



**Deutsche Region der Internationalen
Biometrischen Gesellschaft (IBS-DR)**

Arbeitsgruppe Landwirtschaftliches Versuchswesen

On-Farm-Experimente

Leitfaden, Thesen und Vorträge

Workshop 23. – 24. November 2011, Kassel





**Deutsche Region der Internationalen
Biometrischen Gesellschaft (IBS-DR)**

Arbeitsgruppe

Landwirtschaftliches Versuchswesen

On-Farm-Experimente

Leitfaden, Thesen und Vorträge

Workshop vom 23. bis 24. November 2011 in Kassel

Impressum

Herausgeber: Deutsche Region der Internationalen Biometrischen Gesellschaft,
Arbeitsgruppe Landwirtschaftliches Versuchswesen

Ort: Halle (Saale)

Jahr: 2012

Druck: Druck-Zuck GmbH, Seebener Straße 4, 06114 Halle (Saale), Deutschland

ISBN: 978-3-86829-494-1

© Copyright Deutsche Region der Internationalen Biometrischen Gesellschaft,
Arbeitsgruppe Landwirtschaftliches Versuchswesen
Alle Rechte vorbehalten.

Inhaltsverzeichnis

Einführung	5
Leitfaden zur Einordnung, Planung, Durchführung und Auswertung von Versuchen unter Produktionsbedingungen (On-Farm-Experimente)	7
Thesen für Landwirte	26
Thesen für Versuchsdurchführer	27
Thesen für Entscheider	28
<i>Hans-Georg Schön</i> Fehlerquellen bei der Erfassung von teilflächenspezifischen Ertragsdaten und Korrekturmöglichkeiten	29
<i>Hans-Peter Piepho, Jens Möhring und Emlyn R. Williams</i> Zu den Grundsätzen der Versuchsplanung (insbesondere Randomisation)	36
<i>Christel Richter</i> Anlagemethoden für On-Farm-Experimente	52
<i>Joachim Spilke</i> Entwicklung des Auswertungsmodells	69
<i>Christel Richter und Heinrich Thöle</i> On-Farm-Experiment mit georeferenzierten Daten und teilflächenspezifischer Applikation	84
<i>Dominik Dicke und Andreas Büchse</i> Analyse von On-Farm-Experimenten am Beispiel der Wirkung von Fungizidbehandlungen im Stadium der Rapsblüte (BBCH 65)	103
<i>Ulfried Obenauf, Imke Borchardt, Christoph Lubkowitz und Carsten Kock</i> On Farm Research – ein Gemeinschaftsprojekt der Landwirtschaftskammer Schleswig-Holstein und der Gutsverwaltung Helmstorf	114
Arbeitsergebnisse des gemeinsamen Workshops "On-Farm-Experimente"	135
Verzeichnis der Autoren	138

Einführung

Versuche spielen als Erkenntnismittel in der Agrarforschung und der Umsetzung von Forschungsergebnissen in die landwirtschaftliche Praxis eine zentrale Rolle. Dabei verfügt die Agrarforschung über mehrere Versuchstypen, von denen jeder für die einzelnen Stufen des Erkenntnisprozesses eine spezifische Bedeutung und Rechtfertigung hat. Bei der Auswahl des jeweils geeigneten Versuchstyps geht es stets darum, die Einhaltung der biometrischen Prinzipien zur Durchführung von Versuchen mit einer größtmöglichen Nähe zur landwirtschaftlichen Produktion zu verbinden. Gerade unter dem Gesichtspunkt einer kontrollierten Präzision und Konstanzhaltung von Störgrößen haben Parzellenversuche eine herausragende Bedeutung. Das wird auch in Zukunft so bleiben. Daneben rücken aber Untersuchungen unter Produktionsbedingungen, sog. „On-Farm-Experimente“ (OFE), zunehmend in das Blickfeld des Interesses. Ausgelöst vor allem durch neue technische Entwicklungen bietet sich damit die Möglichkeit, eine abschließende Bewertung von Forschungsergebnissen direkt im Produktionsfeld vorzunehmen. Das ist ein großer Vorteil von OFE. Die damit verbundenen Erwartungen werden sich aber nur erfüllen, wenn man OFE als einen Versuchstyp begreift, für den die biometrischen Prinzipien der Planung und Auswertung vom Grundsatz ebenso gelten wie für jeden anderen Versuchstyp auch. Dieser Sachverhalt wird oft nicht ausreichend beachtet.

