

# **Räumliche Modelle in landwirtschaftlichen Feldversuchen – Abhängigkeit der Ergebnisse von Versuchsplan, Randomisationsplan und Position auf der Versuchsfläche**

Christel Richter

Humboldt-Universität zu Berlin

Invalidenstraße 42

10115 Berlin

ch.richter@agrار.hu-berlin.de

Bärbel Kroschewski

Humboldt-Universität zu Berlin

Invalidenstraße 42

10115 Berlin

b.kroschewski@agrار.hu-berlin.de

## **Zusammenfassung**

Die mögliche Berücksichtigung räumlicher Korrelationen bei der Anlage und Auswertung landwirtschaftlicher Feldversuche ist in den letzten Jahren verstärkt ins Blickfeld gerückt. Grundlage dessen ist die Entwicklung der Theorie des gemischten linearen Modells einerseits und die Implementierung der Auswertung in SAS PROC MIXED (Littell, et al., 2006) in Kombination mit einer Vielzahl möglicher Kovarianzstrukturen andererseits. Untersuchungen zu Vorzügen, Nachteilen und Problemen dieser Ansätze werden in der Literatur anhand von realen Versuchen, simulierten Daten oder Blindversuchen vorgenommen. Wir haben einen 5-jährigen Blindversuch von derselben Fläche mit teilweise unterschiedlichen Fruchtarten analysiert. In der Regel werden bei Untersuchungen auf der Basis von Blindversuchen alle Parzellen dieses Versuchs einbezogen und Vor- und Nachteile verschiedener Versuchspläne und Auswertungsmodelle analysiert. Im Gegensatz dazu haben wir verschiedene Versuchspläne mit je 1000 Randomisationsplänen auf verschiedene Positionen der Fläche projiziert. Ziel der Untersuchung ist es, den Einfluss von Position, Versuchs- und Randomisationsplan auf die Auswahl des am besten angepassten Modells zu erkennen, eventuelle Unterschiede zwischen den Fruchtarten herauszuarbeiten und die Konsequenzen der Modellwahl für die Ergebnisse der Prüfgliedvergleiche zu betrachten.

**Schlüsselwörter:** räumliches Modell, Feldversuch, gemischtes lineares Modell, PROC MIXED, MIXED-Prozedur

## **1 Problemstellung**

Es ist im landwirtschaftlichen Feldversuchswesen allgemein anerkannt, dass zwischen benachbarten Parzellen Kovarianzen existieren. Schon Wiebe [13] wies darauf hin, dass in Getreideversuchen zwischen den Parzellen eine positive, mit zunehmender Distanz fallende Korrelation der Kornerträge besteht. Das Prinzip der Randomisation ist die klassische Herangehensweise, das Problem korrelierter Fehler „zu überwinden“ [14] oder „zu neutralisieren“ [15]. Gleichzeitig dient dieses Prinzip dazu, einseitige Vor- oder Nachteile für einzelne Prüfglieder durch nicht erkannte Störgrößen zu vermeiden. Vor Versuchsbeginn erkannte Störgrößen werden durch Bildung von Blocks erfasst [3] und durch die Zuordnung der Prüfglieder zu den Blocks in ihrer Wirkung egalisiert. Gemeinsam bilden beide Prinzipien bis heute das Rückgrat der Versuchsplanung. Diese zwei Grundsätze bewirken im einfachen linearen Modell, dass innerhalb eines Blocks

die Beobachtungen als unkorreliert (Blocks: fix) betrachtet werden oder von einer konstanten Korrelation (Blocks: zufällig) ausgegangen wird. Zwischen Beobachtungen verschiedener Blocks gibt es keine Korrelation. Die positiven Effekte der Blockbildung und Randomisation (unverzerrte Schätzungen der Prüfglieddifferenzen und reduzierte Restvarianz) hängen jedoch von der korrekten Einschätzung der Bodenheterogenität ab. So traten immer wieder Zweifel auf, ob die Vorinformation zur korrekten Anlage eines Blockversuchs immer existiert bzw. richtig genutzt wird oder ob nicht die in der Nachbarschaft der Parzellen enthaltene weitere Information es wert ist, genutzt zu werden.

Parallel und zeitgleich zu obigem Ansatz entwickelten zunächst heuristisch geprägte Konzepte, die die räumliche Abhängigkeit berücksichtigten - die sogenannten Nächst-Nachbar-Methoden ([6], [1]). Als Lehrmeinung konnten sie sich jedoch nur begrenzt durchsetzen.

Durch die Verallgemeinerung des einfachen linearen Modells zum gemischten linearen Modell und der entsprechenden Implementation in SAS PROC MIXED ist prinzipiell die Möglichkeit gegeben, räumliche Kovarianzen in die Auswertung einzubeziehen. Daraus ergab sich eine Vielzahl von Untersuchungen zu Vorzügen, Nachteilen und Problemen dieser Ansätze anhand von

- realen Versuchen mit unbekannter räumlicher Struktur und unbekanntem Prüfgliedeffekten (z.B. [8], [7], [10], [12])
- simulierten Daten mit vorgegebener räumlicher Struktur und bekannten (existierenden oder nicht existierenden) Prüfgliedeffekten (z.B. [4], [11],) und
- Blindversuchen (z.B. [15]),

einschließlich theoretischer Untersuchungen, u.a. zu den Beziehungen zwischen geostatistischen Ansätzen und den Nächst-Nachbar-Methoden ([8]).

Blindversuche zeichnen sich dadurch aus, dass alle Faktoren im Versuch konstant gehalten werden, also keine Prüfgliedeffekte existieren. Sie wurden und werden mit dem Ziel angelegt, etwas über die räumliche Struktur der Fläche zu erfahren, so dass künftige Versuche dem bestmöglich angepasst werden können. Dazu werden Merkmale (häufig der Ertrag) auf einem möglichst engmaschigen Raster (z.B. 1m x 1m) auf der Fläche erfasst. Sodann werden verschiedene Versuchspläne auf die Gesamtfläche projiziert, die sich in der Teilstückgröße, Prüfgliedanzahl, Anordnung der Teilstücke innerhalb der Blocks und der Anordnung der Blocks auf der Fläche unterscheiden. Verschiedene Teilstückgrößen werden durch Zusammenfassung benachbarter Rasterelemente gebildet. In unseren Untersuchungen steht nicht die Diskussion der Vor- und Nachteile verschiedener Versuchspläne im Vordergrund, sondern wie sich diese Versuchspläne auf das bestangepasste Auswertungsmodell auswirken. Bei derartigen Betrachtungen werden in der Regel alle Parzellen des Blindversuchs gleichzeitig einbezogen. Dieses Vorgehen haben wir modifiziert (siehe Material und Methoden), da zu vermuten ist, dass nicht nur der Versuchsplan, sondern auch seine Position auf der Fläche und unter Umständen der gewählte Randomisationsplan einen Einfluss auf die relative Vorzüglichkeit eines Auswertungsmodells hat. In ausgewählten Fällen werden die Auswirkungen der Modellwahl auf die Ergebnisse der Prüfgliedvergleiche betrachtet.

## 2 Material und Methoden

Grundlage unserer Untersuchungen ist ein Blindversuch, der in den Jahren 1954 bis 1958 auf einer Fläche der damaligen Versuchsstation Blumberg (diluvialer Standort) der Humboldt-Universität zu Berlin mit 4 verschiedenen Fruchtarten durchgeführt wurde ([2], [9]). Die Fläche hatte eine Größe von 40 m x 48 m. Auf einem Raster von 2 m x 2 m wurden die Erträge der Fruchtarten Hanf (1954), Winterroggen (1955), Rüben(1956), Hafer (1957) und Hanf (1958) erfasst. Die Ertragskarten sind Abb. 1 zu entnehmen

Die Semivariogramme in Abb. 2 zeigen die existierende räumliche Korrelation. Bei allen Fruchtarten liegt in x-Richtung eine räumliche Abhängigkeit bis zu einer Distanz von ca. 12 – 15 m vor, in y-Richtung ist - mit Ausnahme von Hanf 1958 - die Abhängigkeit stärker und über größere Distanzen ausgeprägt.

