



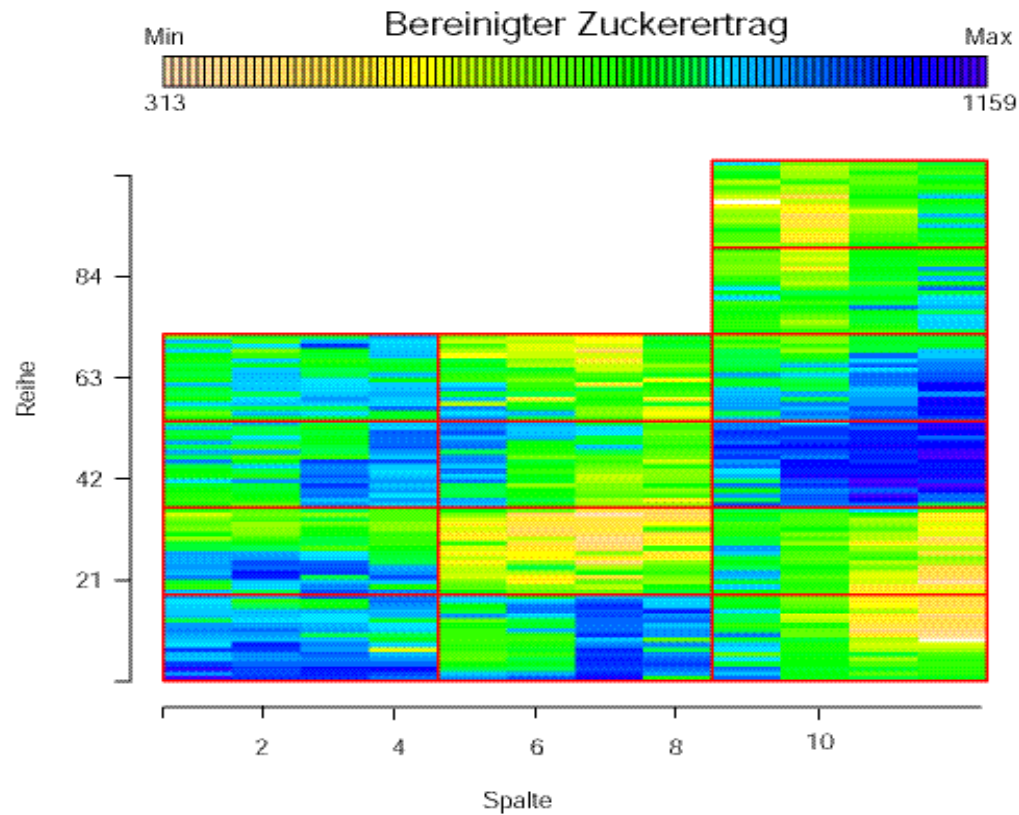
Genauigkeit von Nächst- Nachbarmodellen in pflanzenzüchterischen Versuchen

Bettina Müller, Hans-Peter Piepho
Institut für Pflanzenbau & Grünland
Fachgebiet Bioinformatik

Gliederung

1. Genauigkeit der verschiedenen geostatistischen Modelle
 - Methoden, Datensatz, Ziel
 - Bewertungskriterien
 - Ergebnisse Bewertung
2. Simulation unterschiedlicher augmented Design
 - Unterschiedlichste Anordnungen
 - Eindimensionale Auswertung
 - Zweidimensionale Auswertung

Warum Nächst - Nachbar Modelle?