Davon abgeleitet hat die Arbeitsgruppe Landwirtschaftliches Versuchswesen der Deutschen Region der Biometrischen Gesellschaft federführend in Zusammenarbeit mit der Arbeitsgruppe Versuchswesen der Gesellschaft für Pflanzenbauwissenschaften, dem Arbeitskreis Biometrie und Versuchsmethodik der Deutschen Phytomedizinischen Gesellschaft, der Arbeitsgruppe Biometrie und Bioinformatik der Gesellschaft für Pflanzenzüchtung einen **Leitfaden** mit den biometrischen Grundsätzen für die Planung und Auswertung von OFE erarbeitet. Die Erarbeitung erfolgte in enger Abstimmung mit dem Fachausschuss "Versuchswesen in der Pflanzenproduktion" der Deutschen Landwirtschaftsgesellschaft.

Der Leitfaden richtet sich vorrangig an den Personenkreis, der diese Versuche plant, durchführt und auswertet. Die wichtigsten Aussagen wurden in 10 Thesen (**Thesen zum Leitfaden**) zusammengefasst. Entsprechend dieses Adressatenkreises wurde der Schwerpunkt bewusst auf die biometrische Methodik gelegt. Damit lässt sich eine Spezifik und Darstellung von Details nicht vermeiden, um mit dem Leitfaden für diesen Nutzerkreis tatsächlich eine wirksame Unterstützung zu liefern.

Natürlich sind die allgemeinen Grundsätze auch für andere Zielgruppen von Interesse. Hier sind zunächst die Führungskräfte in Bundes- und Länderbehörden, Landwirtschaftskammern und Landesanstalten zu nennen, die selbst keine Versuche durchführen, deren Durchführung aber anweisen und kontrollieren. Für diese Zielgruppe wurde ebenfalls ein Thesenpapier erarbeitet (**Thesen für Entscheider**). Als weitere sehr bedeutsame Zielgruppe ergeben sich die Landwirte, die an der Durchführung von OFE in ihrem Betrieb oder an der Nutzung von OFE-Ergebnissen für die eigene Entscheidungsunterstützung interessiert sind. Auch für diese Zielgruppe sind die Aussagen in Thesen (**Thesen für Landwirte**) zusammengefasst.

Leitfaden und Thesen bilden den ersten Schwerpunkt dieser Publikation zu On-Farm-Experimenten. Die **Vorträge** und zusammengefassten **Arbeitsergebnisse** eines gemeinsamen Workshops des Arbeitskreises „Koordinierung im Versuchswesen“ des Verbandes der Landwirtschaftskammern und der Arbeitsgruppe "Landwirtschaftliches Versuchswesen" am 23. und 24. November 2011 in Kassel bilden den zweiten Schwerpunkt. Der Workshop hatte das Ziel, zu besonders interessierenden Themen eine detaillierte Diskussion zu ermöglichen, tiefgründiger als das im Leitfaden möglich ist. Den Abschluss bilden die zusammengefassten Arbeitsergebnisse des Workshops. Damit sollen gleichzeitig die einzuhaltenden methodischen Mindestforderungen bei der Planung, Durchführung und Auswertung von On-Farm-Experimenten dokumentiert werden. Mit dieser Publikation sollen neben den methodischen Grundlagen auch praktische Erfahrungen zusammengetragen werden, um damit eine solide Basis für eine fruchtbare weiterführende Diskussion zu schaffen.

Die Autoren danken der Deutschen Region der Internationalen Biometrischen Gesellschaft und der Gesellschaft zur Förderung der Agrar- und Ernährungswissenschaften an der Martin-Luther-Universität Halle-Wittenberg für die finanzielle Unterstützung des Drucks. Weiterhin bedanken wir uns bei Frau Diplom-Agraringenieur Heike Braunsdorff für die umfangreiche technische Unterstützung bei der Drucklegung.