Für die Untersuchungen haben wir verschiedene Versuchspläne so gewählt, dass sie nicht die Gesamtfläche einnehmen, sondern nur Ausschnitte daraus. Dieses ermöglicht, den Einfluss der Position auf die Ergebnisse zu analysieren. Dabei muss in Kauf genommen werden, dass durch diese Vorgehensweise die Versuche nicht sehr groß sein konnten, was bei der Interpretation zu beachten ist. Generell wurden Anlagen mit vier vollständigen Blocks (n = 4) gewählt. Die Parameter der gewählten Versuchspläne sind Tab. 1 zu entnehmen. Die Anzahl der möglichen Positionen ergibt sich aus der Verschiebung des Versuchsplanes um jeweils eine Parzellenbreite oder –länge.

**Tabelle 1:** Parameter der untersuchten Versuchspläne und Anzahl der Positionen auf der Gesamtfläche

Prüfglied- anzahl	Versuchs- nummer	Dimension Teilstück	Dimension Block	Dimension Versuch	Anzahl Positionen auf der Gesamtfläche
4	1	2 x 2	2 x 8	8 x 8	357
	2	2 x 2	8 x 2	32 x 2	120
	3	2 x 2	2 x 8	2 x 32	180
	4	2 x 8	8 x 8	32 x 8	30
	5	2 x 8	8 x 8	8 x 32	51
	6	2 x 8	2 x 32	8 x 32	51
	7	8 x 2	32 x 2	32 x 8	42
	8	8 x 2	8 x 8	32 x 8	42
	9	8 x 2	8 x 8	8 x 32	45
10	10	2 x 2	20 x 2	40 x 4	23
	11	2 x 8	20 x 8	40 x 16	5
	12	2 x 8	20 x 8	20 x 32	33

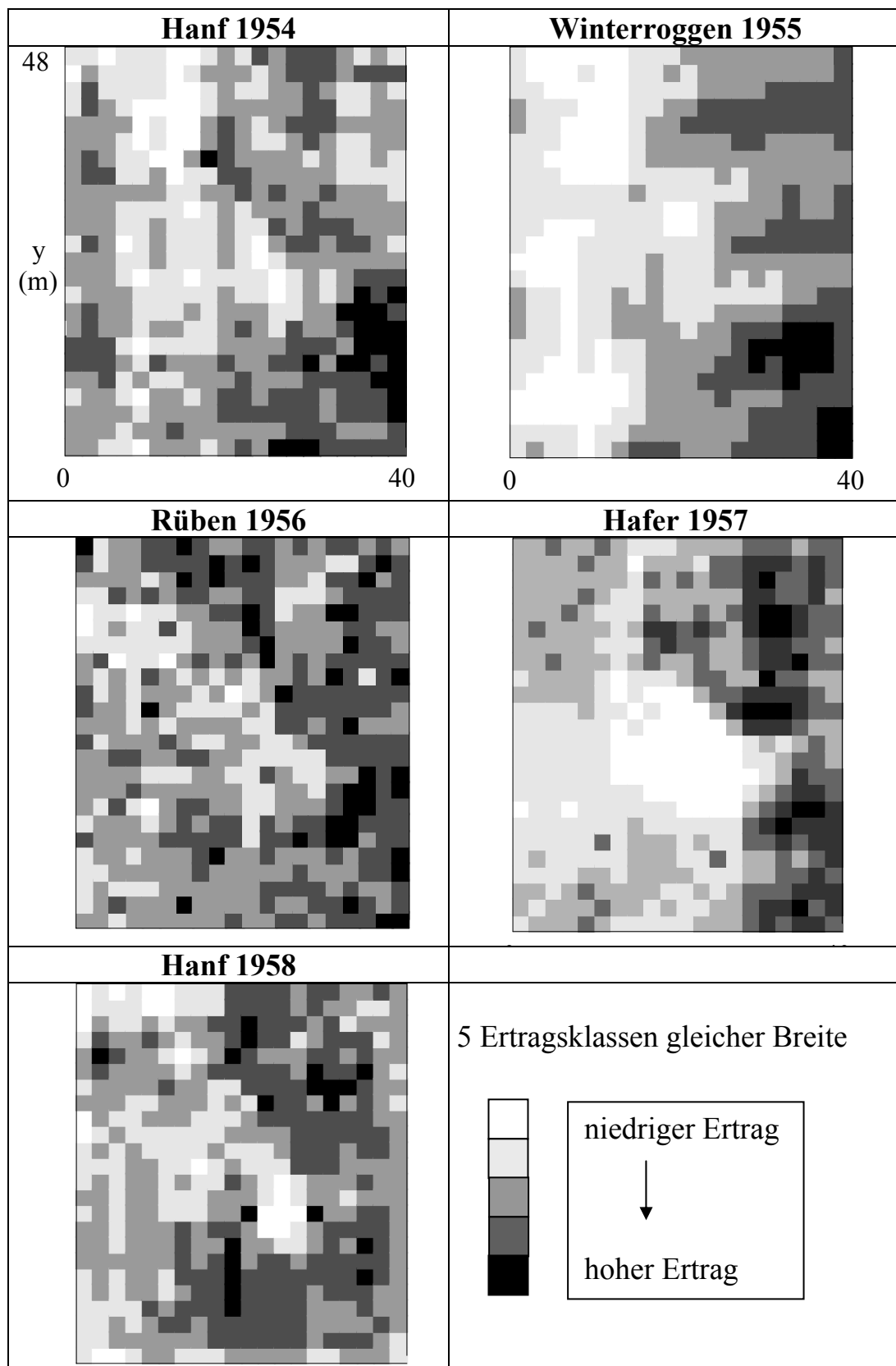
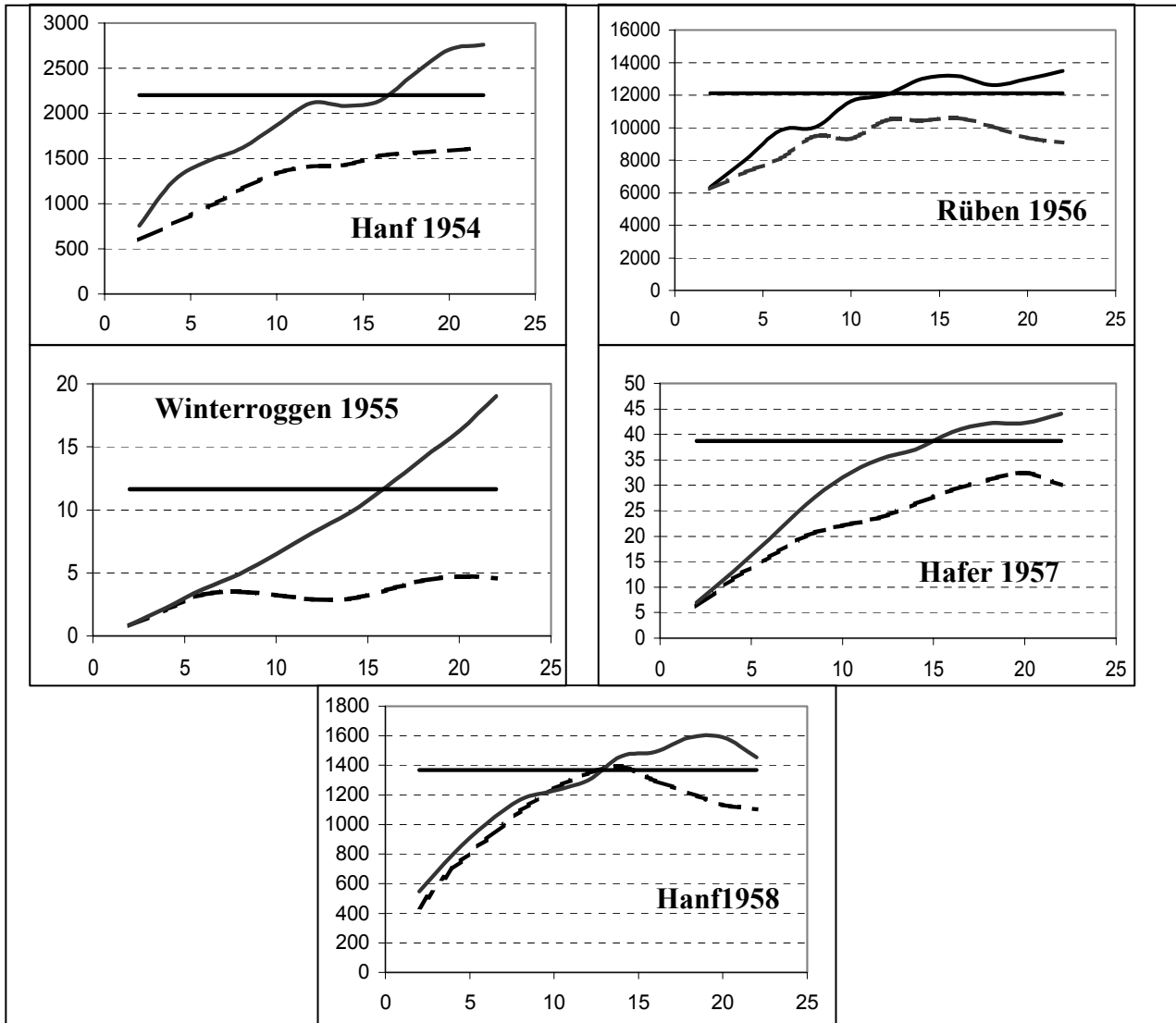