Räumliches Modell

$$v_i = \mu + g_i + t_i + \varepsilon_i$$

v_i : beobachtete Ertrag auf der i ten Parzelle

g_i : Effekt des Genotyps auf der i ten Parzelle

t_i : räumliche Umwelteffekt

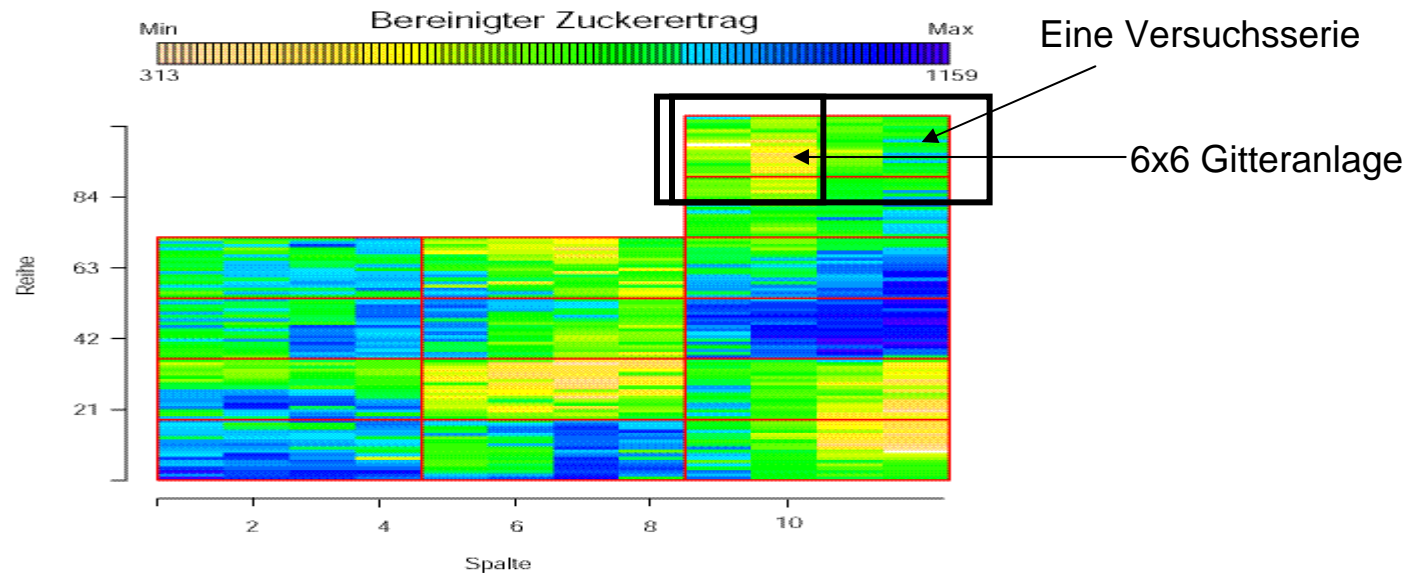
ε_i : Restfehler

μ : Achsenabschnitt

Modelle

- Baseline Modell [Base]
- Erste Differenzen [FD]
- Zweite Differenzen [SD]
- Lineare Varianz nach Williams [LV]
 - Mit und ohne Nugget
 - Mit festem und zufälligem Spalteneffekt
- Random Walk (mit und ohne Nugget) [Ran]
- Autoregressive
Mit festem und zufälligem Spalteneffekt [AR(1)]
- Exponentiell [Exp]
- Sphärisch [Sph]

Datensatz Zuckerrübe



- 293 Zuckerrübenversuche auf 21 Standorten in einem Jahr
- 6x6 Gitteranlagen mit je zwei Wiederholungen
- 72 Parzellen, aufgeteilt auf 4 Reihen und 18 Spalten
- 837 Zuckerrüben-genotypen

Ziel

Wie effizient und präzise sind die unterschiedlichen Nächst- Nachbar Modelle?

Bewertungskriterien

1. AIC (Akaike Information Kriterium)
→ Für Modelle mit demselben festen Effekt
2. Phänotypische Korrelation (Präzision)

Phänotypische Korrelation für i ten Genotypen in j ter Umwelt (bei Gitteranlagen)

$$y_{ij} = \mu_i + g_{ij} + e_{ij}$$

$$\text{corr}(y_{ij}, y_{ij'}) = \frac{\text{cov}(y_{ij}, y_{ij'})}{\sqrt{\text{var}(y_{ij}) \times \text{var}(y_{ij'})}}$$



$$\text{corr}(y_{ij}, y_{ij'}) = \frac{\text{cov}(g_{ij}, g_{ij'})}{\sqrt{\text{var}(g_{ij} + e_{ij}) \times \text{var}(g_{ij} + e_{ij'})}}$$