Leitfaden zur Einordnung, Planung, Durchführung und Auswertung von Versuchen unter Produktionsbedingungen (On-Farm-Experimente)

1 Einleitung und Problemstellung

Versuche unter Produktionsbedingungen stellen seit jeher eine bedeutsame Komponente des landwirtschaftlichen Versuchswesens in der Pflanzenproduktionsforschung dar (SVAB, 1957; SPECHT und MÜLLER, 1960; SPECHT und SCHULZE, 1963; BÄTZ et al., 1982; THOMAS, 2006). Im Folgenden sollen solche Versuche unter Produktionsbedingungen, die mit einer wissenschaftlichen Zielstellung verbunden sind und damit die Einhaltung von Grundsätzen einer wissenschaftlichen Arbeitsweise erfordern, als „On-Farm-Experimente“ (OFE) bezeichnet werden. Die Erprobung eines Verfahrens/einer Variante, die ein einzelner Landwirt unter seinen konkreten Bedingungen durchführt, soll zur Abgrenzung dazu als „Demonstrationsversuch“ bezeichnet werden. Bei vorhandener versuchsmethodischer und statistischer Expertise kann natürlich auch für einen einzelnen Betrieb ein OFE konzipiert werden, das dann aber einen entsprechend eingeschränkten Aussagebereich hat.

Dabei sollen „Produktionsbedingungen“ die Situation der Bewirtschaftung von Ackerflächen mit einer praxisüblichen Größe und Maschinenausstattung beschreiben. Bei Beachtung dieser Bedingungen können OFE auch auf Versuchsstationen durchgeführt werden. Andere in der Literatur übliche Begriffe für Versuche unter Produktionsbedingungen sind: Praxisversuch, Produktionsexperiment, Großversuch oder real world experiment.

Mit OFE zu beantwortende Versuchsfragen beziehen sich häufig auf die Bewertung der Anbaueignung von Sorten, Prüfung von Produktionsverfahren bei Aussaat, Düngung, Ausbringung von Pflanzenschutzmitteln und Ernte oder auf Probleme des Maschineneinsatzes.

Damit wird die wissenschaftliche Zielstellung verfolgt, Forschungs- und Entwicklungsergebnisse unter Praxisbedingungen zu bewerten, die sich aus zeitlich vorgelagerten Versuchen im Labor, Gewächshaus, auf Versuchsstationen oder an Maschinen-Modellen (Prototypen) ergeben haben, und damit zusätzliche Fragestellungen zu prüfen, die nur unter Produktionsbedingungen untersucht werden können. Dazu gehören Fragen zur Einordnung und Realisierbarkeit der zu prüfenden Varianten hinsichtlich

- der in der Praxis üblichen Technik und der Einordnung in technologische Abläufe (agrotechnisch-technologische Repräsentativität),
- sozio-ökonomischer Aspekte (Prüfung der Akzeptanz bestimmter Verfahren, deren Wirtschaftlichkeit und ihre Beziehung zum gesamten sozio-ökonomischen Umfeld) und/oder
- ökosystemarer Aspekte (Prüfung von ökologischen Effekten in gegenüber den vorgelagerten Versuchstypen offeneren Systemen).

Hier ordnet sich z. B. auch das fallspezifische, hypothesengeleitete Monitoring zur Analyse und Bewertung von Effekten von Genetisch Modifizierten Organismen (GMO) als Teil des anbaubegleitenden Monitorings ein (TRAXLER et al., 2000).

Die Beantwortung dieser zusätzlichen Fragestellungen erlaubt einen erweiterten Aussagebereich und damit einen potentiellen Vorteil gegenüber den vorgelagerten Versuchstypen.

Neben einer wissenschaftlichen Zielstellung haben OFE auch eine Demonstrations- und Überleitungsfunktion und werden zunehmend zur Generierung acker- und pflanzenbaulicher Beratungsinhalte eingesetzt. Deshalb sollten die für die Durchführung gewonnenen Betriebe auch für ihre konkreten Bedingungen letztlich eine Bewertung der geprüften Varianten erhalten bzw. vornehmen können. Daraus ergibt sich, dass meist eine der geprüften Varianten die betriebsübliche Variante ist.

Eine Prüfung unter Produktionsbedingungen erfordert in der Regel die Durchführung eines OFE unter verschiedenen Standort- und Witterungsbedingungen. Hinsichtlich der dadurch bestimmten ökologischen Repräsentativität gelten im Wesentlichen die gleichen Aussagen wie im Parzellenfeldversuch (siehe Abschnitt 2.1).