Abbildung 1: Ertragskarten des Blindversuchs

Um den Einfluss des Randomisationsplanes zu prüfen, wurden für die Prüfgliedanzahlen  $v = 4$  und  $v = 10$  mit PROC PLAN je 1000 Randomisationspläne in



**Abbildung 2:** Semivariogramme der 2x 2-Teilstücke des Blindversuchs in Richtung:  
 ——— der x-Achse      - - - der y-Achse.  
 Die waagerechte Linie entspricht der Stichprobenvarianz der 480 Teilstücke.

vollständigen Blocks erzeugt und jeder durch die Parameter in Tab. 1 festgelegte Versuchplan 1000 mal an jeder möglichen Position auf der Fläche platziert. Somit waren je Fruchtart und Versuchplan 1000 x Anzahl Positionen „Pseudoversuche“ auszuwerten. Die Auswertung erfolgte mit PROC MIXED der SAS Version 9.1 ohne Parametervorgabe. Neben dem Basismodell (Blockanlage mit fixen Blocks ohne räumliche Korrelation) wurden verschiedene räumliche Kovarianzmodelle gewählt, wobei räumliche Abhängigkeit über die Fläche des jeweiligen Pseudoversuchs oder ausschließlich innerhalb der Blocks Berücksichtigung fand. In jedem Falle wurde mit fixen Blocks gearbeitet. Generell zeigte sich, dass die Modelle mit Nugget-Varianz selten konvergierten oder sich auf das entsprechende Modell ohne Nugget-Varianz reduzierten. Gleichfalls zeigten das Matern-Modell und geometrisch anisotrope Modelle schlechte Konvergenzei-

enschaften. Für alle in Tab. 1 aufgeführten Versuchsanlagen wurden deshalb folgende Modelle durchgängig angepasst (später jeweils genutzte Abkürzung in Klammern):

### 1. Basismodell (**b**, **basis**)

```
proc mixed data = datafile;
class pruefglied block;
model ertrag = pruefglied block;
lsmeans pruefglied / pdiff;
run;
```

### 2. Modelle mit nichtlinearer räumlicher Kovarianzfunktion

- **Modellauswahl:**

*isotrop: sphärisches (sph), Power- (pow), Gauss-Modell (gau)*

```
%let mod=sph; /*%let mod=pow; %let mod=gau;*/
```

*anisotrop, sofern möglich: separables anisotropes Power-Modell (powa)*

```
%let mod=powa;
```

- **Bereich der Wirksamkeit einer Kovarianz:**

*über den gesamten Versuch:* %let bereich = intercept;

*nur innerhalb der Blocks (Modelle erhalten den Zusatz “-bl“):*

```
%let bereich = block;
```

- **Freiheitsgradkorrektur nach Satterthwaite oder Kenward-Roger:**

```
%let korr= satterth; /*%let korr = kr;*/
```

```
proc mixed data = datafile;
class pruefglied block;
model ertrag = pruefglied block /ddfm = &korr;
repeated / type = sp(&mod) (x y) subject = &bereich;
lsmeans pruefglied / pdiff;
run;
```

### 3. random-walk-Modell (bzgl. der Prüfglieddifferenzschätzungen identisch dem linearen Varianzmodell innerhalb fixer Blocks [8]) (**rw**)

```
%let v = pg_anz; /*pg_anz = Anzahl der Pruefglieder*/
data rw;
parm = 1;
array col coll-col&v;
do row = 1 to &v; do i = 1 to &v;
  col[i]=min(row,i);
end; drop i; output; end;
run;
```

```
proc mixed data = datafile;
class pruefglied block;
model ertrag = pruefglied block;
repeated / type = lin(1) ldata=rw subject = block;
```

```
lsmeans pruefglied / pdiff;  
run;
```

Wenn nur die räumliche Abhängigkeit innerhalb der Blocks berücksichtigt wurde, war eine Auswertung als anisotropes Modell nicht möglich, da in allen Versuchen die Parzellen innerhalb der Blocks ausschließlich neben- oder übereinander angeordnet waren. Wurde die Korrelation über die gesamte Fläche betrachtet, dann konnte für die Versuche 1, 5 bis 8 und 10 bis 12 auch der anisotrope Ansatz gewählt werden (siehe auch Tab. 2a und 2b).

Der Modellvergleich wurde anhand des Akaike-Informationskriteriums (AIC) vorgenommen. Da in keinem Fall bekannt ist, welches räumliche Kovarianz-Modell das wahre Modell ist, wurde auf andere Anpassungskriterien verzichtet. Es werden die prinzipiellen Konsequenzen aufgezeigt, wenn stattdessen mit dem BIC-Kriterium gearbeitet wird.

Zum Zwecke der Bewertung der Rolle von Position und Randomisation wurde jeder der 12 Versuche aus Tab. 1 für jede Fruchtart auf folgende Weise ausgewertet:

- (1) Je Kombination von Position und Randomisation wurden die oben beschriebenen 8 oder 9 Modelle angepasst und der AIC-Wert registriert.
- (2) Je Modell wurde für die AIC-Werte der Varianzanteil, der aus den Positionen, den Randomisationen und deren Wechselwirkung resultiert, bestimmt.
- (3) Je Kombination von Position und Randomisation wurden die AIC-Werte der Modelle rangiert (beste Anpassung = kleinste Rangzahl).
- (4) Für die Rang-Werte wurde je Modell der Varianzanteil, der aus den Positionen, den Randomisationen und deren Wechselwirkung resultiert, errechnet.
- (5) Über alle Kombinationen von Position und Randomisation wurde die relative Häufigkeit ermittelt, mit der jedes Modell als bestes aufgetreten ist (das Modell, das die kleinste Rangzahl bzw. den kleinsten AIC-Wert hat).
- (6) Je Position (im Mittel der 1000 Randomisationen) wurde die mittlere Rangzahl jedes Modells berechnet. Diese Maßzahl benutzen wir im Folgenden zur Charakterisierung der relativen Vorzüglichkeit eines Modells an einer bestimmten Position. Sodann wurde bestimmt, an wie vielen Positionen von den jeweils möglichen Positionen ein Modell die beste mittlere Rangzahl hat.
- (7) Je Randomisationsplan (im Mittel der jeweils vorhandenen Positionen) wurde die mittlere Rangzahl jedes Modells berechnet. Diese Maßzahl benutzen wir zur Charakterisierung der relativen Vorzüglichkeit eines Modells in Verbindung mit einem gegebenen Randomisationsplan im Mittel aller Positionen. Sodann wurde bestimmt, für wie viele Randomisationspläne von den jeweils 1000 Plänen ein Modell die beste mittlere Rangzahl hat.

