

Von Streifenmittelwerten zu Punktdaten

Effizienzgewinne durch geostatistische Modellierung

Damon Raeis-Dana

BASF Digital Farming / xarvio

Die große Versprechung

Die große Versprechung georeferenzierter Ertragskarten:

- hohe räumliche Auflösung
- viele Daten = viel Information
- Vorliegen als digitale Datei = schnelle automatisierte Auswertung
- bessere Grenzdifferenzen - einfach(er) zum Ziel
“Signifikanz”?

Die Ernüchterung?

- *out of the box* kaum zu gebrauchen
 - Kalibration, Bereinigung, ...
- viele *räumlich autokorrelierte* Daten
 - effektives n häufig viel kleiner
- Selten vollständig automatisierbar
- **Aber:**
 - im Idealfall *statistisch effizienter*

1 Der Versuch

Versuchsfläche

- Feld um Würzburg
- Variabler Aussaatversuch im **Silo-Mais**
- Aussaat 12.04.2025
- Ernte 17.09.2025



Hypothese

*Das Erhöhen der **Aussaatrate** propotional zum **Ertragspotential** steigert die Erträge im Silomais signifikant verglichen mit einer konstanten Aussaatrate.*

Ertragspotential	Körner pro qm
Niedrig	7.6
Mittel	9.0
Hoch	10.4

Der Versuch

Setting

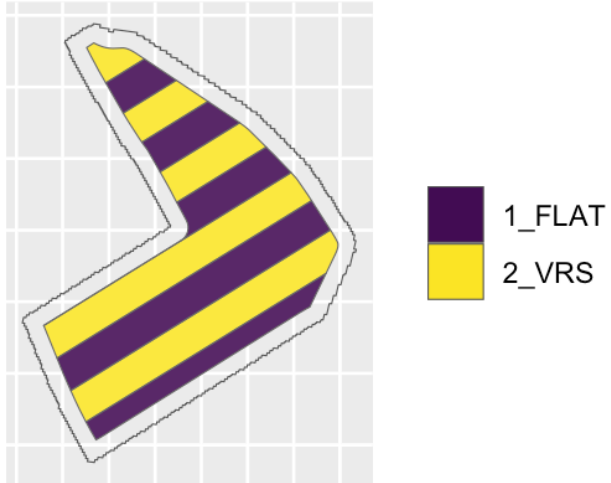
- Effektive Feldgröße: 12.7 ha
- Streifenanlage mit 10 Streifen
- 2 Behandlungen - VRS vs FLAT
- 5 Wiederholungen
- Geblockt mit systematischem bzw alternierendem Design

Technik

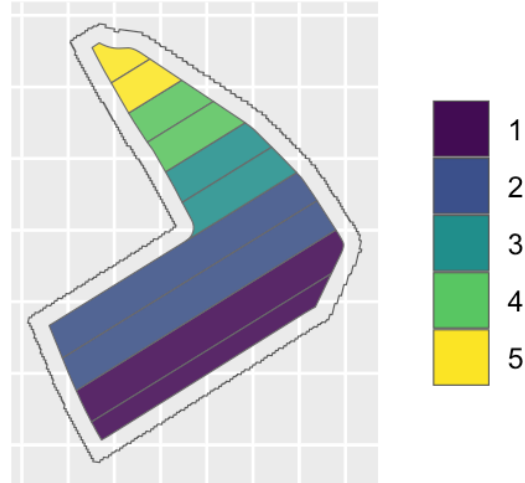
- Streifenbreite 54m
- Häcksler mit 9m
- 6 Häckslerdurchfahrten pro Versuchsstreifen

Versuch - visuell

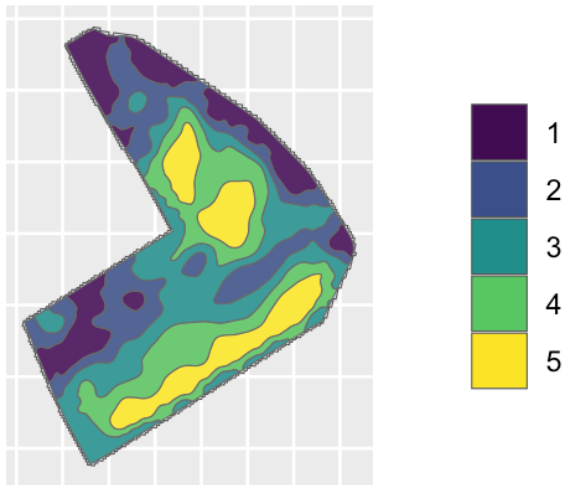
Trial - Treatment No.



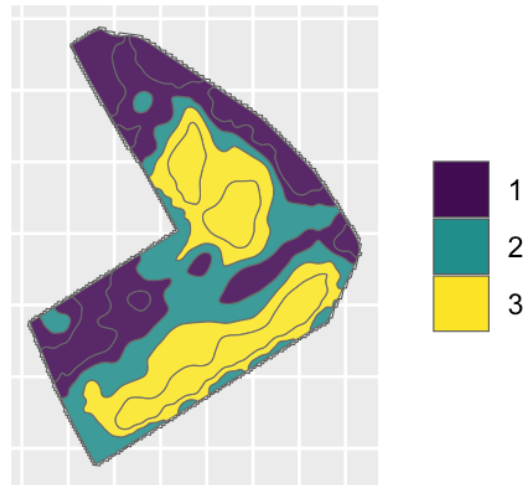
Trial - Replicate No.



Power Zones (5 Zones)



Power Zones (3 Zones)



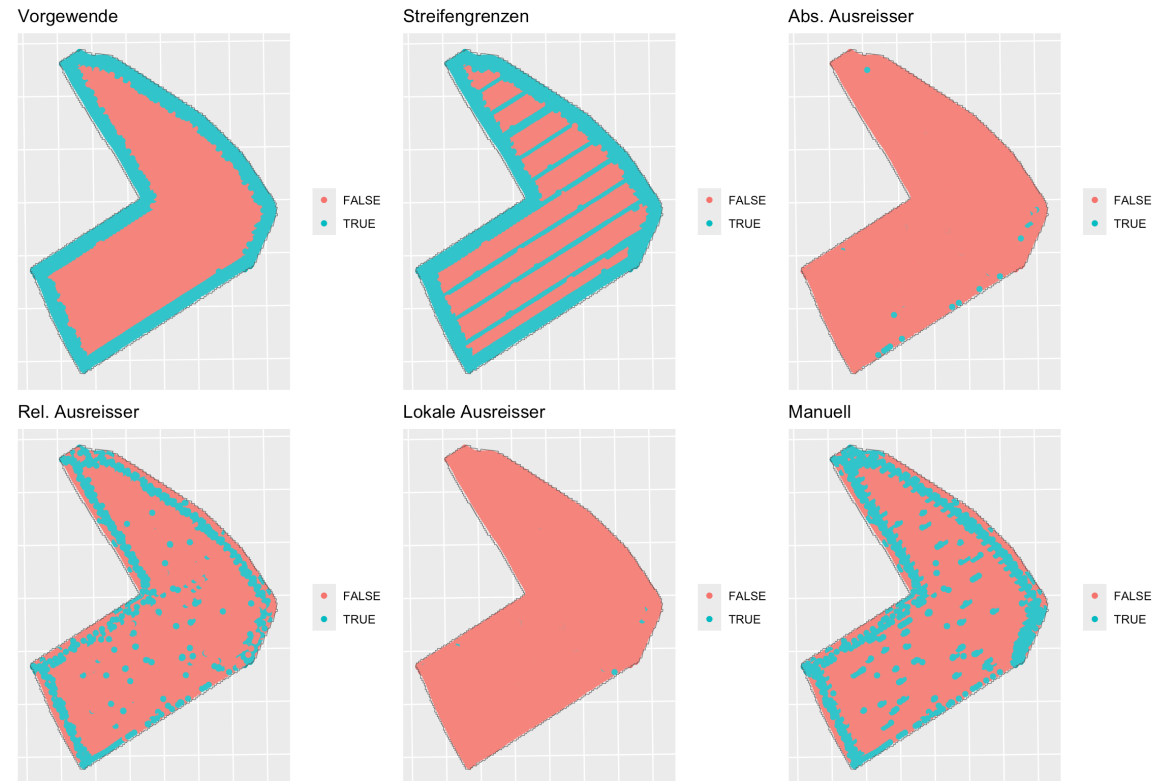
Randomisierung?