Effizienz

$$\overline{v.e.d.} = \frac{1}{t(t-1)} \sum_{k \neq k'} [s.e.d. (k, k')]^2$$

$$\text{Relative Effizienz} = \left(\frac{\overline{v.e.d. \text{ Model ohne Geostatistik}}}{\overline{v.e.d. \text{ Model mit Geostatistik}}} \right) \times 100\%$$

k = Genotyp

t = Anzahl der Genotypen

s.e.d = Standardfehler der Differenz

v.e.d = Varianz einer Differenz

Ergebnisse

UNIVERSITÄT HOHENHEIM



Ergebnisse - AIC Kriterium

Tabelle 1: Prozentualer Anteil (%) des Modells mit den niedrigsten AIC-Werten für Modelle mit gleichem festen Effekt.

Modell	Prozent
Model ohne räumlichen Trend und fixem Spalteneffekt	46%
Lineares Varianz Modell mit Nugget und Random Walk mit Nugget	28%
Ar(1) und Exponentielles Modell	17%
Gaussches Model	6%
Lineares Varianz Modell ohne Nugget Random Walk ohne Nugget	1%
Sphärisches Modell	1%

Ergebnisse – Phänotypische Korrelation über die Modelle

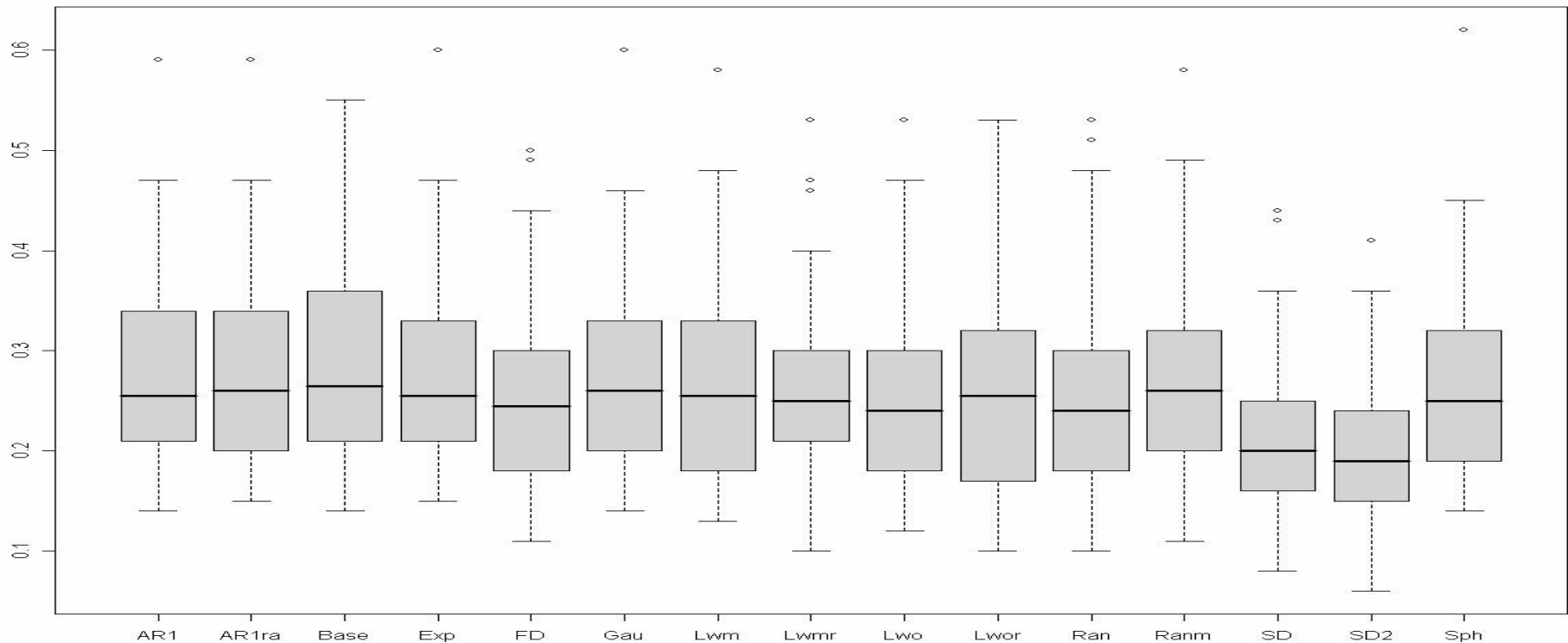


Abbildung 1: Spannweite der phänotypischen Korrelation über die einzelnen Modelle in allen Versuchsserien. (Sph= Sphärisches Modell, Ranm= Random Walk mit zufälligem Spalteneffekt, Ran= Random Walk, Lwor= Lineares Modell nach Williams ohne Nugget mit zufälligem Spalteneffekt, Lwo= Lineares Modell nach Williams ohne Nugget und festem Spalteneffekt, Lwmr= Lineares Modell nach Williams mit Nugget und zufälligem Spalteneffekt, Lwm= Lineares Modell nach Williams mit Nugget, Gau= Gaußsches Modell, Exp= Exponentielles Modell, AR1ra= AR(1)-Modell mit zufälligem Spalteneffekt, AR1= AR(1)-Modell, Base= Basis Modell ohne Geostatistik, SD = Zweite Differenzen nach Schwarzbach, SD2 = Zweite Differenzen nach Williams, FD = Erste Differenzen)