Gegenüber Versuchen unter reproduzierbaren, nahezu reproduzierbaren oder praxisnahen Versuchsbedingungen ist es bei der Planung und Durchführung von OFE schwieriger, die Konstantfaktoren zu definieren, während der Versuchsdurchführung einzuhalten und zu kontrollieren und damit dem *ceteris-paribus*-Prinzip zu entsprechen. Dieses hat zur Folge, dass im Vergleich zu den anderen Versuchstypen bei gleicher Wiederholungsanzahl die Präzision eines Prüfgliedmittelwertes oder einer Mittelwertdifferenz geringer ausfallen wird. Das heißt, die Schätzungen sind oft mit einem vergleichsweise höheren Zufallsfehler behaftet und in diesem Sinne weniger präzise, als das bei Parzellenversuchen der Fall ist. Oft noch schwerwiegender ist die mögliche Einschränkung der Treffgenauigkeit, beispielsweise durch fehlerhaft arbeitende Sensoren für die Merkmalerfassung oder systematische und nicht erkannte Bodenunterschiede auf der Versuchsfläche. Dann repräsentieren die gemessenen Merkmalswerte nicht die eigentlich interessierenden Prüfglieder, sondern beinhalten auch die Effekte der verzerrten Merkmalerfassung oder der Bodenunterschiede. Ein OFE muss gerade deshalb so geplant, durchgeführt und ausgewertet werden, dass beiden Komponenten der biometrischen Genauigkeit – Präzision und Treffgenauigkeit – Rechnung getragen wird. Die Unverzichtbarkeit der Einhaltung einer definierten Präzision (d. h., der Zufallsfehler einer Schätzung ist kleiner als eine vorgegebene Grenze) und Treffgenauigkeit (d. h., das Mittel der Schätzwerte eines Parameters trifft den Parameter selbst) bedingt die Anerkennung der vollen Gültigkeit der Grundprinzipien der Versuchsplanung (siehe Abschnitt 2.1). Aus der angeführten Darstellung folgt auch, dass OFE keinesfalls Parzellenversuche in Versuchsstationen ersetzen, sondern nur ein Versuchstyp neben anderen in einem abgestimmten Prüfsystem sein können.

Unter dem Schlagwort „Precision Farming“ hat sich insbesondere im letzten Jahrzehnt ein neuer Arbeitskomplex herausgebildet, der für die pflanzliche Erzeugung neue Möglichkeiten einer präziseren Wirtschaftsweise bietet. Damit sind auch für landwirtschaftliche Versuche unter Produktionsbedingungen neue Möglichkeiten verbunden. Ausgelöst wurde das

- durch technische Entwicklungen mit der Möglichkeit einer präzisen Positionsbestimmung von mobilen Arbeitsmaschinen durch das Global Positioning System (GPS) und
- die Bereitstellung von Prozesstechnik mit der Möglichkeit einer positionsabhängigen Ausbringung von Saatgut, Dünge- und Pflanzenschutzmitteln, aber auch einer georeferenzierten Ertragsfassung und weiteren Entwicklungen im Hinblick auf Inhaltsstoffe, beispielsweise Protein.

Abbildung 1 zeigt eine mögliche Klassifikation von OFE, die nach Aspekten vorgenommen wurde, die für die Planung und/oder Auswertung der Versuche relevant sind.