Offensichtlicher Einwand - warum nicht randomisiert?

- “Landwirt wollte es so”
- Keine Versuche oder variable Applikation in Vorjahren
- Kein Trend parallel zur Streifenanlage
- Kein Confounding mit Maschineneffekten
 - 6 Überfahrten pro Streifen, angrenzende entfernt
 - 4 effektive Überfahrten - 2 pro Richtung

Bereinigung

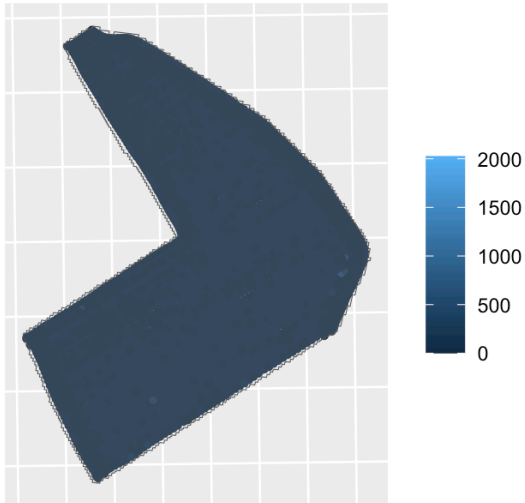
- Aussaat verhältnismäßig präzise
- Keine “Kontamination” zwischen Streifen
- Entfernen angrenzender Überfahrten, um Vermischung beim Ernteprozess auszuschließen



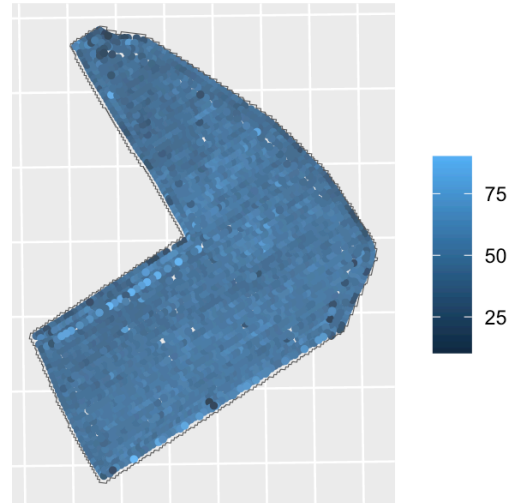
Ertragskarte

- Häcklser produzieren oft exzessiv hohe und niedrige Werte
- *Im Durchschnitt* korrekt
- Produziert aber unnötig breite Verteilungen bzw Standardabweichungen

