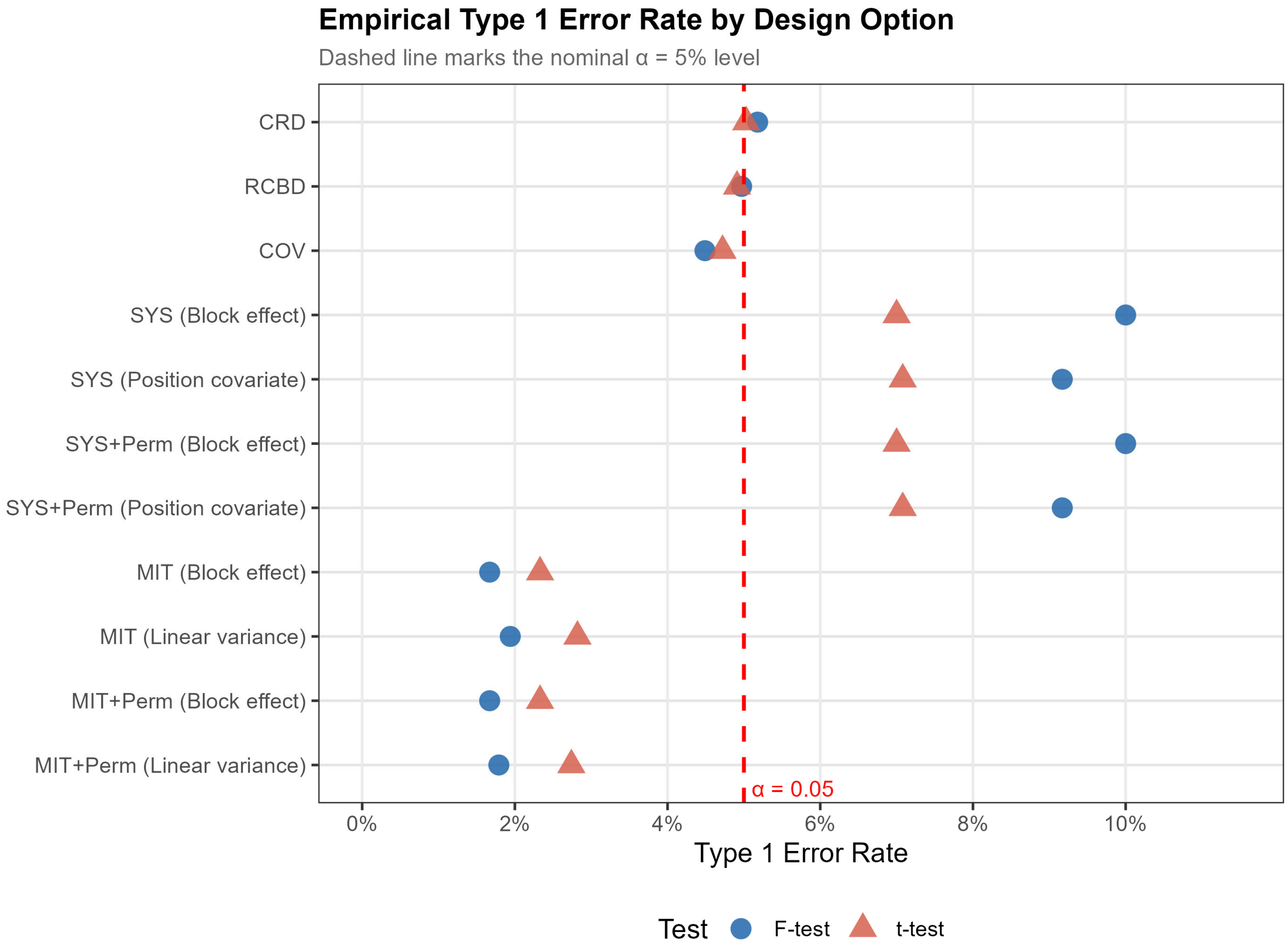


Figure: Empirical Type 1 error rates using Mercer and Hall (1911, Table V) wheat uniformity data for different design and analysis options. (Piepho *et al*. 2026)