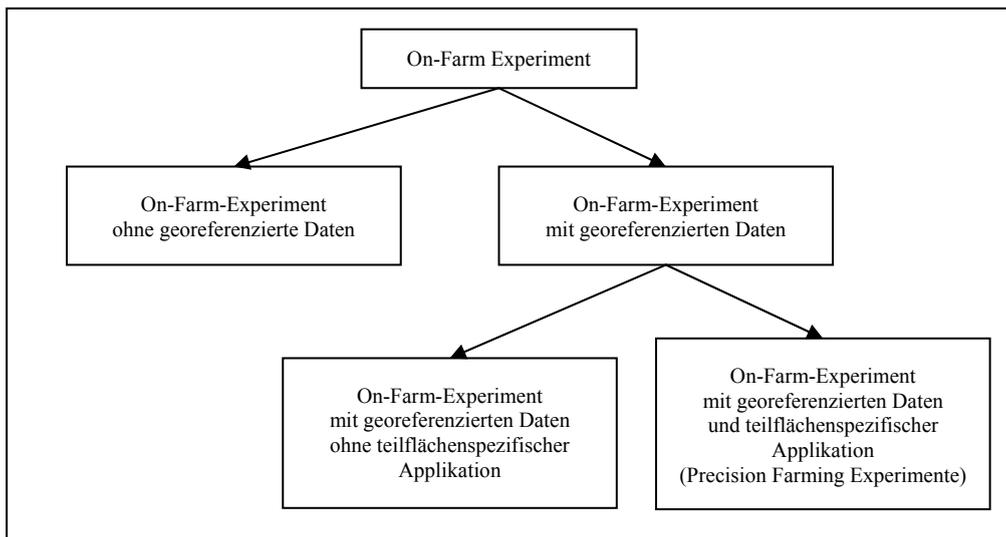


Abbildung 1: Übersicht zur Begriffsbildung und Klassifikation von On-Farm-Experimenten

Insbesondere im Rahmen der Precision Farming-Forschung haben sich mit den erweiterten technischen Möglichkeiten die durch OFE zu beantwortenden Fragestellungen verändert. Generell liefern OFE mit georeferenzierten Daten die Möglichkeit einer detaillierteren Analyse als solche ohne diese Möglichkeit. Ob diese Georeferenzierung einen zusätzlichen Wert für die Beantwortung der Versuchsfrage besitzt, hängt zum einen von der Versuchsfrage selbst ab, aber auch von den natürlichen Gegebenheiten auf dem Feld. Ziel sollte es sein, diese Daten im Zuge der Auswertung zu nutzen und gegebenenfalls in die Interpretation einfließen zu lassen. Aus Sicht der Auswertung kann diesen veränderten Bedingungen aufgrund der methodischen Entwicklungen der räumlichen Statistik und deren Umsetzung in Software Rechnung getragen werden. Unabhängig davon sind jedoch auch bei der Anlage derartiger Versuche die Prinzipien der Versuchsplanung uneingeschränkt gültig und für die Gewinnung aussagefähiger Ergebnisse anzuwenden. Bei der Auswertung stellen Fragen der Modellwahl sogar erhöhte Anforderungen (siehe Abschnitt 2.4). Neben den eigentlich interessierenden Untersuchungsfaktoren sind die relevanten Störgrößen zu erkennen und ein Modell für die adäquate, meist räumlich bedingte Kovarianzstruktur zu entwickeln. Das zwingt zur Anwendung von Methoden der Modellselektion, um zu einer begründeten Wahl des mathematisch-statistischen Auswertungsmodells zu kommen.

Abgeleitet von der kurz skizzierten Situation ist es Gegenstand dieses Leitfadens, die Schrittfolge bei der Anlage und Auswertung von OFE aus biostatistischer Sicht darzustellen. Das erscheint vor allem deshalb dringend, weil die gewachsenen technischen Möglichkeiten oft zu einer unkritischen Anwendung von OFE führen. Dabei wird leicht vergessen, dass es sich hierbei nur um einen möglichen Versuchstyp im Rahmen des landwirtschaftlichen Versuchswesens handelt. Parzellenversuche in landwirtschaftlichen Versuchsstationen können dadurch nicht ersetzt werden.

Weiterhin geht es darum zu verdeutlichen, dass bei OFE dieselben biometrischen Grundsätze anzuwenden sind wie bei anderen Versuchstypen auch. Eine umfassende Darstellung der biometrischen Gesichtspunkte von OFE und des aktuellen Sachstandes ist auch bei PIEPHO et. al (2011) zu finden.