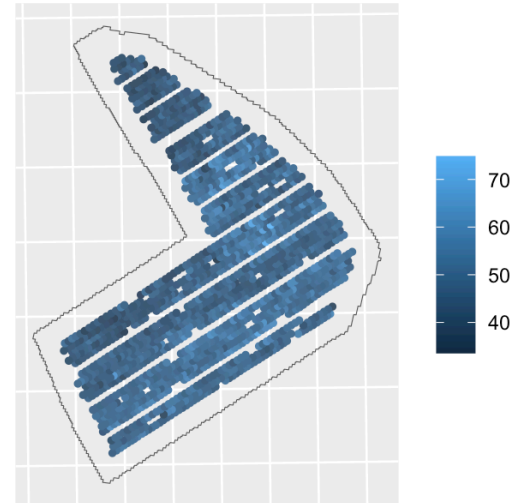
Ertrag - Roh



Ertrag - ohne min/max 1 %

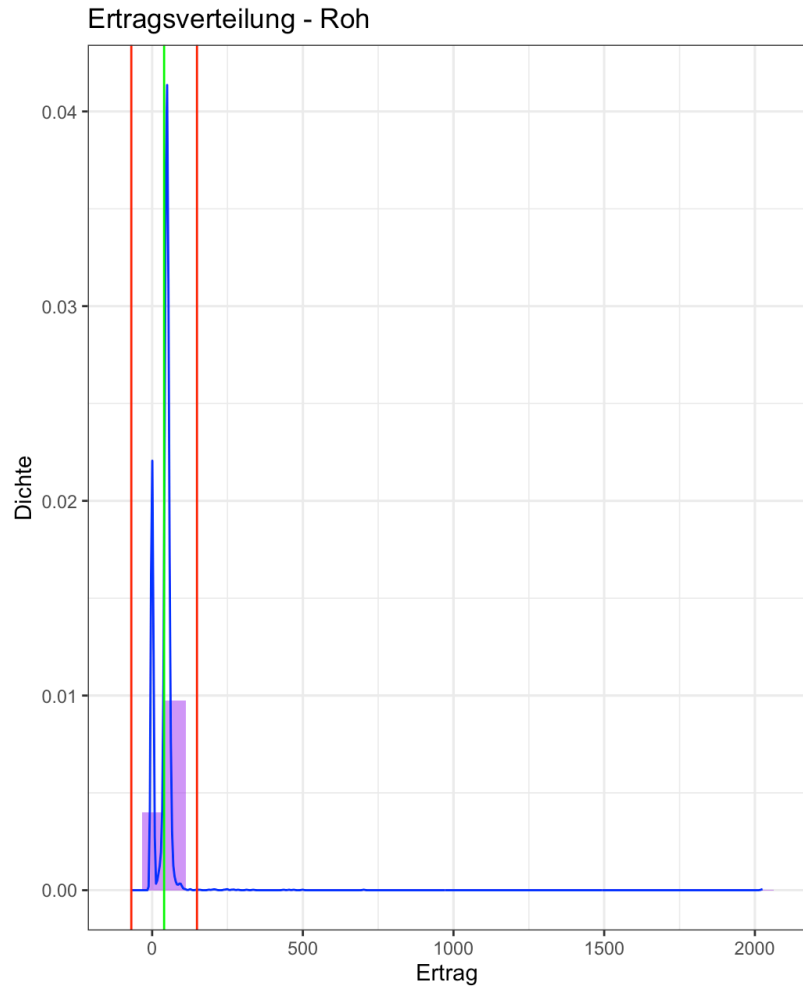


Ertrag - bereinigt

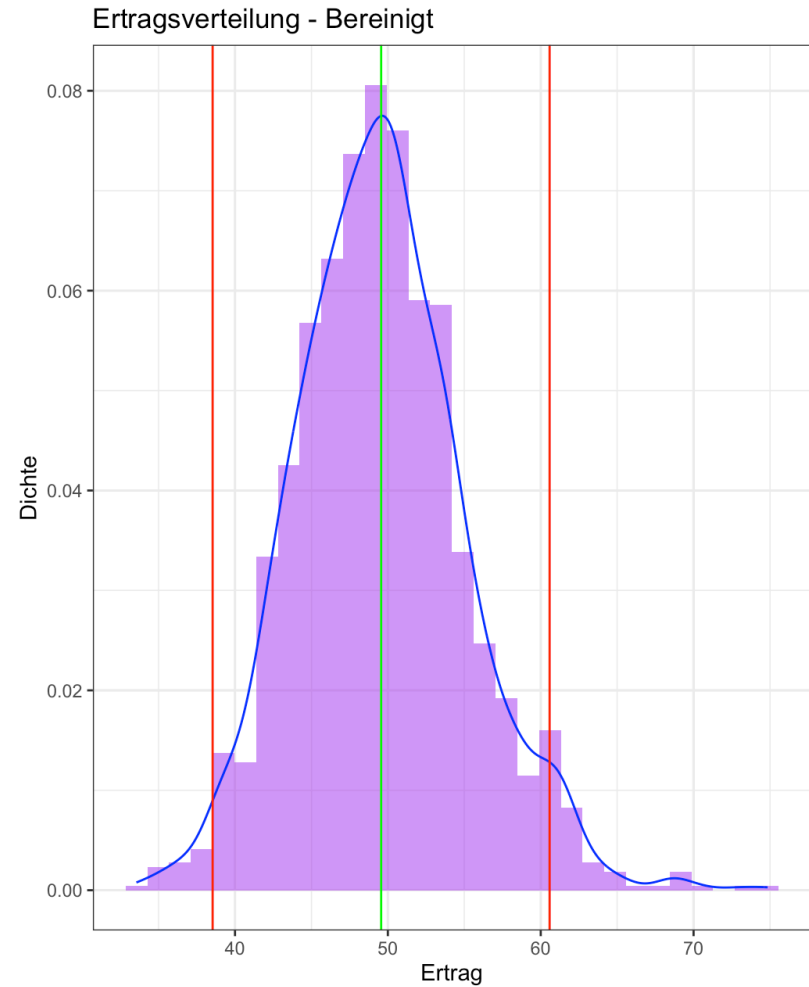


Ertragsverteilung

Konventionell normalverteilt (ohne Test).