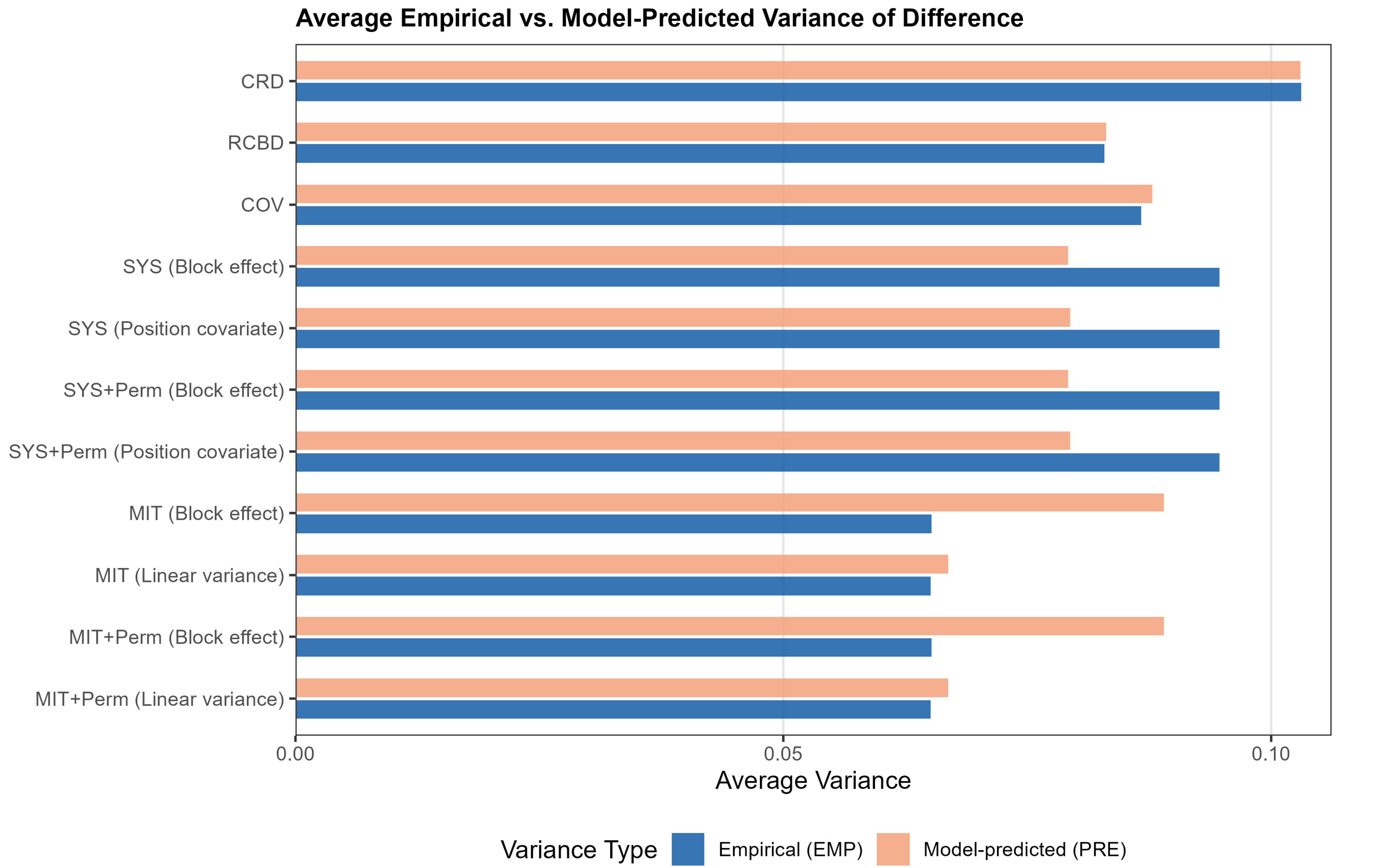


Figure: Empirical variance of a difference (EMP*) and model-based variance of a difference (PRE*) using Mercer & Hall (1911, Table V) data. (Piepho *et al*. 2026)

Table: Empirical variance of a difference (EMP*), model-based predicted variance of a difference (PRE*), and empirical Type 1 error rate at α=5%.

MIT = Mitscherlich design

* Besag and Kempton (1986)
Baird and Mead (1991)

Design / analysis options	A – B	A – C	A – D	A – E	B – C	B – D	B – E	C – D	C – E	D – E
MIT										
Block effect	0.0563 (0.0890) [0.0250]	0.0642 (0.0890) [0.0083]	0.0806 (0.0890) [0.0250]	0.0658 (0.0890) [0.0333]	0.0521 (0.0890) [0.0333]	0.0642 (0.0890) [0.0167]	0.0742 (0.0890) [0.0250]	0.0577 (0.0890) [0.0167]	0.0681 (0.0890) [0.0083]	0.0691 (0.0890) [0.0417]
Linear variance[1]	0.0568 (0.0650) [0.0167]	0.0618 (0.0669) [0.0083]	0.0845 (0.0687) [0.0250]	0.0598 (0.0707) [0.0250]	0.0527 (0.0649) [0.0250]	0.0640 (0.0668) [0.0167]	0.0707 (0.0687) [0.0333]	0.0618 (0.0649) [0.0250]	0.0682 (0.0667) [0.0167]	0.0708 (0.0650) [0.0500]
MIT + Permutation										
Block effect	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0652 (0.0890) [0.0233]	0.0655 (0.0890) [0.0233]	0.0654 (0.0890) [0.0233]
Linear variance[1]	0.0655 (0.0669) [0.0273]	0.0652 (0.0669) [0.0276]	0.0651 (0.0669) [0.0273]	0.0651 (0.0669) [0.0275]	0.0650 (0.0669) [0.0274]	0.0650 (0.0669) [0.0273]	0.0651 (0.0669) [0.0275]	0.0651 (0.0669) [0.0277]	0.0651 (0.0669) [0.0274]	0.0654 (0.0669) [0.0275]

[1] Spatial model on Slide 16

(Piepho *et al*. 2026)

Disadvantages of systematic arrangements

(1) There can be no assurance that the estimate of error is unbiased, however this estimate is arrived at, and the objectivity of the tests of significance is consequently lost.