Neben der Zielstellung, den Einfluss von Position und Randomisationsplan auf die Modellauswahl zu zeigen, werden für zwei ausgewählte Beispiele die Konsequenzen auf die nachfolgenden Prüfgliedvergleiche demonstriert. Zur Schätzung der Varianz-Kovarianz-Matrix wird die REML-Methode genutzt. Diese Schätzung geht in die Kontrastschätzung der fixen Effekte ein. Da es sich um geschätzte Varianzen und Kovarian-

zen und nicht um bekannte fixe Größen handelt, werden ohne Berücksichtigung dieses Sachverhalts die Standardfehler der Kontraste unterschätzt. Durch die Nutzung der geschätzten Varianz-Kovarianz-Matrix folgen die Teststatistiken nicht mehr exakt der F- oder t-Verteilung. Dem wird durch die Freiheitsgradkorrektur nach Satterthwaite (SW) bzw. Kenward-Roger (KR) Rechnung getragen. Mit der letzteren Korrektur wird darüber hinaus eine asymptotische Korrektur der geschätzten Varianz-Kovarianz-Matrix der fixen Effekte vorgenommen. Diese zwei Korrekturmöglichkeiten sind bei den Auswertungsmodellen zu beachten, in denen mehr als ein Varianz-Kovarianz-Parameter zu schätzen ist (also in allen, außer im Basis- und random walk-Modell). Für die Versuche 4 und 12 werden die Verzerrungen des geschätzten Standardfehlers bei nicht erfolgter Korrektur und bei Korrektur nach KR gegenübergestellt, sowie die empirischen Typ I-Fehler der F- und t-Tests nach beiden Korrekturvarianten (SW und KR) dargestellt.

### 3 Ergebnisse und Diskussion

#### 3.1 Bewertung der Rolle von Position und Randomisation

Bei den AIC-Werten zeigen die Varianzanteile für die einzelnen Versuchspläne, Auswertungsmodelle und Fruchtarten ein relativ einheitliches Bild. In der Mehrzahl aller Untersuchungen liegt der durch die Randomisationspläne (im Mittel der Positionen) erklärte Varianzanteil bei 0 bis 4 % und für die Positionen (im Mittel der Randomisationen) bei 70 – 80%. Der verbleibende Rest, der aus einer unterschiedlichen Reaktion der Randomisationspläne an den einzelnen Positionen resultiert, liegt mehrheitlich bei 20-30%. Dieses zeigt, dass für alle Modelle die Güte der Anpassung vorwiegend durch die Position bestimmt ist, die Randomisationspläne haben im Mittel der Positionen sehr geringen Anteil.

Die gleichen Betrachtungen für die Rangzahlen zeigen, dass der relative Anteil der Randomisationspläne (im Mittel der Positionen) noch weiter zurückgedrängt wird, aber auch der der Positionen (im Mittel der Randomisationen) abnimmt, so dass mit 50 bis 70 % der Hauptanteil der Varianz bei der Wechselwirkung liegt. Somit müssen Unterschiede in der relativen Vorzüglichkeit der einzelnen Modelle vorwiegend aus unterschiedlichen Reaktionen der Randomisationspläne an den einzelnen Positionen erklärt werden. Auch hier konnte keine Spezifik der Versuchspläne und Fruchtarten erkannt werden.

In Tab. 2a und 2b ist für die einzelnen Fruchtarten und Versuche aufgeführt, welche Modelle sich am häufigsten als bestangepasste im obigen Sinne durchgesetzt haben. Generell wurden nur jene Modelle aufgenommen, auf die dieses in mehr als 10% der jeweiligen Fälle zutreffend war. Mit (I), (II) bzw. (III) seien die Bestimmungsgrundlagen nach (5), (6) bzw. (7) aus dem Kapitel Material und Methoden bezeichnet.

Die in Tab. 2a aufgeführten Versuche zeichnen sich dadurch aus, dass das Basismodell und das random walk-Modell dominieren. Hier sind die Blocks relativ klein und kompakt. Durch die Blockbildung gelingt es, einen großen Teil der räumlichen Heterogenität über die Blockeffekte zu erfassen. Entweder ist es nicht notwendig, eine zusätzliche räumliche Struktur zu berücksichtigen (Basismodell) oder es ist ausreichend, die Kor-



relation innerhalb der Blocks zu beachten, wobei auf den kurzen Distanzen nur der lineare Anstieg im Semivariogramm eine Rolle spielt (vgl. Abb. 2). Letzteres wird durch das random walk-Modell abgebildet. Eine Ausnahme stellt der Versuch 1 dar, der sich von den anderen Versuchen aus Tab. 2a durch eine kompakte Form des Versuchs unterscheidet. In der Regel finden sich die für die Gesamtheit aller Kombinationen aus Position und Randomisation (I) am häufigsten bestangepassten Modelle auch in der Verteilung über die Positionen im Mittel der Randomisationen (II) und in der über die

**Tabelle 2a:** Bestangepasste Modelle mit einem Anteil > 10% in absteigender Folge der relativen Häufigkeiten. = gleicher Anteil. **fett:** Anteil > 50%. fett: Anteil > 90%. (A) anisotroper Ansatz möglich. Bestimmungsgrundlage I – III (siehe Text)

Vers. Nr.	Hanf 54	Roggen 55	Rübe 56	Hafer 57	Hanf 58
<b>1</b> (A)	I b/rw/sph = gau	<b>gau</b> /sph/rw	<b>b</b> /rw	b/rw = gau/sph	b/rw/gau/sph
	II <b>b</b> /rw = sph	<b>gau</b> /sph	<b>b</b> /rw/sph	b/sph/gau = rw	b/rw/sph
	III <b>sph</b> /b/gau	<b>gau</b>	<b>b</b>	<b>gau</b> /sph	<b>gau</b> /sph
<b>2</b>	I b/rw/gau	rw/gau-bl/b/ gau	<b>b</b> /rw	rw/b/gau = gau-bl	b/rw/gau
	II b/rw	<b>rw</b> /gau-bl/b	<b>b</b> /rw	<b>rw</b> /b	b/rw
	III <b>b</b> /gau	<b>rw</b>	<b>b</b>	<b>rw</b>	<b>b</b> /gau
<b>3</b>	I <b>b</b> /rw	rw = gau/b	<b>b</b> /rw	rw/b/gau	b/rw/gau
	II <b>b</b> /rw	<b>rw</b> /gau/b	<b>b</b> /rw	<b>rw</b> /b	b = rw
	III <b>b</b>	<b>rw</b>	<b>b</b>	<b>rw</b> /gau	gau/b
<b>4</b>	I rw = b/gau	<b>rw</b> /gau-bl/ gau	b/rw	rw/gau-bl/gau	rw/b/gau
	II b = rw/gau	<b>rw</b> /gau-bl	<b>b</b> /rw	<b>rw</b> /gau-bl/b	<b>rw</b> /b
	III <b>rw</b>	<b>rw</b>	<b>rw</b> /b	<b>rw</b>	<b>rw</b>
<b>5</b> (A)	I rw/b = powa/gau	<b>rw</b> /gau	b/rw	<b>rw</b> /gau/b	b/rw
	II rw/b/powa = sph	<b>rw</b> /gau	<b>b</b> /rw	<b>rw</b>	<b>b</b> /rw
	III <b>rw</b> /gau	<b>rw</b>	<b>rw</b> /b	<b>rw</b>	rw = gau/b
<b>8</b> (A)	I rw/b	<b>rw</b> /gau/ gau-bl	<b>b</b> /rw	rw/gau/b	<b>rw</b> /b
	II <b>rw</b> /b	<b>rw</b>	<b>b</b> /rw	rw/b = gau	<b>rw</b> /b
	III <b>rw</b>	<b>rw</b>	<b>b</b> /rw	<b>rw</b> /gau	<b>rw</b>
<b>9</b>	I rw/b/gau	gau = rw	b/rw	rw/b/gau	rw/b/gau
	II <b>rw</b> /b/gau	<b>rw</b> /gau	<b>b</b> /rw	rw/b/gau	<b>rw</b> /b
	III <b>rw</b>	<b>rw</b>	<b>b</b> /rw	<b>rw</b>	<b>rw</b>