Ergebnisse – Phänotypische Korrelation über die Versuchsserien

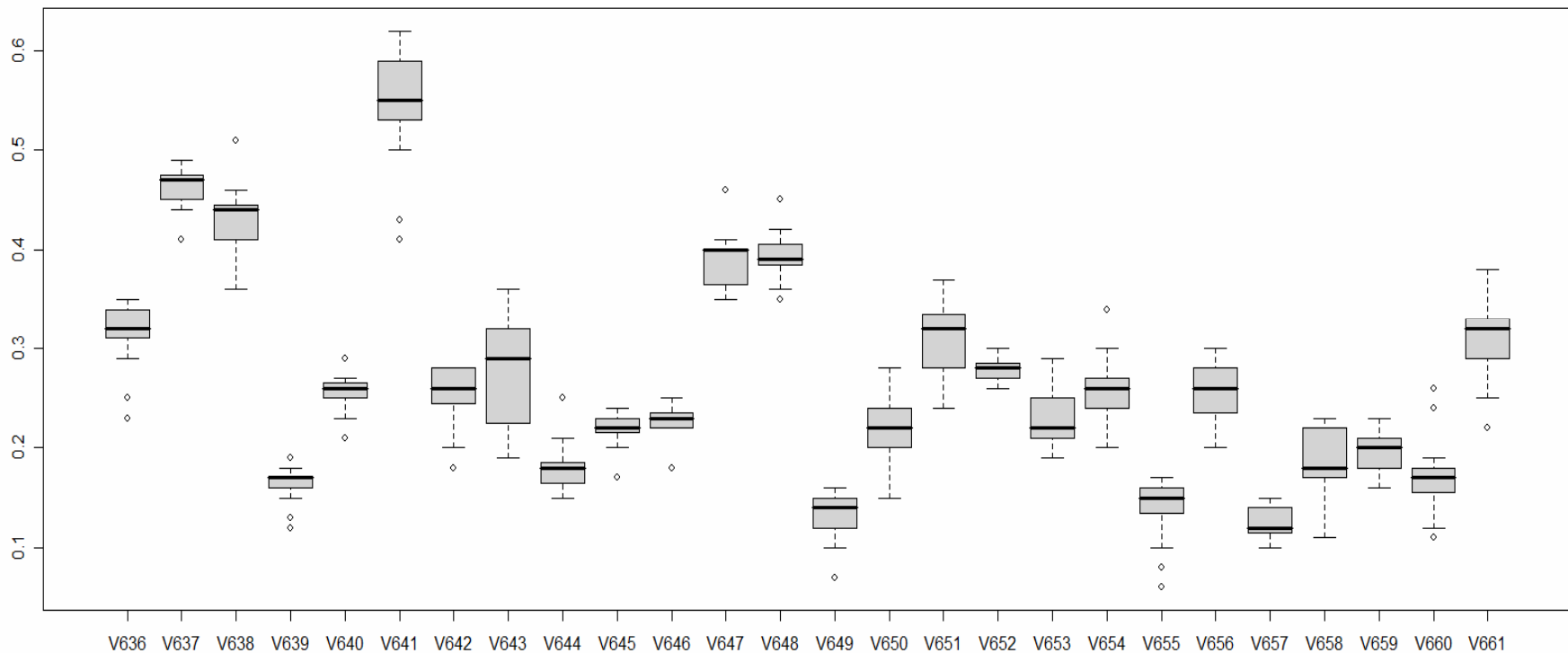


Abbildung 2: Spannweite der phänotypischen Korrelation über die einzelnen Versuchsserien in allen Modellen.

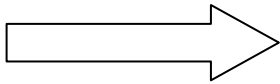
Ergebnisse - Effizienz

Tabelle 2: Schätzwerte der berechneten Effizienz in % modelliert über alle Standorte und Versuche.

Nr.	Modell	Schätzwert
2a	Zweite Differenzen (Endparzelle nach Schwarzbach)	163.17
2b	Zweite Differenzen (Endparzelle nach Williams)	161.20
7	Lineares Modell nach Williams ohne Nugget und zufälligen Spalteneffekt	153.82
8	Lineares Modell nach Williams mit Nugget und zufälligen Spalteneffekt	152.22
3	Random Walk ohne Nugget	143.91
1	Erste Differenzen	143.19
5	Lineares Modell nach Williams ohne Nugget und festem Spalteneffekt	143.01
10	AR(1) mit zufälligem Spalteneffekt	130.12
9	AR(1)	128.53
6	Lineares Modell nach Williams mit Nugget und festem Spalteneffekt	126.54
11	Exponentielles Modell	126.49
12	Sphärisches Modell	125.65
4	Random Walk mit Nugget	123.47
13	Gaußsches Modell	119.93

Fazit

- Das Baseline Model zeigt in den meisten Fällen den besten AIC Wert
- Das Baseline Modell hat ähnlich hohe Präzision wie die Nächst- Nachbarmodelle