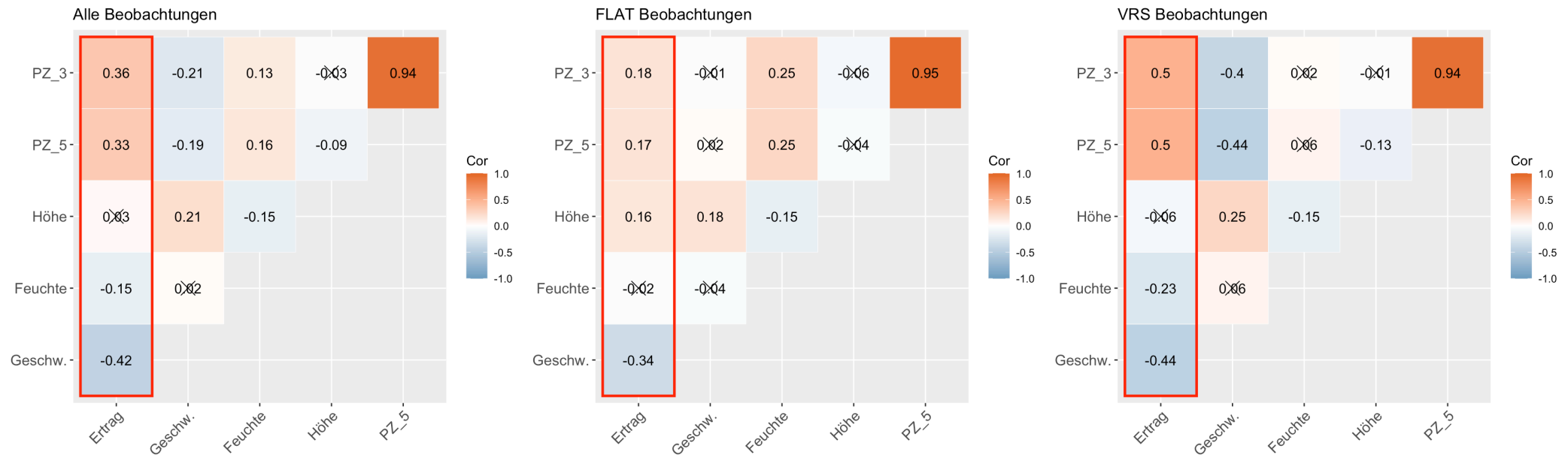
Copyright xarvio



Copyright xarvio

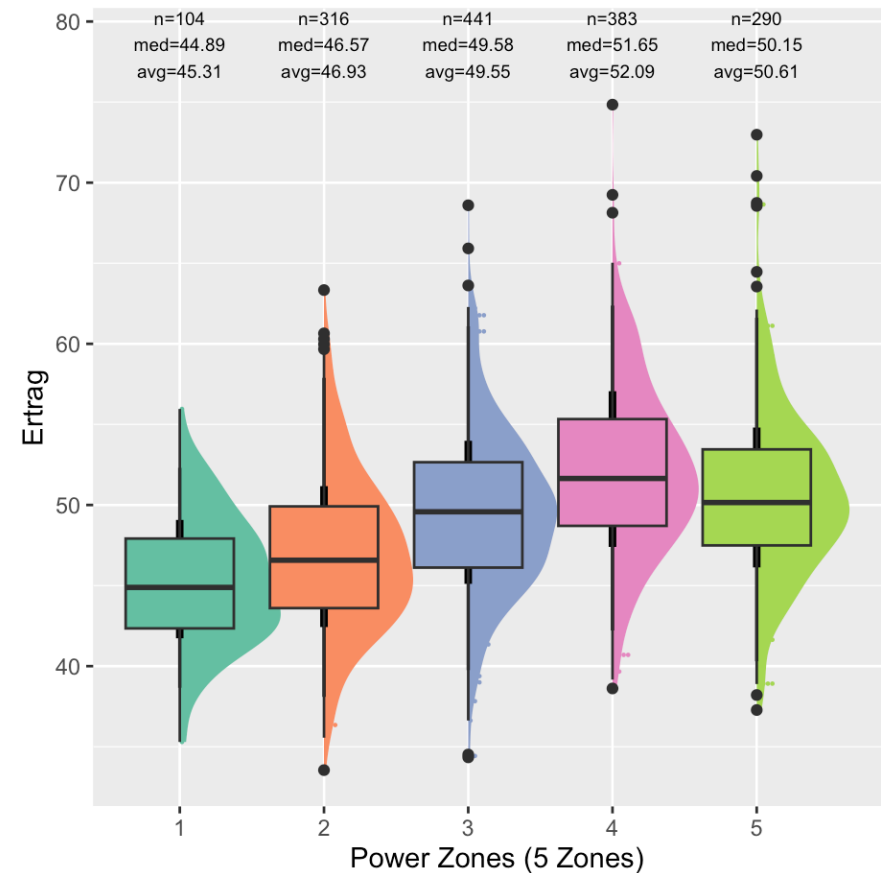
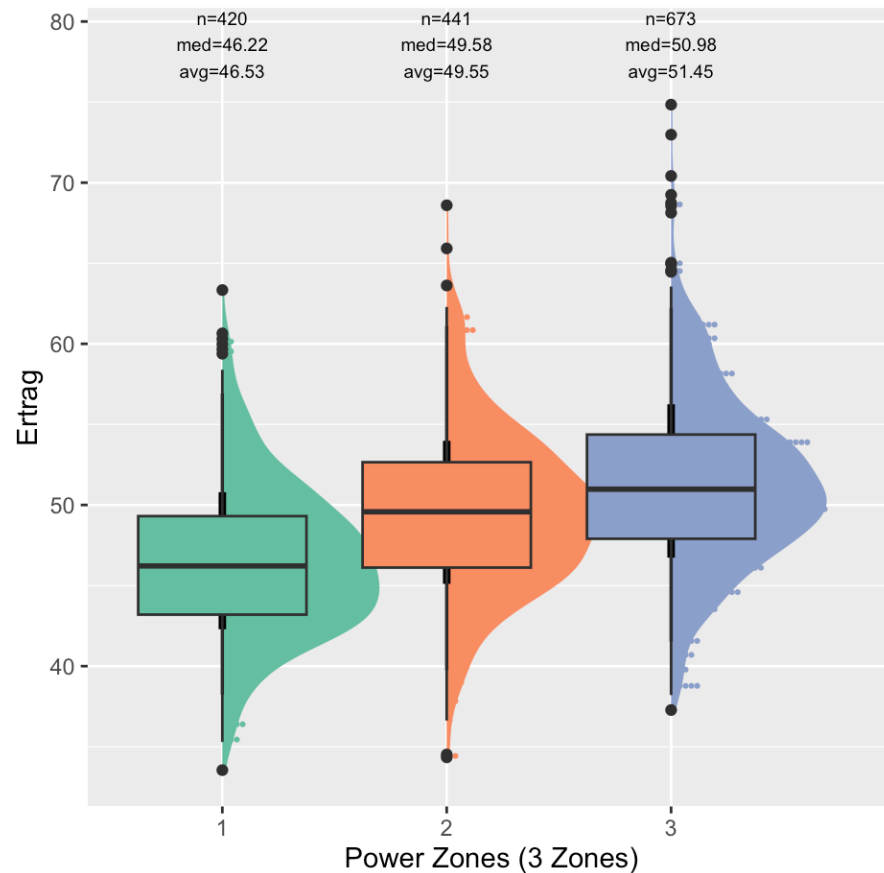
Korrelationen

- Lineare Pearson-Korrelationen
- VRS verstärkt - konstante Rate homogenisiert?