**Tabelle 2b:** Fortsetzung von Tab. 2a

Vers.Nr.	Hanf 54	Roggen 55	Rübe 56	Hafer 57	Hanf 58
6	I gau/sph/rw =b	gau/sph/rw=b	b/gau=sph/rw	sph/gau/ powa	gau/sph=b/pow= rw
(A)	II sph/gau=b	sph/gau	<b>sph/b</b>	<b>sph/gau</b>	sph/gau
	III <b>sph</b>	<b>sph/gau</b>	<b>sph</b>	<b>sph</b>	<b>sph</b>
7	I sph/gau	<b>sph</b> /gau	gau=rw=sph/ pow	gau=sph	sph/gau/pow
(A)	II <b>sph/gau</b>	<b>sph/gau</b>	sph/rw/gau	<b>sph/gau</b>	<b>sph/gau=b</b>
	III <b>sph</b>	<b>sph</b>	<b>sph</b>	<b>sph/gau</b>	<b>sph</b>
10	I gau/pow/ sph/powa	sph/powa/gau	pow/b	sph/pow	pow=sph/ gau=powa
(A)	II gau/pow/ sph	<b>sph/gau=</b> powa	<b>pow/b</b>	<b>sph/pow</b>	sph/gau=pow
	III <b>pow/gau</b>	<b>sph</b>	<b>pow</b>	<b>sph</b>	<b>pow/sph</b>
11	I <b>powa/gau</b>	<b>sph/gau-bl/</b> powa	rw/pow/sph/po wa	<b>sph/gau-bl</b>	powa=sph
(A)	II <b>powa/gau =</b> sph-bl	<b>sph</b>	<b>pow/rw</b>	<b>sph/gau-bl</b>	sph/pow-bl = gau-bl=pow
	III <b>powa</b>	<b>sph</b>	rw/sph/pow	<b>sph</b>	<b>pow/powa/sph</b>
12	I <b>powa/rw</b>	<b>sph/rw/gau</b>	rw/pow/powa/ sph-bl	sph/rw/ powa	powa/gau=rw = pow=sph-bl=sph
(A)	II <b>powa/rw/</b> sph	<b>sph/rw</b>	pow/rw/sph-bl	<b>sph/rw/</b> sph-bl	pow/rw/powa = gau=sph-bl
	III <b>powa</b>	<b>sph</b>	<b>pow/rw</b>	<b>rw/sph</b>	<b>pow</b>

Randomisationspläne im Mittel der Positionen (III) wieder. Auffallend ist, dass in (III) eine starke Konzentration der relativen Häufigkeit auf oft nur ein Modell erfolgt. Da zwischen den Rangmittelwerten der Randomisationspläne praktisch keine Varianz vorliegt, sind diese Modelle auch jene, die im Mittel über alle Positionen und Randomisationen die kleinsten mittleren Rangzahlen haben. Dieses gilt auch für Tab. 2b. Im Gegensatz zu Tab. 2a dominieren bei den in Tab. 2b dargestellten Versuchen nichtlineare räumliche Kovarianzmodelle. Hier sind zum einen die Blocks größer und auch nicht kompakt. Die alleinige Blockbildung ist in keinem Fall ausreichend, die räumliche Heterogenität adäquat zu berücksichtigen. Das Basismodell tritt fast nicht mehr auf, das random walk-Modell selten. Insbesondere bei der Betrachtung über alle Positionen und Randomisationen nimmt bei den Fruchtarten Rübe und Hanf 58 die Anzahl der verschiedenen bestangepassten Modelle zu. In den seltensten Fällen ist ein nichtlineares Modell, das nur die Kovarianz innerhalb der Blocks berücksichtigt, in mehr als 10 % der Fälle das beste Modell. Verstärkt tritt in Tab. 2b mit Zunahme der Versuchsgröße das anisotrope Modell in Erscheinung. Auffallend ist - insbesondere bei den Getreidearten - die durchgängige Ähnlichkeit der Modellbevorzugung.

Stellt man die je Position besten Modelle räumlich dar, so ist ein räumliches Muster zu erkennen. In Abb. 3 (links) ist beispielhaft für den Versuch 4 Hanf 54 der räumliche

Bezug hergestellt. Hier sind an den 30 Positionen in je 44.3% das Basis- und das random walk-Modell die bestangepassten, das Gauss-Modell ist an 10% der Positionen vertreten (vgl. Tab. 2a). Das Modell gau-bl ist nur an einer Position (<10%) das beste und deshalb nicht in Tab. 2a aufgeführt. Dass in x-Richtung räumliche Cluster gleicher Modelle existieren, ergibt sich zum Teil aus den existierenden Überschneidungen zwischen den Versuchen. In y-Richtung gibt es diese Überschneidungen jedoch nicht, die Clusterung ist allein aus der räumlichen Korrelation in dieser Richtung zu erklären.

### 3.2 Vergleich der Modellauswahl nach AIC- und BIC-Kriterium

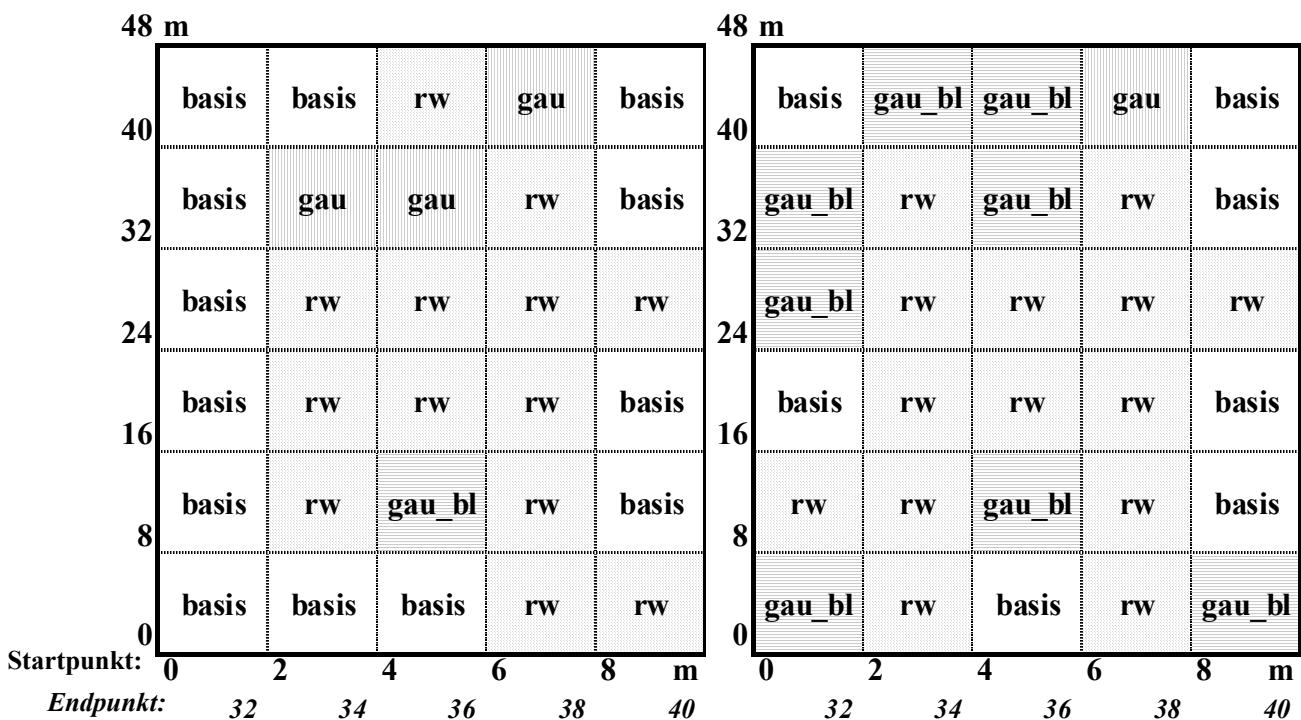
Die prinzipiellen Konsequenzen des Wechsels vom AIC- zum BIC-Kriterium können anhand der in PROC MIXED vorgenommenen Berechnung der Kriterien hergeleitet werden:

$$AIC = - 2 \text{ Res loglikelihood} + 2p$$

$$BIC = - 2 \text{ Res loglikelihood} + p \cdot \ln(k),$$

mit  $k = (v-1)(n-1) = \text{FG Rest}$ , wenn *Anzahl subjects* = 1

und  $k = \text{Anzahl subjects}$ , wenn *Anzahl subjects* > 1.