2 Allgemeine Grundsätze für OFE

2.1 Ökologischer Aussagebereich von OFE

Die Wahl von Standort und Flächen bedingt ebenso wie bei Parzellenversuchen in Versuchstationen den ökologischen Aussagebereich. Daraus folgt, dass eine hohe Aussagekraft von OFE nur bei einer überbetrieblichen Planung der Standorte und Jahre zu erreichen ist. Wird beispielsweise ein Sortenvergleich einer bestimmten Fruchtart nur auf einem Schlag in einem Jahr durchgeführt, so beziehen sich die Ergebnisse genau auf diesen Schlag im jeweiligen Untersuchungsjahr. Der Wert eines solchen Versuches ist gering. Das folgt schon allein daraus, dass auf diesem Schlag diese Fruchtart i. A. in den darauf folgenden Jahren nicht angebaut wird. Ein solcher Einzelversuch hat keine verallgemeinerungsfähige Aussagekraft. Daraus folgt konsequenterweise, dass nur eine geplante Serie und entsprechende Auswertung bedeutsam ist. Der Zusammenstellung in Tabelle 1 (in Anlehnung an THOMAS, 2006) ist zu entnehmen, wie die ökologische Repräsentativität eines OFE hinsichtlich Standort- und klimatischer Repräsentanz zu bewerten ist. Diese unterscheidet sich nicht von Parzellenfeldversuchen. Die über die ökologische Repräsentativität hinausgehenden Aspekte der Repräsentativität der OFE wurden bereits in Abschnitt 1 erläutert.

Tabelle 1: Ökologische Repräsentativität eines OFE (in Anlehnung an THOMAS, 2006)

	Standortwahl	Laufzeit des Versuchs	ökologischer Aussagebereich
Einzelversuch	ein Ort ¹⁾	ein Jahr	Ackerfläche(n) des Schlages (der Schläge) und die Witterung des Versuchsjahres an dem Ort
	ein Ort ^{1) 2)}	mehrere Jahre ³⁾	Ackerflächen des Versuchsortes (Schlages) ²⁾ und seine klimatischen Bedingungen ³⁾
Versuchsserie	mehrere Orte	fest ausgewählt	Ackerflächen der Versuchsorte (Schläge) und die Witterung des Versuchsjahres an diesen Orten
		zufällig ausgewählt	Anbauggebiet und Witterung des Versuchsjahres
		fest ausgewählt	Ackerflächen der Versuchsorte und ihr Klima ³⁾
		zufällig ausgewählt	Anbauggebiet und seine klimatischen Bedingungen ³⁾

¹⁾ An dem einen Ort wird der Versuch auf einem oder mehreren Schlägen mit gleichen Standortbedingungen durchgeführt.

²⁾ Auf Grund einer Fruchtfolge wird es sich bei einer Untersuchung über mehrere Jahre i. A. um mehrere Schläge eines Ortes handeln.

³⁾ Obwohl bei mehreren Jahren die Witterungsverläufe als zufällig betrachtet werden können, muss bei der Auswertung und Interpretation beachtet werden, ob sie tatsächlich als repräsentative Stichprobe aus den jeweiligen klimatischen Verhältnissen betrachtet werden können. Ist dieses nicht der Fall, so kann nur eine Bewertung für die Witterungsbedingungen der Versuchsjahre erfolgen.

Neben den in Tabelle 1 aufgeführten Fällen sei auf einige spezielle Situationen hingewiesen. Wird der Versuch an einem Ort in einem Jahr auf einem oder mehreren Schlägen mit gleichen Standortbedingungen angelegt (die verschiedenen Schläge stellen Wiederholungen dar – siehe Abschnitt 2.2), so handelt es sich, wie oben angegeben, um einen Einzelversuch. Repräsentieren die verschiedenen Schläge des Ortes jedoch unterschiedliche Standortbedingungen, so kann der Versuch als Versuchsserie betrachtet werden. Dabei können die nach Standortbedingungen klassifizierten Schläge in der Auswertung formal wie verschiedene Orte behandelt werden. Der Unterschied zu einer „echten“ mehrortigen Serie besteht hier in den einheitlicheren Konstantfaktoren (z. B. Witterung, Betriebsmanagement). Für diese beiden Fälle ist die Bezeichnung *Betriebsgroßversuch* gebräuchlich. Mehrortige Serien werden häufig als *Großflächenstreuversuch* bezeichnet (THOMAS, 2006). Abhängig von der Auswahl der Orte (aus einem oder mehreren Anbaugebieten) und/oder der Schläge innerhalb der Orte (gleiche oder ungleiche Standortbedingungen) entstehen u. U. hierarchische Strukturen, die bei der Auswertung und Interpretation zu beachten sind.