Für die untersuchten Versuche liefert eine geostatistische Modellierung nur geringen Gewinn.

Augmented Design

- Verschiedene Anordnung von Standards
- Unterschiedliche Anzahl von Standards
- Unterschiedliche Blockgröße
- Ein- oder zweidimensionale Auswertung

Unterschiedliche Anordnungen mit je 2 Standards

- 12 er Block

- nach jeder 6.ten Position
- 3 verschiedene Anordnungen
(zufällige Verteilung)

EEEECEEECEEEE
EEEECEEECEEEE
EEEECEEECEEEE

- 6 verschiedene Anordnungen
(zufällige Verteilung)

EEEECEEECEEEE
EEEECEEECEEEE
EEEECEEECEEEE
EEEECEEECEEEE
EEEECEEECEEEE
EEEECEEECEEEE

- 18 er Block

- nach jeder 9.ten Position
- 3 verschiedene Anordnungen
(zufällige Verteilung)

EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE

- 6 verschiedene Anordnungen
(zufällige Verteilung)

EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE
EEEEEECECEEEEEEEEE

Zwei vs. drei Standards (18er Block)

ZWEI STANDARDS

- nach jeder 9.ten Position
- 3 verschiedene Anordnungen (zufällige Verteilung)

```
EEEEEEECCEEEEEEEEE  
EEEEEECECEEEEEEEEE  
EEEEEECEEECEEEEEEEE
```

- 6 verschiedene Anordnungen (zufällige Verteilung)

```
EEEEEEECCEEEEEEEEE  
EEEEEECECEEEEEEEEE  
EEEEEECEEECEEEEEEEE  
EEEEEECEEECEEEEEEEE  
EEEEEECEEECEEEEEEEE  
EEEEEECEEECEEEEEEEE
```