**Abbildung 3:** Räumliche Anordnung der je Position bestangepassten Modelle – Versuch 4 Hanf 54. links: nach AIC-Kriterium. rechts: nach BIC-Kriterium

Für die untersuchten Versuchspläne ergibt sich damit:

Wird mit `subject = intercept` gearbeitet, ist *Anzahl subjects* = 1 und  $\ln(k) > 2$ . Wird mit `subject = block` gearbeitet, dann ist *Anzahl subjects* = n (hier: 4) > 1 und  $\ln(k) < 2$ . Das bedeutet, dass alle mit `subject = intercept` arbeitenden Modellansätze mit dem BIC-Kriterium stärker bestraft werden als solche, die mit `subject = block` arbeiten. Damit ist zu erwarten, dass durch den Wechsel zum BIC-

Kriterium die Modelle sph-bl, pow-bl, gau-bl und rw häufiger als bestangepasste ausgewählt werden. Die Gegenüberstellung in Abb.3 demonstriert dieses.

### 3.3 Vergleich beobachtete und geschätzte Standardfehler

Beispielhaft sind in Tab. 3 die Verzerrungen der geschätzten Standardfehler in Prozent der beobachteten Standardfehler für die Versuche 4 und 12 im Mittel aller Prüfgliedvergleiche dargestellt. Alle Auswertungsmodelle, die den Power-Ansatz enthalten, haben ohne Korrektur Verzerrungen, die denen anderer Modelle vergleichbar sind, bei der KR-Methode sind sie jedoch teilweise größer als  $10^6$  % und deshalb in Tab. 3 nicht aufgeführt. Damit bestätigen sich die Aussagen in Hu et al. ([4]), dass im Falle des Power-Modells ohne Nugget-Varianz die Verzerrungen extrem groß sind. Während sich bei Hu et al. ([4]) dieses auch für das sphärische Modell zeigte, konnte hier keine einheitliche Tendenz festgestellt werden. Beim Gauss-Modell wurde die starke Verzerrung (Unterschätzung) durch die Korrektur verringert. Das beim Power-Modell geschilderte Problem ist eng mit dem Konvergenzverhalten des Modells im Randbereich verknüpft: Wenn keine räumliche Abhängigkeit besteht, geht dieses Modell - im Gegensatz zum sphärischen und Gauss-Modell - nicht automatisch in das Basismodell über. Besonders große geschätzte Standardfehler wurden in den Fällen beobachtet, wo die Korrelation benachbarter Parzellen  $< 0.2$  oder  $> 0.9999$  ist (das entspricht einem praktischen Range von  $< 2$  bzw.  $> 29956$  m). Beim Ausschluss dieser Fälle nimmt die Verzerrung zwar ab, liegt aber in den meisten Fällen noch in der Größenordnung von  $10^4$ %. Insbesondere trat dieses Problem bei Winterroggen auf, wo teilweise nur in 30% der insgesamt 30.000 bzw. 33.000 Auswertungsläufe eine Schätzung der Varianz-Kovarianz-Parameter erfolgte.

Generell ist beim Basismodell eine sehr kleine Verzerrung zu beobachten. Die bei den anderen Modellen allgemein erwartete Unterschätzung zeigte sich für Hanf 54, Rübe und Hanf 58; bei Roggen und Hafer liegt eine Tendenz zur Überschätzung vor. Wir konnten nicht feststellen, dass die Vorzüglichkeit eines Modells an einer Position mit einer dort geringeren Verzerrung verbunden ist.

**Tabelle 3:** Verzerrung des geschätzten Standardfehlers in Prozent des beobachteten Standardfehlers im Mittel aller Prüfgliedvergleiche (ohne Korrektur (-) und mit Korrektur nach Kenward-Roger (KR))

Vers. Nr.	Modell	Korr.	Hanf 54	Roggen 55	Rübe 56	Hafer 57	Hanf 58
4	basis	-	-0.43	0.82	0.28	-1.00	0.14
	rw	-	-2.97	10.50	-6.76	7.85	-3.63
	pow <sup>1)</sup>	-	-4.55	3.22	-6.68	7.47	-4.16
	sph	-	-7.04	6.86	-9.84	3.00	-7.32
		KR	-6.59	6.87	-9.58	4.48	-5.30
	gau	-	-16.80	-14.78	-16.05	-19.69	-16.05
		KR	6.92	9.64	8.97	2.54	8.63
	pow-bl <sup>1)</sup>	-	-2.62	-3.59	-4.87	-1.63	-2.52
12	sph-bl	-	-5.65	6.90	-8.95	5.42	-6.18
		KR	-6.71	-1.60	-10.86	-0.20	-8.10
	gau-bl	-	-18.04	-21.55	-16.65	-20.44	-17.35
		KR	8.96	-3.55	12.42	-2.78	8.77
	basis	-	0.26	0.31	-0.01	0.66	-0.02
	rw	-	-2.63	18.07	-9.06	9.94	-4.45
	pow <sup>1)</sup>	-	0.80	17.25	-4.25	9.43	0.09
	powa <sup>1)</sup>	-	-4.75	11.69	-7.81	3.49	-5.25
sph	-	-3.01	15.33	-8.38	7.66	-4.00	
	KR	-2.80	15.07	-8.05	8.53	-3.41	
gau	-	-12.82	-28.03	-11.67	-19.63	-11.50	
	KR	-7.65	-26.36	-6.03	-16.65	-5.79	
pow-bl <sup>1)</sup>	-	1.25	20.97	-4.26	9.30	-0.06	
sph-bl	-	-3.60	18.43	-9.93	9.39	-5.06	
	KR	-5.34	15.80	-11.12	6.99	-5.64	
gau-bl	-	-12.55	-27.02	-11.64	-19.12	-11.43	
	KR	-7.38	-25.34	-6.00	-16.13	-5.72	

<sup>1)</sup> Bei diesen Modellen ist mit  $ddf_m = kr$  die Verzerrung  $> 10^3\%$ , teilweise  $> 10^6\%$ .

### 3.4 Empirischer Typ I-Fehler des F- und t-Tests

In Tab. 4 sind für die beiden obigen Versuche die empirischen Typ I-Fehler für den F- und t-Test bei einem nominalen Typ I-Fehler von 0.05 angegeben. Wird der nominale Fehler unterschritten, so führt dieses zu einer konservativen, bei Überschreitung zu einer anti-konservativen Testentscheidung. Generell wird im Basismodell der nominale Fehler relativ gut eingehalten, bei Roggen und Hafer im Versuch 4 etwas stärker unterschritten. In Abhängigkeit davon, ob eine positive oder negative Verzerrung des geschätzten Standardfehlers vorliegt, sind die empirischen Fehler im random walk-Modell zu niedrig oder zu hoch. Beim sphärischen Modell liegen die empirischen Fehler nach KR-Korrektur stets über denen nach SW korrigierten. Dieses kann zu einer Verbesse-

rung führen, hier aber in der Mehrzahl zu einer Verschlechterung. Insbesondere im Versuch 4 führt die KR-Korrektur im Modell sph-bl zu erheblichen Überschreitungen.