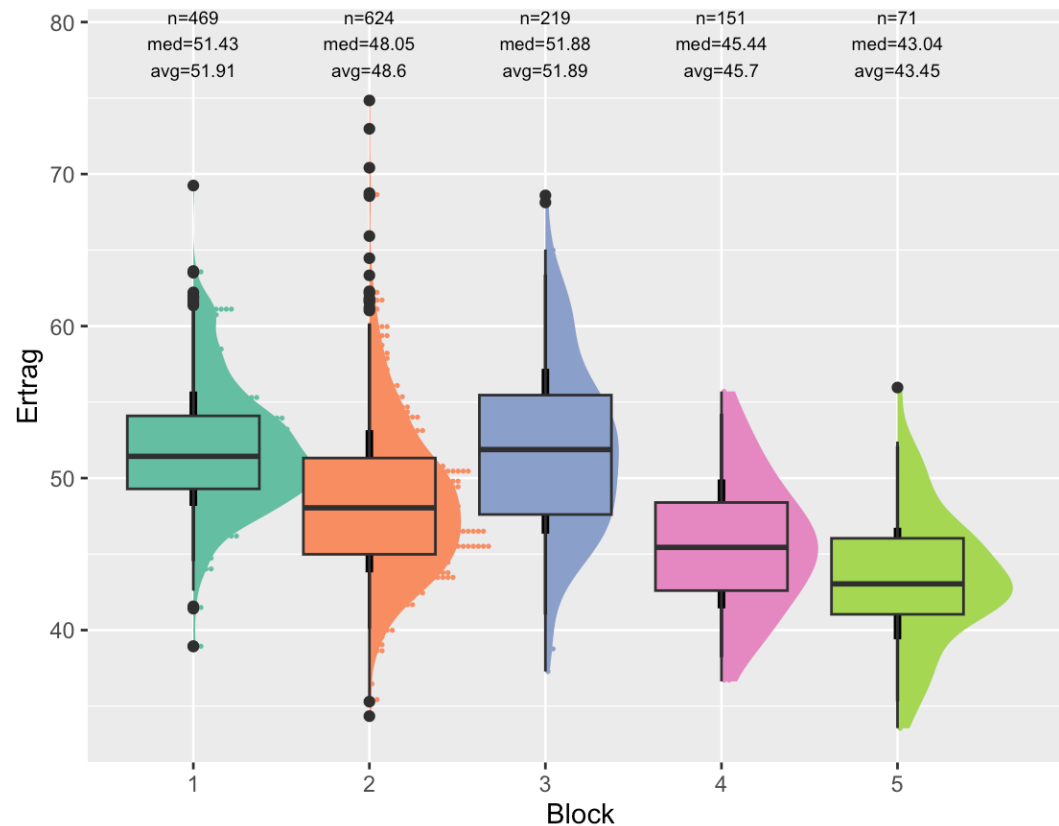
Boxplots - Haupteffekte

- Sättigung in hohen Zonen
- Differenzierung von etwa 3 t/ha in 3-Zonen-Modell

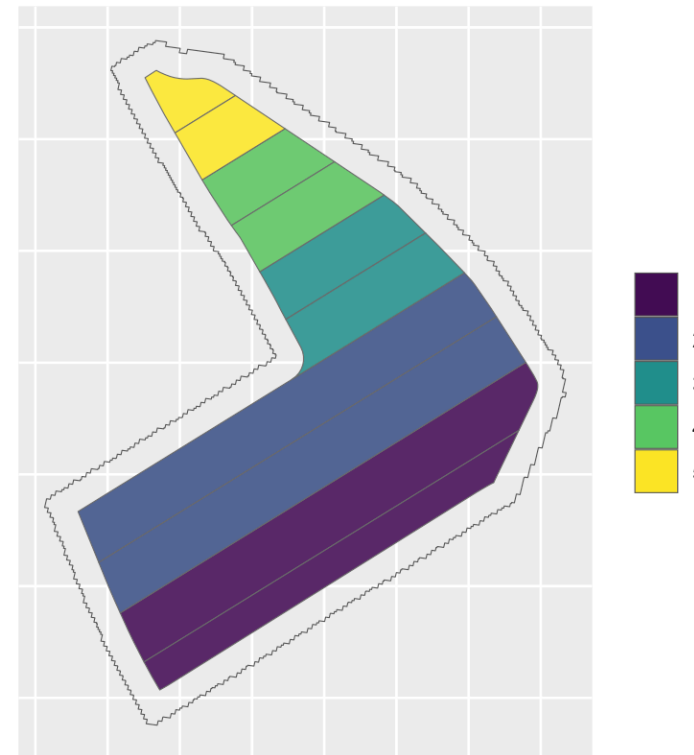


Boxplots - Haupteffekte II

- Nord-Süd-Trend?
- immer kleineres n

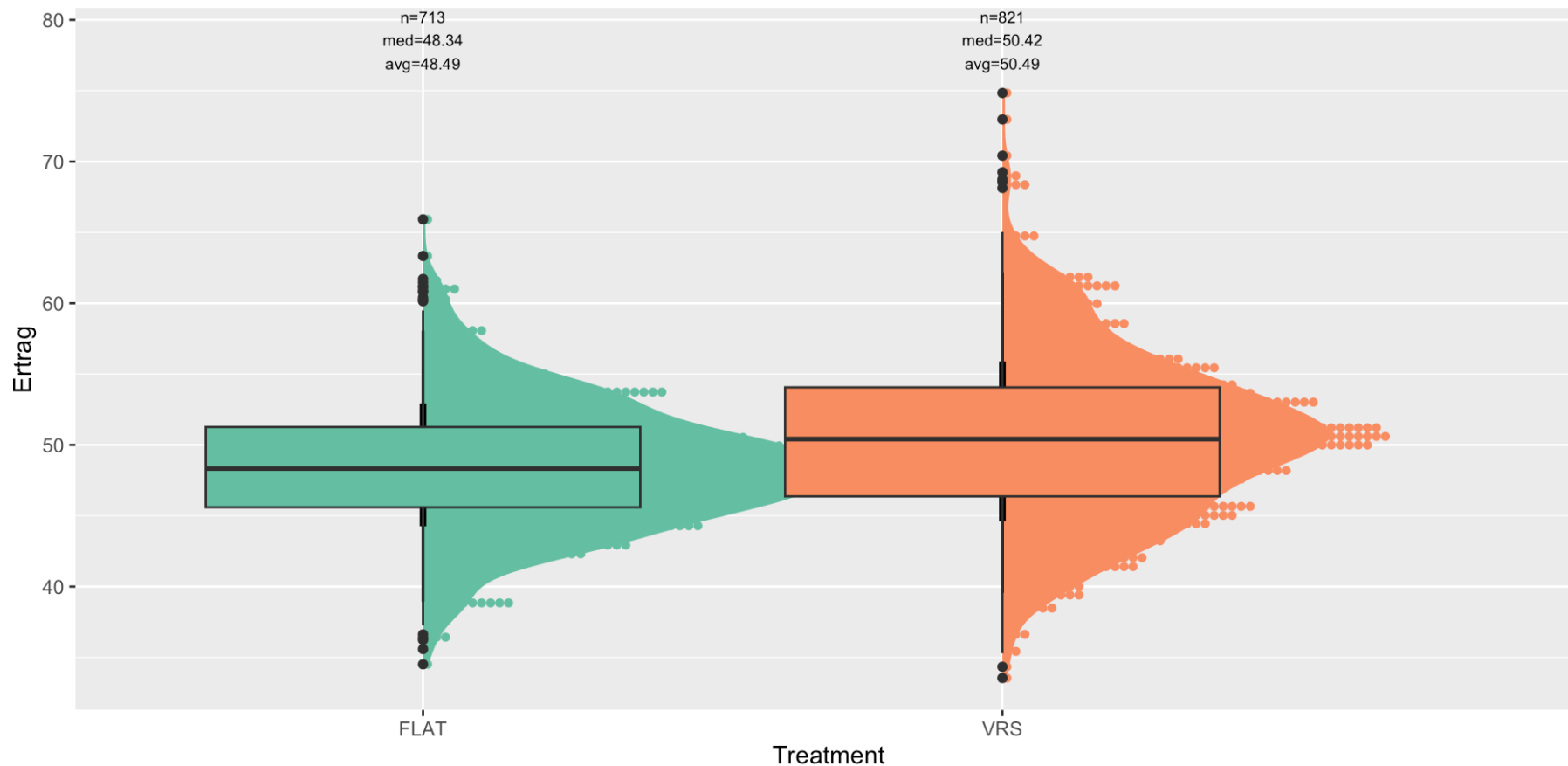


Trial - Replicate No.