DREI STANDARDS

- nach jeder 6.ten Position
- 6 verschiedene Anordnungen symmetrisch (zufällige Verteilung)

```
EEEEEECCCEEEEEEEEE  
EEEEEECECEEEEEEEEE  
EEEEEECEEECEEEEEEEE  
EEEECEEECEEECEEEEE  
EEEECEEECEEECEEEEE  
EECEEEEECEEEEECEEE
```

- 6 verschiedene Anordnungen unsymmetrisch (zufällige Verteilung)

```
EEEEECCEEEEEEEEEEEE  
EEEECECECEEEEEEEEE  
EEEECEEECEEECEEEEE  
EEECCEEECEEECEEEEE  
EEECCEEECEEECEEEEE  
EECEEEEECEEEEECEEE
```



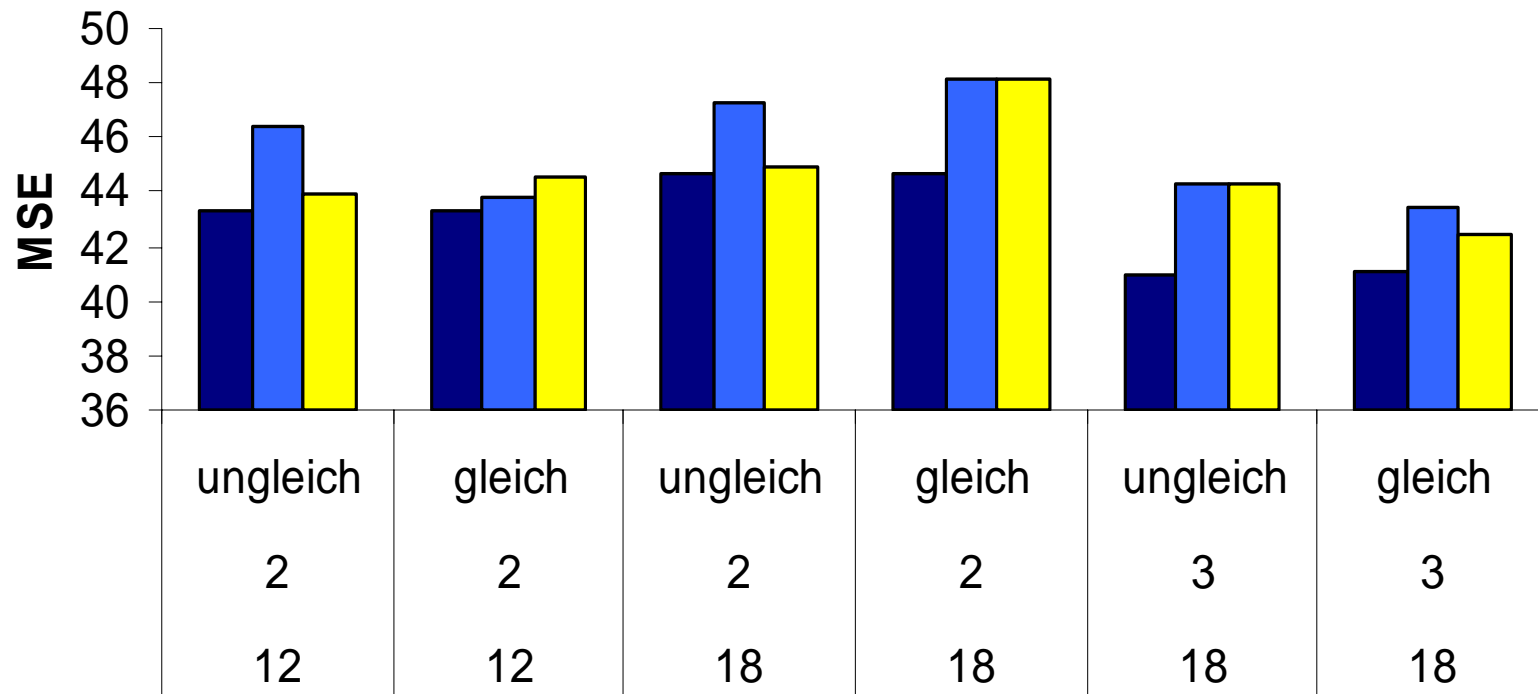
MSE Werte der Modelle im Überblick

Block	Anzahl	Standard	Anordnung	Ohne Geo	Baseline	LV+ EV	LV o. EV	Ar1 + EV	Exp + EV	Sph + EV
12	2	ungleich	6. te Positon	56,83	43,56	43,28	54,16	87,06	44,42	50,83
12	2	ungleich	6Kombinationen	56,70	45,79	46,36	60,25	48,56	48,56	45,79
12	2	ungleich	3Kombinationen	55,65	43,95	43,95	62,64	44,01	43,95	43,95
18	2	ungleich	9. te Position	55,28	44,50	44,63	61,00	51,04	45,46	55,48
18	2	ungleich	6Kombinationen	55,39	45,10	47,22	63,86	48,48	50,72	53,76
18	2	ungleich	3Kombinationen	55,75	43,90	44,87	62,94	46,80	50,43	54,39
12	2	gleich	6. te Positon	43,35	43,61	43,35	54,23	87,26	44,49	50,92
12	2	gleich	6Kombinationen	55,75	43,85	43,85	67,52	43,85	45,32	43,85
12	2	gleich	3Kombinationen	56,09	44,58	44,58	63,04	47,34	47,33	51,43