**Tabelle 4:** Empirischer Typ I-Fehler für den F-Test und den t-Test im Mittel aller Prüfungsgliedvergleiche (nominaler Typ I-Fehler = 0.05) im Versuch 4 und 12

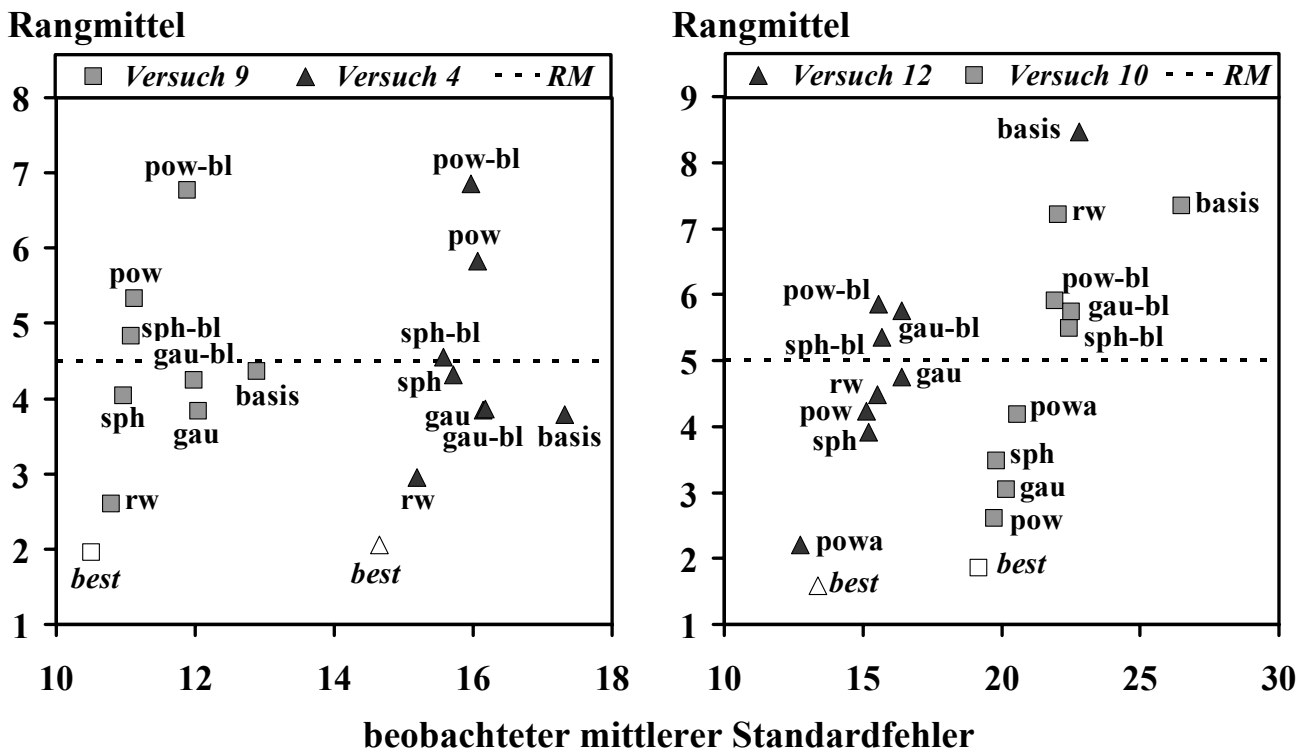
Vers. Nr.	Modell	Korr.	Hanf 54		Roggen 55		Rübe 56		Hafer 57		Hanf 58	
			F	t	F	t	F	t	F	t	F	t
4	basis		.048	.049	.035	.042	.045	.049	.030	.038	.043	.047
	rw		.066	.059	.022	.028	.081	.068	.026	.032	.062	.058
	pow	SW	.059	.059	.035	.040	.067	.063	.024	.030	.055	.056
		KR	.094	.092	.048	.058	.091	.089	.059	.066	.067	.065
	sph	SW	.074	.067	.028	.033	.088	.075	.031	.037	.069	.065
		KR	.107	.094	.047	.051	.135	.113	.068	.070	.104	.086
	gau	SW	.098	.085	.129	.099	.115	.092	.138	.113	.113	.090
		KR	.071	.070	.086	.091	.087	.078	.096	.106	.090	.078
	pow-bl	SW	.046	.051	.039	.051	.053	.056	.035	.040	.044	.049
		KR	.048	.044	.023	.024	.043	.041	.018	.016	.042	.040
sph-bl	SW	.064	.062	.027	.033	.077	.070	.027	.034	.060	.060	
	KR	.130	.125	.102	.123	.157	.146	.084	.092	.151	.146	
gau-bl	SW	.074	.075	.101	.090	.091	.080	.097	.091	.084	.076	
	KR	.060	.057	.093	.075	.073	.061	.089	.075	.068	.059	
12	basis		.045	.048	.042	.048	.044	.048	.043	.048	.043	.049
	rw		.057	.056	.003	.019	.118	.075	.018	.034	.066	.060
	pow	SW	.038	.047	.003	.020	.069	.060	.012	.032	.034	.047
		KR	.110	.115	.056	.083	.025	.031	.135	.155	.030	.038
	powa	SW	.068	.059	.015	.029	.084	.069	.020	.040	.051	.059
		KR	.015	.020	.062	.073	.020	.022	.039	.041	.011	.016
	sph	SW	.061	.057	.005	.023	.109	.073	.017	.035	.065	.059
		KR	.069	.059	.009	.025	.123	.074	.028	.037	.077	.061
	gau	SW	.119	.087	.463	.175	.116	.084	.212	.117	.094	.082
		KR	.161	.077	.576	.174	.130	.071	.278	.111	.117	.070
pow-bl	SW	.035	.046	.002	.016	.068	.060	.015	.034	.034	.047	
	KR	.060	.067	.113	.123	.031	.038	.150	.157	.012	.021	
sph-bl	SW	.061	.058	.003	.019	.120	.077	.019	.035	.069	.062	
	KR	.094	.074	.016	.030	.163	.093	.047	.051	.099	.073	
gau-bl	SW	.114	.086	.439	.166	.115	.084	.204	.114	.093	.082	
	KR	.159	.076	.561	.166	.130	.071	.272	.109	.117	.070	

Während im Versuch 4 die KR-Korrektur des Gauss-Modells zu einer geringeren Überschätzung führt, wird im Versuch 12 insbesondere bei den Getreidearten die ohnehin äußerst starke Überschätzung nach der SW-Methode noch weiter verstärkt. Betrachtet

man nur die Korrekturen nach SW, sowie das Basis- und random walk-Modell, so wird fast durchgängig im t-Test der nominale Fehler besser als im F-Test eingehalten.

### 3.5 Vergleich ausgewählter Versuchsanlagen

In Abb. 4 werden für Hanf 54 und ausgewählte Versuchspläne die je Modell beobachteten Standardfehler im Mittel aller Prüfgliedvergleiche in Beziehung zu den entsprechenden Rangmittelwerten über alle Kombinationen von Position und Randomisation dargestellt. Auf Grund der Verzerrung wurde auf die Darstellung der Beziehung zu den geschätzten Standardfehlern verzichtet.



**Abbildung 4:** Hanf 54, ausgewählte Versuchspläne: beobachteter Standardfehler im Mittel aller Prüfgliedvergleiche und Rangmittelwert über alle Kombinationen von Position und Randomisation für die jeweiligen Auswertungsmodelle. RM: Gesamtrangmittel über alle 8 bzw. 9 Modelle.