Die Begriffe Betriebsgroßversuch und Großflächenstreuversuch sollten nur noch zur Charakterisierung der Ein- oder Mehrortigkeit des Versuchs benutzt werden. Die in der Literatur damit verknüpften Beschreibungen der Versuchsanlage sind teilweise überholt und/oder widersprüchlich.

Die Erhaltung der Möglichkeit einer Serienauswertung stellt hohe Anforderungen an die Definition der Prüfglieder und an die Disziplin bei deren Einhaltung. Gerade der letztgenannte Aspekt zwingt zu der Forderung, wonach eine Durchführung von OFE ohne externe verfahrensmethodische Unterstützung im Allgemeinen nicht möglich ist.

2.2 Prinzipien der Versuchsplanung und ihre Umsetzung

Die aus Sicht der Anlage des Versuches wesentlichen Unterschiede von OFE gegenüber Parzellenversuchen auf Versuchsstationen sind:

- (a) die oft erheblich größere Fläche der Parzellen und
- (b) die Tatsache, dass pro Parzelle meist mehrere, oft sehr viele Einzelmessungen vorliegen, die u. U. georeferenziert sind und somit eine geostatistische Auswertung ermöglichen.

Ein einzelner Wert pro Parzelle liegt beispielsweise vor, wenn die Gesamtparzelle oder eine Mährescherbreite zur Ertragsermittlung herangezogen wird. Erfolgt die Ertragsfeststellung mittels entsprechender Sensoren am Mährescher, so können sehr viele Einzelwerte vorliegen, die bei einer Kopplung mit GPS georeferenziert sind. Auch für einige Merkmale der Produktqualität existieren geeignete Sensoren. Steht diese Technik nicht zur Verfügung, so werden Merkmale häufig von Hand auf mehreren Ausschnitten aus der Parzelle ermittelt. Auf spezielle Verhältnisse bei Versuchen mit georeferenzierten Daten wird in Abschnitt 3 eingegangen.

Die Ergebnisse von OFE dienen der Entscheidungsunterstützung bei der Wahl zwischen mehreren Varianten/Verfahren (Beispiele siehe Abschnitt 1). Oft sind diese alternativen Varianten mit erheblich unterschiedlichen festen und variablen Kosten verbunden. Das zwingt zur Forderung nach Ergebnissen mit hoher und kontrollierter Präzision und Treffgenauigkeit, um zu einer begründeten Varianten-/Verfahrenswahl zu kommen. OFE weisen daher bei der Kontrolle der statistischen Genauigkeit keinen grundsätzlichen Unterschied zu Parzellenversuchen in Versuchsstationen auf. Zur Sicherung der Genauigkeitsforderungen sind daher die Grundprinzipien der Versuchsplanung

- Randomisation,
 - Wiederholung und
 - Blockbildung
- grundsätzlich einzuhalten.

Randomisation bedeutet die zufällige Zuordnung der Wiederholungen der Prüfglieder zu den Randomisationseinheiten. Damit wird das Ziel verfolgt, Zufallseinflüsse von Umweltwirkungen sowie von Nachbarschaftswirkungen der Prüfglieder auszugleichen. Die Vermeidung einer systematischen Anordnung der Prüfglieder erlaubt, eventuell existierende und im Zuge der Versuchsplanung nicht erkannte Störgrößen im Mittel aller möglichen Randomisationen in ihrer Wirkung zu egalisieren. Wie bei Kleinparzellenversuchen ist bei OFE die Parzelle die wesentliche Randomisationseinheit. Falls zur Beantwortung der Versuchsfrage ein Parzellenmittelwert hinreichend und gerechtfertigt ist (siehe Abschnitt 3.2 „gleichmäßige Heterogenität“), ermöglicht die randomisierte Zuordnung von Behandlungen zu Parzellen immer eine klassische varianzanalytische Auswertung auf der Basis von Parzellenmittelwerten entsprechend der Versuchsanlage, d. h., zum Beispiel entsprechend einer randomisierten Blockanlage, falls eine solche Anlage gewählt wurde (vgl. Absatz Blockbildung).