18	2	gleich	9. te Position	55,34	44,64	44,70	61,02	44,64	45,58	55,56
18	2	gleich	6Kombinationen	58,46	48,82	48,12	54,25	48,82	51,03	48,82
18	2	gleich	3Kombinationen	57,10	48,90	48,87	n.c	48,90	48,90	48,90
18	3	ungleich	6.te Position	56,77	40,93	40,93	50,57	n.c	40,93	n.c
18	3	ungleich	6Kombinationen_unsy	57,32	43,71	44,26	62,82	51,09	51,08	53,83
18	3	ungleich	6Kombinationen_symn	58,04	41,64	41,89	51,00	45,22	45,22	44,54
18	3	gleich	6.te Position	56,90	41,02	41,02	n.c	41,07	41,02	46,68
18	3	gleich	6Kombinationen_unsy	56,11	43,10	43,43	67,84	45,52	45,52	43,10
18	3	gleich	6Kombinationen_symn	55,92	40,55	42,46	59,59	n.c	46,95	40,55

MSE= Mean Square Error

Lineares Varianz- Modell mit Nugget

■ 6.- 9.te Position ■ 6 Kombinationen ■ 3 Kombinationen



Ergebnisse – Eindimensionale Auswertung

- Kleine Blöcke waren besser
- Je mehr Standards desto besser
- Besser verschiedene Standards als gleicher Standard
- LV + EV und Baseline mit zufälligem Spalteneffekt hatten geringsten MSE Wert.

Zweidimensionale Auswertung (9 x 10 Versuchsserie)

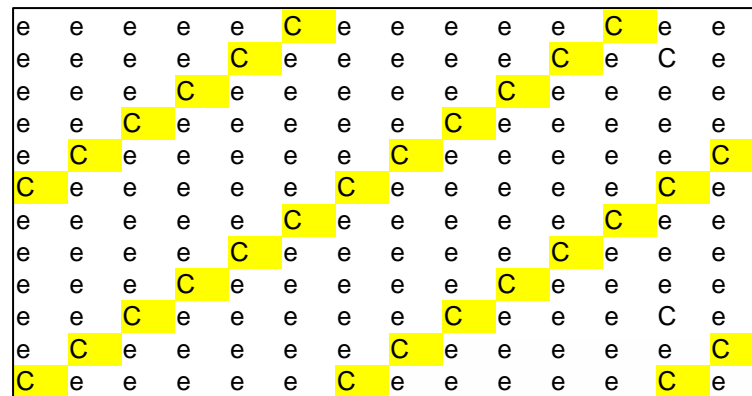
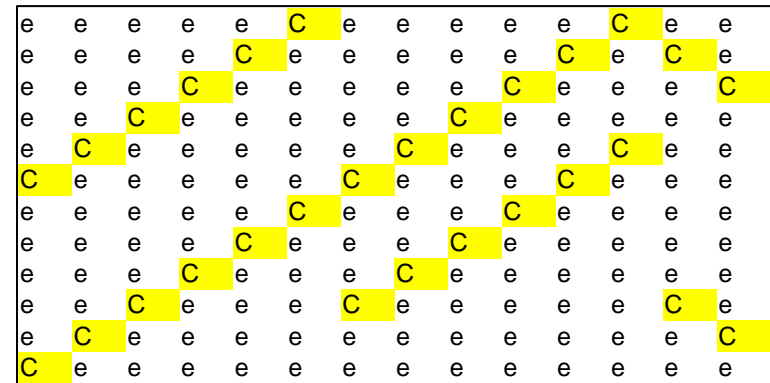
e	e	e	C	e	e	e	e	e	C
e	e	C	e	e	e	e	e	C	e
e	C	e	e	e	e	e	C	e	e
C	e	e	e	e	e	C	e	e	C
e	e	e	e	e	C	e	e	C	e
e	e	e	C	C	e	e	C	e	e
e	e	C	e	e	e	C	e	e	e
e	C	e	e	e	C	e	e	e	e
C	e	e	e	C	e	e	e	e	e