Die Versuche 4 und 9 haben jeweils 4 Prüfglieder, gleiche Blockdimensionen und zeigen nach Tab. 2a eine ähnliche Modellbevorzugung. Sie unterscheiden sich nur in der Richtung der Teilstückbildung. Ausgehend vom Basismodell, das jeweils den höchsten beobachteten Standardfehler hat, ist die Situation im Versuch 4 ungünstiger als im Versuch 9, da hier ein größerer Teil der Heterogenität innerhalb der Blocks liegt (vgl. Abb.1). Durch die Nutzung räumlicher Modelle verbessert sich die jeweilige Situation, jedoch nicht so weit, dass dadurch die beobachteten Standardfehler des Versuchs 4 in die Größenordnung des Versuchs 9 kommen. Dass das Basismodell dennoch im Rangmittel, das auf den AIC-Werten basiert, besser als einige räumliche Modelle abschneidet, ist teilweise im kleineren Strafterm bei der Berechnung des AIC-Wertes begründet. Das random walk-Modell hat in beiden Fällen die kleinste mittlere Rangzahl und führt

zum kleinsten mittleren Standardfehler. Generell ist aber keine strenge Relation zwischen den Rangmitteln und den Standardfehlern zu erwarten. Wird je Position  $x$  Randomisation das bestangepasste Modell (best) gewählt, so ist eine kleine Verbesserung möglich.

Die Versuche 10 und 12 haben je 10 Prüfglieder und unterscheiden sich im Wesentlichen in der Teilstückgröße. In beiden Fällen sind die Blocks so angelegt, dass ein großer Teil der Heterogenität innerhalb der Blocks liegt. Dieses äußert sich in den hohen beobachteten Standardfehlern der Basismodelle, wobei der Versuch 10 mit der kleineren Teilstückgröße ungünstiger ist - bei größeren Teilstücken liegt ein größerer Varianzanteil innerhalb der Teilstücke. Das Basismodell hat jeweils den höchsten Rangmittelwert. Aus der Sicht der beobachteten Standardfehler ist das isotrope bzw. anisotrope Power-Modell das jeweils beste (vgl. Tab. 2b); sie haben auch den jeweils kleinsten Rangmittelwert. Im Versuch 10 verringern sich die Standardfehler bei der räumlichen Auswertung so stark, dass sie kleiner werden als der des Basismodells im Versuch 12. Dennoch bleibt auch bei den räumlichen Analysen die ungünstigere Situation des Versuchs 10 gegenüber dem Versuch 12 erhalten. Im Versuch 12 ist zu erkennen, dass das je Kombination bestangepasste Modell nicht unbedingt an den kleinsten mittleren Standardfehler gekoppelt ist.

## 4 Abschließende Bemerkungen

In unserem Blindversuch zeigte sich, dass durch die existierende räumliche Korrelation in der Regel die räumlichen Modelle besser als das Basismodell angepasst sind. Ausnahmen existieren nur, wenn die Blocks sehr klein sind. Bei kleinen Blocks ist häufig die Berücksichtigung des linearen Verhaltens des Variogramms für kleine Distanzen ausreichend; je größer die Blocks und der Versuch werden, umso eher sind nichtlineare Kovarianzmodelle besser angepasst. Dabei ist zu beachten, dass bei Nutzung des AIC anstelle des BIC-Kriteriums generell das Basismodell und Modelle, die die Kovarianz über die Fläche des Versuchs berücksichtigen, bevorzugt werden. Während die Güte der Anpassung jedes Modells vorwiegend durch die Position bestimmt wird, ist die Vorzüglichkeit eines Modells gegenüber den anderen (gemessen an den rangierten AIC-Werten je Kombination von Position und Randomisationsplan) vorwiegend durch eine spezifische Reaktion der Randomisationspläne an den einzelnen Positionen bedingt. Im Mittel der Positionen zeigen die Randomisationspläne, sowohl in der Modellanpassung als auch in der relativen Vorzüglichkeit im Prinzip keinen Unterschied. Bei welchem Auswertungsmodell welche Randomisationspläne der jeweiligen Position am besten angepasst sind, wurde hier nicht untersucht. Diese Problemstellung beinhaltet die gezielte Wahl des Randomisationsplans in Abhängigkeit von der existierenden Kovarianzstruktur auf einer ausgewählten Fläche. Die Ergebnisse zur Verzerrung der geschätzten in Relation zu den beobachteten Standardfehlern zeigen sehr widersprüchliche Ergebnisse. Die KR-Methode ist in den Power-Modellen ohne Nugget-Varianz in jedem Fall ungeeignet. Sowohl SW- als auch KR-Methode zeigen bei der Einhaltung des nominalen Typ I-Fehlers teilweise erhebliche Defizite. Eine Ursache kann sein, dass die Versuche noch relativ klein sind. Bei Hu et al. ([4]) wurde für den t-Test anhand von



simulierten Daten bei bekannter Kovarianzstruktur gezeigt, dass bei größeren Prüfgliedanzahlen (Blockgrößen) die Übereinstimmung größer ist. Eine andere Ursache könnte darin liegen, dass die mit PROC PLAN erzeugten Versuchspläne eigentlich dem Basismodell Rechnung tragen und nicht auf die Spezifik der wirklichen (hier unbekannt) Kovarianzstruktur abgestimmt sind.

## **Literatur**

- [1] M.S. Bartlett: The approximate recovery of information from replicated field experiments with large blocks. *Journal of Agricultural Science*. 1938. 28, 418-427.
- [2] G. Bätz: Untersuchungen zur Erhöhung der Aussagekraft von Feldversuchen. 1968. Habilitation. Friedrich-Schiller-Universität Jena.
- [3] R.A. Fisher: *The design of experiments*. 1935. Oliver and Boyd.
- [4] X. Hu, J. Spilke & C. Richter: The influence of spatial covariances on the type I error and the power for different evaluation models. *Biometrical Letters*. 2006, 43 (1), 19-37.
- [5] R.C. Littell, G.A. Milliken, W.W. Stroup, R. D. Wolfinger & O. Schabenberger: *SAS for mixed models*. 2nd. Edition. SAS Institute Inc., Cary, NC, USA, 2006.
- [6] J.S. Papadakis: Méthode statistique pour des expériences sur champ. *Bull. Inst. Amél. Plantes à Salonique*, 1937. No. 23.
- [7] W. Pilarczyk: The extent and prevailing shape of spatial relationship in Polish variety testing trials on cereals. *Proceedings of the International Symposium "Agricultural Field Trials – Today and Tomorrow"*. Stuttgart. Grauer-Verlag. 2007, 153-159.
- [8] H.-P. Piepho, C. Richter & E. Williams: Nearest neighbour adjustment and linear variance models in plant breeding trials. *Biometrical Journal*. 2008, 50 (2), 164-189.
- [9] C. Richter, I. Kuzyakova & B. Kroschewski: Trendidentifikation und Heterogenität in beprobten Flächen. *Zeitschrift für Agrarinformatik*. 1997, 5 (3), 60-7.
- [10] C. Richter, B. Kroschewski, V. Michel & A. Zenk: Analysis of standard designs using spatial models. *Proceedings of the International Symposium "Agricultural Field Trials – Today and Tomorrow"*. Stuttgart. Grauer-Verlag. 2007, 176-181.
- [11] J. Spilke, C. Richter: Comparison of the effectiveness of model selection methods in the presence of spatial covariance. *Proceedings of the International Symposium "Agricultural Field Trials – Today and Tomorrow"*. Stuttgart. Grauer-Verlag. 2007, 221-226.
- [12] W.W. Stroup: Power analysis based on spatial effects mixed-models: a tool for comparing design and analysis strategies in the presence of spatial variability. *Journal of Agricultural, Biological, and Environmental Statistics*. 2002, 7(4), 491-511.

- [13] G.A. Wiebe: Variation and correlation in grain yield among 1,500 wheat nursery plots. *Journal of Agricultural Research*. 1935, 50, 331-357.
- [14] F. Yates: The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments. *Biometrika*. 1938, 30, 440-466.
- [15] D.L. Zimmerman, D.A. Harville: A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics*. 1991, 47, 223 – 239.