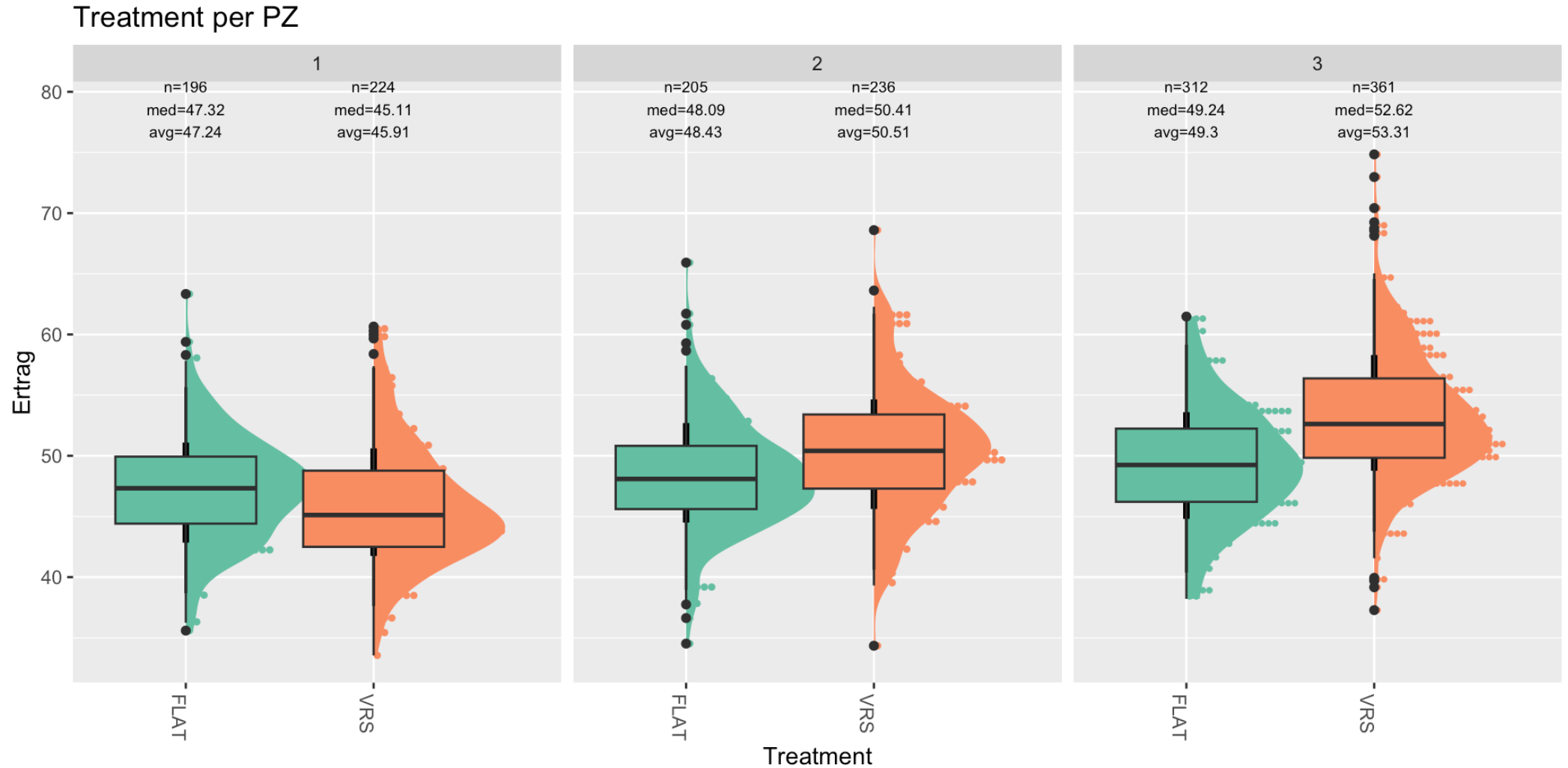


Boxplots - Haupteffekte III

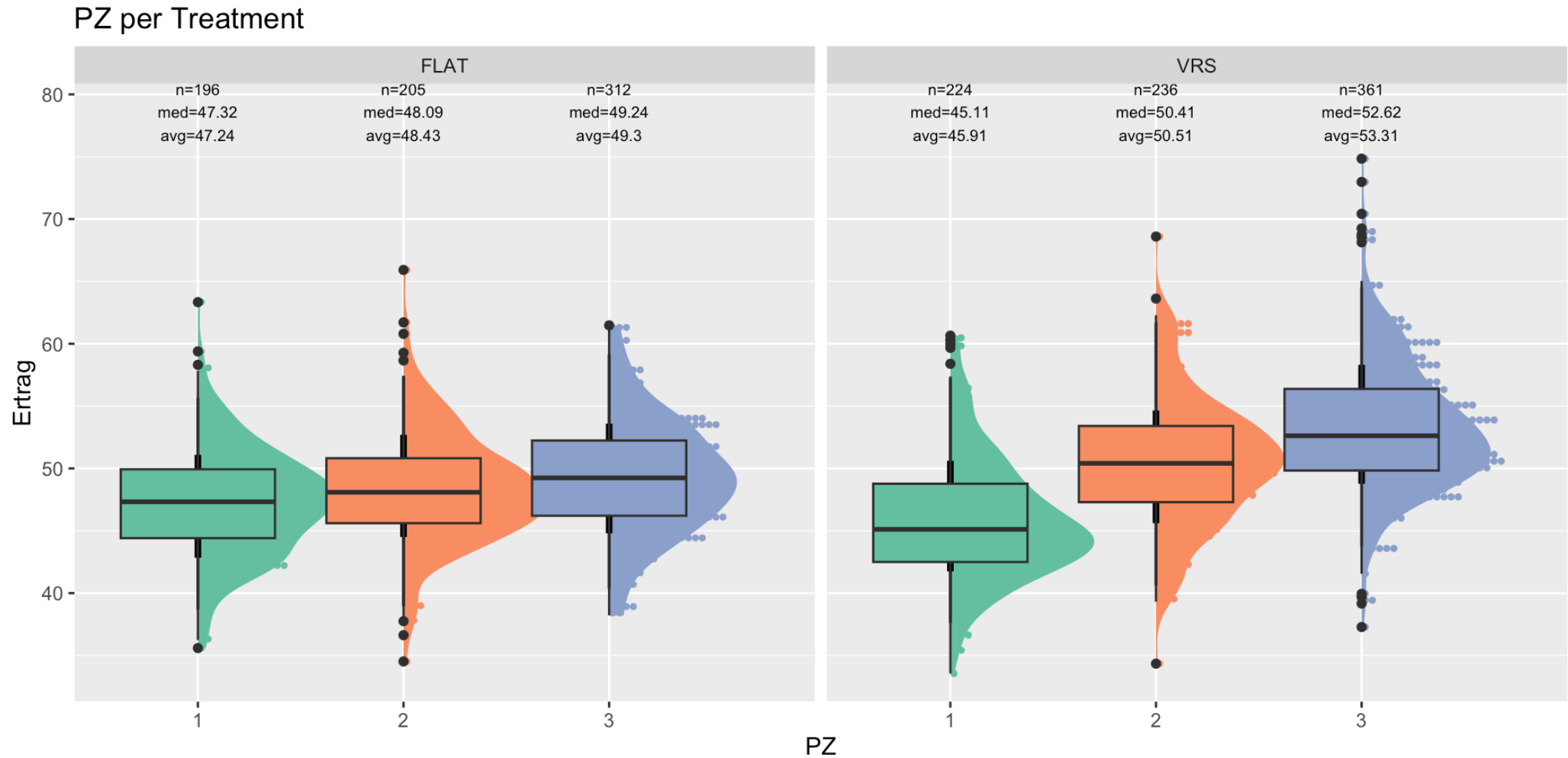
- VRS deskriptiv + 3.7 % bzw + 2.0 t/ha



Boxplots - Interaktionen I



Boxplots - Interaktionen II



2 Modellierung

Metriken

- **Least Significant Difference**
 - Trennschärfe für Gruppeneffekte in natürlicher Einheit
 - kleiner = besser
- **AIC**
 - Modellgüte vs Komplexität
 - kleiner = besser
- Beide Metriken verhalten sich nicht zwangsweise gleichförmig
- Aufnahme weiterer Parameter *kann* bessere Trennschärfe liefern, aber wird bei AIC penalisiert

Zusammenfassung

- Produktivität folgt Blockstruktur ($r = 0.73$)
- Starke Binnendifferenz in 1 und 4