C	e	e	e	C	e	e	e	C	e
e	C	e	e	e	e	e	e	C	e
e	e	C	e	e	e	e	C	e	e
e	e	e	C	e	C	e	e	e	e
e	e	e	e	C	e	e	e	e	C
e	e	e	C	C	e	e	e	e	e
e	e	C	e	e	e	C	e	e	e
e	C	e	e	e	e	e	C	e	e
C	e	e	e	e	e	e	e	C	C

e	e	C	e	e	e	e	e	e	C
e	C	e	e	e	e	e	e	C	e
C	e	e	e	e	e	e	C	e	C
e	e	e	e	e	e	C	e	C	e
e	e	e	e	e	C	e	C	e	e
e	e	e	C	C	e	C	e	e	e
e	e	C	e	e	C	e	e	e	e
e	C	e	e	C	e	e	e	e	e
C	e	e	C	e	e	e	e	e	e



Zweidimensionale Auswertung (12 x 15 Versuchsserie)



Modelle im Vergleich

Tabelle 3: MSE- Werte für die einzelnen Versuchsserien für zweidimensionale Modelle

Versuchsserie	Anordnung	MSE- Wert (Baseline)	MSE- Wert (LvLv)	MSE- Wert (I_Vc)	MSE- Wert (Vr_I)	MSE- Wert (LV3)	MSE- Wert (AR1 x AR1)
9 x 10	1	39,06	36,89	38,52	41,36	42,41	38,50
9 x 10	2	35,33	36,55	35,71	36,07	41,21	37,01
9 x 10	3	41,4	38,93	41,76	41,39	n.c	43,76
12 x 15	1	36,61	36,84	37,08	40,03	40,40	36,61
12 x 15	2	38,36	35,68	38,26	38,92	40,61	36,61
12 x 15	3	35,33	33,27	34,77	n.c	37,02	37,5
36 x 30	1	40,93	38,80	35,80	36,64	n.c	40,93

n.c= Convergence not met.

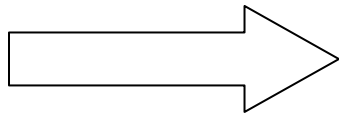
MSE: Mean Square Error

UNIVERSITÄT HOHENHEIM



Fazit

- Für die 2-D Auswertung hat den kleinsten MSE $LV \times LV$
- Generell MSE für 2-D Auswertung besser als für 1-D Auswertung



Gewinn an Genauigkeit der Schätzungen durch 2-D Auswertung.

Vielen Dank für ihre
Aufmerksamkeit!

UNIVERSITÄT HOHENHEIM



Referenzen

- Piepho, H.P, Richter, C. and Williams, E., 2008: Nearest neighbor adjustment and linear variance models in plant breeding, *Biometrical Journal*, 50 (2), 164-189.
- Piepho, H.P. & Williams, E., 2008, Two-dimensional linear variance structures for field trials, submitted
- Williams, E.R., 1986: A neighbor model for field experiments, *Biometrika* 73, 279-287.



Modelle