(2) Many different methods of estimating the error can reasonably be advocated, so that the tests of significance are not even unique.

(3) The comparisons of different pairs of treatments are subject to different errors, so that even if the estimate of error is reasonably unbiased, it cannot be used to test individual differences.

(4) Biases may be introduced into the treatment means, owing to the pattern of the systematic arrangement coinciding with some fertility pattern in the field, and this bias may persist over whole groups of experiments owing to the arrangement being the same in all. Competition between plots with different treatments which always fall next to one another may produce similar effects.

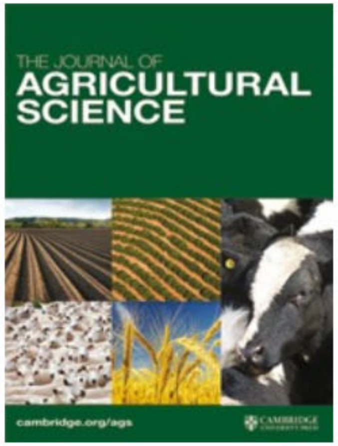
(Yates 1939)

CAMBRIDGE UNIVERSITY PRESS

Discover Content | Products and Services

Cambridge Core | Browse | Services | Open research

Home > Journals > The Journal of Agricultural Science > Accepted manuscript > An empirical comparison of systematic and randomized...



The Journal of Agricultural Science

Accepted manuscript

An empirical comparison of systematic and randomized field experimental designs

Published online by Cambridge University Press: **15 June 2026**

Hans-Peter Piepho, Waqas Ahmed Malik and Emlyn Williams

Show author details

Article | Metrics

Save PDF | Share | Cite | Rights & Permissions

https://doi.org/10.1017/S0021859626100732

Cornell University

Learn about arXiv becoming an independent nonprofit.

We gratefully acknowledge support from the Simons Foundation, member institutions, and all contributors.

Donate

arXiv > stat > arXiv:2605.29758

Search... All fields Search

Help | Advanced Search

Statistics > Methodology

[Submitted on 28 May 2026]

Fisher's ideas and the design of field experiments in agronomy and plant breeding

Hans-Peter Piepho

R. A. Fisher was one of the greatest scientists of the last century. He made many ground-breaking contributions, so many indeed that it seems almost impossible to list all of them. His revolutionary contributions to the design of experiments can mostly be traced to the early part of his academic career, and they are inextricably linked to his involvement with agricultural field experiments at Rothamsted Experiment Station. In this talk I will review Fisher's key ideas on experimental design and relate them to some of the work I am involved in, most of which directly focuses on field experiments in agriculture. Topics covered include systematic designs, row-column designs, augmented row-column designs, multi-environment trials, partially replicated designs, optimal allocation of trials to zones in sub-divided target populations of environments, and the connection of trialling systems across countries.

Comments:	31 pages, 2 tables
Subjects:	**Methodology (stat.ME)**
ACM classes:	G.3
Cite as:	arXiv:2605.29758 **[stat.ME]**

Access Paper:

View PDF

Current browse context:
stat.ME
< prev | next >
new | recent | 2026-05
Change to browse by:
stat

References & Citations
NASA ADS
Google Scholar
Semantic Scholar

Export BibTeX Citation

Bookmark

http://arxiv.org/abs/2605.29758

Thank you!

Fisher-“Student” dispute on the merits and demerits of systematic designs

> *Fisher felt that the estimate of error used by "Student" in this type of experiment was not valid, whilst "Student" maintained to his death that the problem was more theoretical than practical, that the gain in precision which resulted was more than offset by any theoretical lack of validity. There seems to be little doubt that "Student" was correct on this point.*

(Speed 1991)

Our approach leans heavily towards Fisher, including his use of uniformity trial data to assess the validity of estimates of error, but is open to mild restrictions on randomization if there is substantive reason to expect improvements in precision.

(Piepho *et al*. 2026)

Randomization test just a diagnostic device in case of suspected departures

In recent years tests using the physical act of randomization to supply (on the Null Hypothesis) a frequency distribution, have been largely advocated under the name of “Non-parametric” tests. Somewhat extravagant claims have often been made on their behalf. The example of this Chapter, published in 1935, was by many years the first of its class. The reader will realise that it was in no sense put forward to supersede the common and expeditious tests based on the Gaussian theory of errors. The utility of such nonparametric tests consists in their being able to supply confirmation whenever, rightly or, more often, wrongly, it is suspected that the simpler tests have been appreciably injured by departures from normality.

(Fisher 1935; 1971, The Design of Experiments, p. 48)