Block trt	Mean	Median	SD	avg_PZ	avg_Elev	n
1 FLAT	50.05	49.74	3.96	4.79	286.68	148
1 VRS	52.77	52.13	4.03	3.59	284.78	321
2 FLAT	48.55	48.17	4.09	2.57	283.57	322
2 VRS	48.66	47.89	6.36	2.74	281.41	302
3 FLAT	49.94	49.37	5.04	4.21	273.46	111
3 VRS	53.89	53.33	5.94	4.05	271.83	108
4 FLAT	46.61	46.62	4.55	3.87	271.08	83
4 VRS	44.58	45.08	3.67	2.46	270.71	68
5 FLAT	43.22	42.80	4.22	1.98	269.61	49
5 VRS	43.97	44.17	4.82	1.73	268.41	22

Klassische Auswertung

- Linear Mixed Model

$$\bullet \text{ Yield} = \overbrace{\text{Treatment} + \text{PZ}}^{\text{Fixed Effects}} + \overbrace{(1 \mid \text{Block})}^{\text{Random Effect}} + \varepsilon$$

- Keine Nutzung der räumlichen Struktur
- Informationsverlust durch Aggregation?
- **Aber:** Unabhängigkeit der Beobachtungen

Modell	Estimate	SE	AIC	LSD
Treatment + Block	1.1	1.04	55.42	2.52
Treatment + PZ + Block	2.7	1.31	51.05	3.04

Maximum Likelihood estimates

Geostatistik - Überlegungen

- Modellierung nicht trivial - wähle:
 - Räumliches Modell (→ Variogramm)
 - Winkelabhängigkeit (→ Isotropie)
 - Blockstruktur
 - Kovariaten (→ PZ, Elevation, ...)

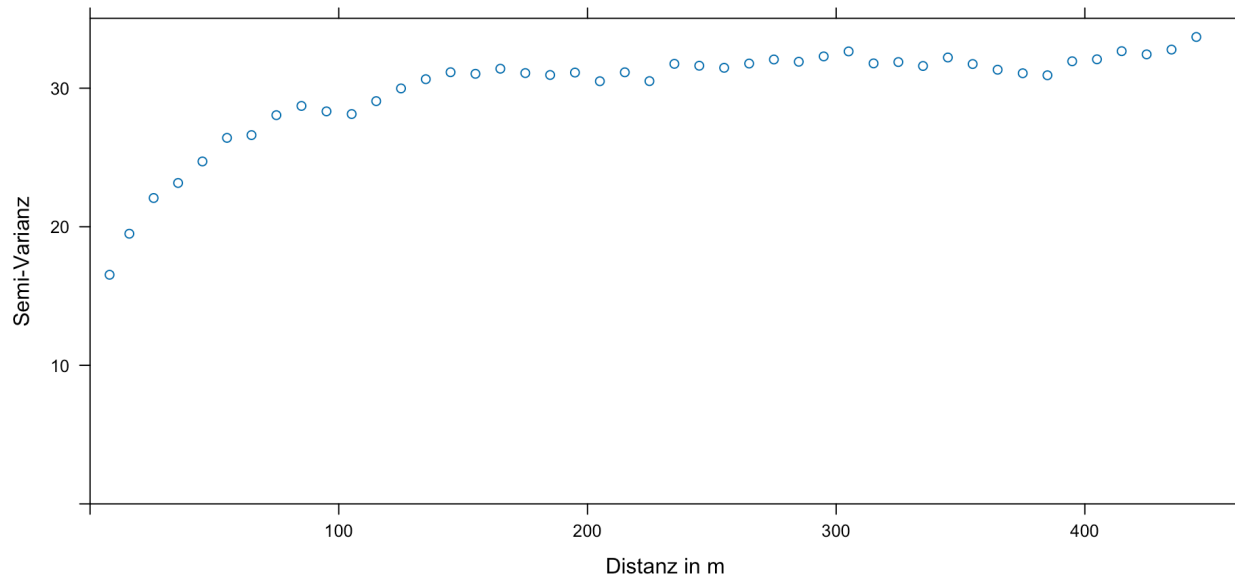
Geostatistik - Eingrenzung

- Matern-Variogramme flexibel
- Starke Richtungseffekte (durch Sensorik) → Anisotropie oft plausibel
- Blockstrukturen und Kovariaten können gleichwertige Informationen liefern

Variogramm

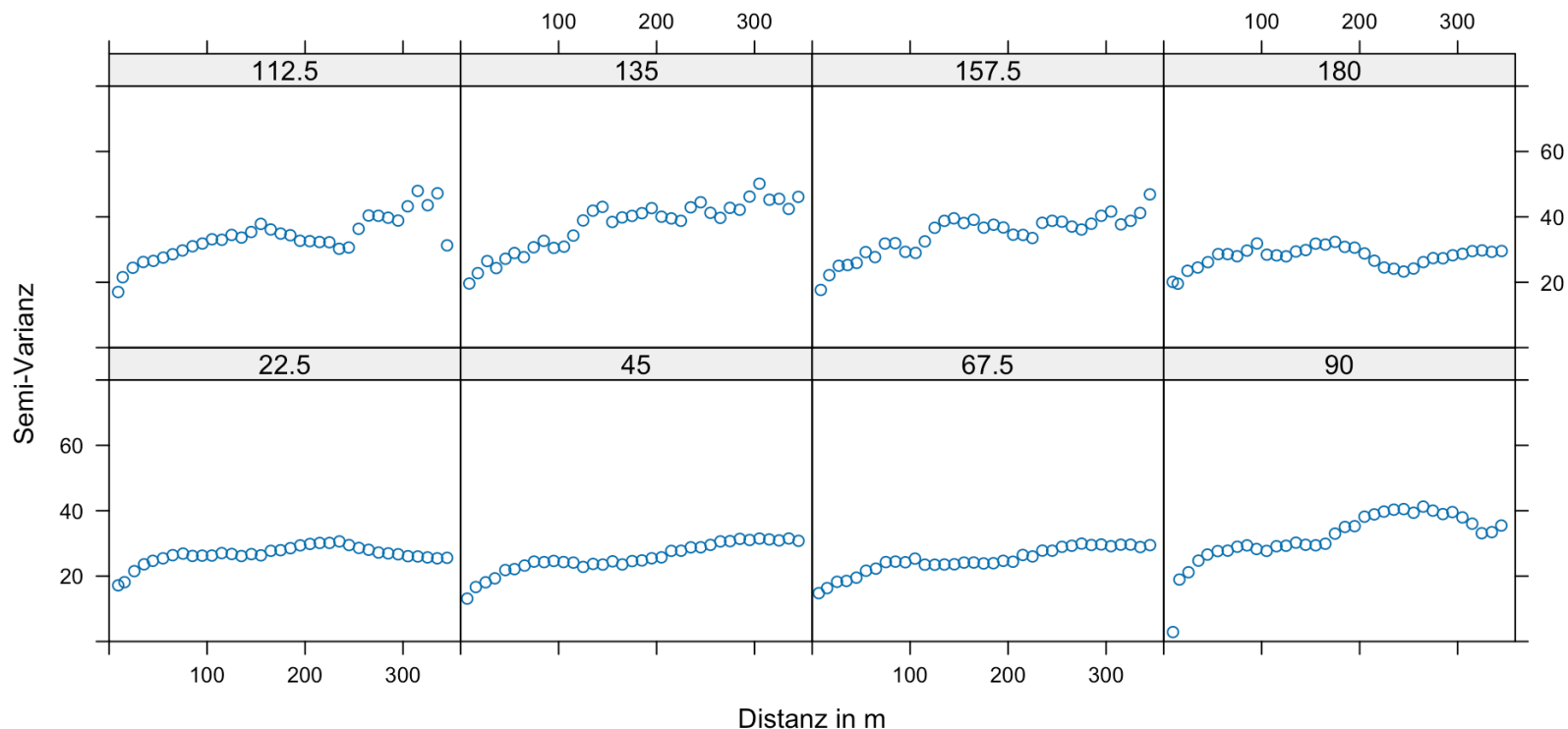
Intuition

- Ertragspunkte sind **nicht unabhängig**:
 - $n_{\text{effektiv}} \ll n$
- Punkte, die nah beieinander liegen, sind sich ähnlicher als weit entfernte Punkte.
- Modelliere die **Korrelation zwischen Punkten** als **Funktion der Distanz**



Anisotropie?

- Anisotropie beschreibt die Winkelabhängigkeit der Distanzen
 - Korrelation zwischen Punkten als **Funktion der Distanz *und* des Winkels**
- Im landwirt. Kontext diktieren z.B. Fahrspuren und Gelände diese Abhängigkeit



Übersicht der Modelle

No.	Name	Kovariaten	Fehler	Isotropie	Block
1	LMM	Treatment	iid $\varepsilon \sim \mathcal{N}(0, \sigma^2)$	–	Ja
2	LMM + PZ	Treatment + PZ	iid $\varepsilon \sim \mathcal{N}(0, \sigma^2)$	–	Ja
3	Geo	Treatment	Räumlich	Ja	Nein
4	Geo + Block	Treatment	Räumlich + RI	Ja	Ja
5	Geo + Interaktion	Treatment x PZ	Räumlich	Ja	Nein
6	Geo + Anisotropie	Treatment	Räumlich + Winkel	Nein	Nein
7	Geo + Anis. + Block	Treatment	Räumlich + Winkel + RI	Nein	Ja
8	Geo + Anis. + Inter.	Treatment x PZ	Räumlich + Winkel	Nein	Nein

Modellvergleich

model	block	anisotropy	covariate	estimate	SE	p	npar	AIC	LSD
Geo	no	no	no	-3.36	0.49	2	4	8922.37	0.95
Geo + Block	yes	no	no	-3.14	0.53	2	5	8921.43	1.05
Geo + Interaction	no	no	yes	-3.11	0.49	6	4	8919.75	0.97
Geo + Anisotropy	no	yes	no	-2.35	0.94	2	6	8882.90	1.85
Geo + Aniso + Block	yes	yes	no	-2.47	0.90	2	7	8858.48	1.76
Geo + Aniso + Inter	no	yes	yes	-3.06	0.94	6	6	8855.89	1.84
LMM	yes	no	no	-1.10	1.04	2	4	55.42	2.52
LMM + Covariate	yes	no	yes	-2.70	1.31	3	5	51.05	3.04

Maximum Likelihood estimates, AIC not comparable between model classes

Beobachtungen I

- Anisotropie
 - verbessert Model-Fit deutlich (ΔAIC zw. 40 und 60)
 - verschlechtert LSD deutlich
- Interaktions-Term
 - verbessert Model-Fit
 - LSD konstant
- Blockeffekt in ähnlicher Größenordnung
- Designeffekte implizit “gleichwertig” (Block $\approx$ PZ)
- Block und Kovariate verbessern Modell-Fit vor allem bei Anisotropie

Beobachtungen II

- Nur LMM-Schätzungen insignifikant
- Standardfehler der Differenz liegt je nach Modellspezifikation zwischen
 - 0.95 und 1.84 t/ha in der Geostatistik
 - 2.52 und 3.02 t/ha “klassisch”
- Statistische Effizienz für geostat. Modell mit Anisotropie und Block gegenüber LMM:
 - $(2.52/1.76)^2 = 2.05 = 105\%$
- Um selbe SE in LMM zu erreichen, bräuchte man
 - $5 * 2.05 = 10$ Wiederholungen (statt 5)

3 Fazit

Fazit - allgemein

- Geostatistische Modelle haben in allen Spezifikationen deutlich kleinere Grenzdifferenzen
- Hierfür mussten allerdings
 - Daten bereinigt & kalibriert
 - komplexe & zeitintensive Modelle berechnet werden
- Wahl des finalen Modells nicht trivial
- Blockstruktur “kostet” wenig und als Backup-Modell stets sinnvoll
 - Nur valide unter Randomisation
- Finale Schätzung sollte mit REML statt ML erfolgen (Bias)

Fazit - konkret

- Finales Modell mit Anisotropie und Interaktion
- Relative Grenzdifferenz **3.5 % statt 5.1 %**
- “Einsparung” von 5 Wiederholungen
- (An-)Isotropie verbessert Modell-Güte deutlich und weist damit auf starke Richtungseffekte hin
 - Druschrichtung als fixer Effekt?
- Modellierung mit Kovariaten hilft in jedem Setup
- Matern-Modelle rechnen am längsten, sind aber am flexibelsten und zeitsparender als ein Vergleich zwischen 5 anderen Modellen

Backup

Modellvergleich - “Klassisch”

- Linear Mixed Model

$$\text{Yield} = \text{Treatment} + (1|\text{Block}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

- Linear Mixed Model + Kovariate

$$\text{Yield} = \text{Treatment} + \text{Productivity} + (1|\text{Block}) + \varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \sigma^2)$$

Modellvergleich - Geostatistisch

- Geostatistisches Modell (Matern, isotropisch)

$$\text{Yield} = \text{Treatment} + \varepsilon(s), \quad \varepsilon(s) \sim \mathcal{GP}(0, C(h)), \quad C(h) = \text{Matern}(h; \nu, \rho)$$

- Geostatistisches Modell (Matern, isotropisch) + Block-Struktur

$$\text{Yield} = \text{Treatment} + (1|\text{Block}) + \varepsilon(s), \quad \varepsilon(s) \sim \mathcal{GP}(0, C(h)), \quad C(h) = \text{Matern}(h; \nu, \rho)$$

- Geostatistisches Modell (Matern, anisotropisch)

$$\text{Yield} = \text{Treatment} + \varepsilon(s), \quad \varepsilon(s) \sim \mathcal{GP}(0, C(h, \theta)), \quad C(h, \theta) = \text{Matern}(h, \theta; \nu, \rho)$$

- Geostatistisches Modell + Interaktion mit Kovariate (Matern, isotropisch)

$$\text{Yield} = \text{Treatment} \times \text{Productivity} + \varepsilon(s), \quad \varepsilon(s) \sim \mathcal{GP}(0, C(h)), \quad C(h) = \text{Matern}(h; \nu, \rho)$$

Terrain

- **Aspect** is the compass direction that a slope faces (also known as *exposure*). Given in degrees.
- **Slope** difference to surrounding cells. Given in degrees.
- **Elevation** height above NN. Given in meters.