Wiederholung bedeutet die mehrfache Anlage desselben Prüfgliedes mit dem Ziel der Schätzung der zufallsbedingten Varianz. Um diesem Ziel entsprechen zu können, müssen die Wiederholungen zeitgleich unter den gleichen Boden- und Witterungsbedingungen angelegt sein. Die Anzahl der Wiederholungen eines Prüfgliedes wird zunächst durch die entsprechende Anzahl der Parzellen bestimmt. Auf Grund der Randomisation der Prüfglieder auf Parzellen können die Effekte der Parzellen als unabhängig betrachtet werden. Die innerhalb einer Parzelle vorliegenden Einzelmessungen sind jedoch wiederholte Messungen (repeated measures) und stellen somit keine „echten“ Wiederholungen dar. Im Allgemeinen sind diese Beobachtungen voneinander abhängig. Wird bei der Auswertung ein Modell für die Einzelmessungen zu Grunde gelegt, so muss diese Abhängigkeit bei der Auswertung berücksichtigt werden (siehe Abschnitt 3). Diese spezielle Kovarianzstruktur – Unabhängigkeit der Parzellen innerhalb und zwischen den Blocks (alternativ ist auch die Pluralform *Blöcke* gebräuchlich; in Österreich und der Schweiz nur *Blöcke*), sowie Berücksichtigung der Korrelation innerhalb der Parzellen – wird in 3.5 (a) als Ansatz 1 bezeichnet und dort weiter ausgeführt. Auswertungen auf der Basis der Einzelmessungen sind durch geostatistische Ansätze möglich (siehe Abschnitt 3). Einfacher strukturiert ist ein Modell, das auf den Parzellenmittelwerten beruht, wozu die Einzelmessungen einer Parzelle/Randomisationseinheit zu einem Mittelwert zusammengefasst werden und nur diese in die anschließende Analyse einfließen. Dabei geht die Streuung der Einzelmessungen innerhalb der Parzellen und deren Kovarianz in die Streuung der Parzellenmittelwerte ein. Durch die Mittelwertbildung geht keinerlei Information bezüglich der Behandlung verloren, falls die Varianz der Mittelwerte bei den weiteren Berechnungen verwendet wird.

Blockbildung verfolgt das Ziel, durch eine räumliche Zusammenfassung solcher Parzellen, von denen eine einheitliche Wirkung auf die Prüfmerkmale erwartet wird (einheitliche Ausgangsbedingungen vor Versuchsbeginn), systematische Einflussfaktoren wie bspw. Bodenunterschiede zu berücksichtigen. Bei Nutzung von Vorinformationen über die ausgewählten Versuchsflächen sind die Blocks so zu wählen, dass bezüglich der natürlichen Bedingungen innerhalb der Blocks eine hohe Einheitlichkeit vorliegt. Durch die gleichzeitige Nutzung des Blocks als Organisationsprinzip bei der Versuchsdurchführung werden auch aus dieser Sicht für die Parzellen eines Blocks einheitliche Effekte angenommen. Auf Grund der Schwierigkeiten bei der Gewährleistung der notwendigen Einheitlichkeit sollte in OFE die Anzahl der Prüfglieder relativ klein sein, und i. A. wird es sich um einfaktorielle Versuche mit zwei bis vier Prüfgliedern handeln. Entspricht die Anzahl der Parzellen je Block der Anzahl der Prüfglieder und die Prüfglieder werden den Parzellen innerhalb der Blocks zufällig zugeordnet, so handelt es sich um eine randomisierte Anlage mit vollständigen Blocks. In diesem Fall ist die

Anzahl der Wiederholungen eines Prüfglieds gleich der Anzahl der vollständigen Blocks. Wenn die Anzahl der Prüfglieder jedoch größer als die Anzahl der in einem Block liegenden Parzellen ist, so sollten randomisierte Anlagen mit unvollständigen Blocks gewählt werden. Bei technisch-technologischer Machbarkeit kann es auch sinnvoll sein, eine zweifache Blockbildung im Sinne einer Zeilen-Spalten-Anlage vorzunehmen. Bei der Konstruktion von unvollständigen Blockanlagen und Zeilen-Spalten-Anlagen sind einige Besonderheiten zu beachten, welche die Einholung methodischer Expertisen erfordert (JOHN und WILLIAMS, 1995). Die meist genutzte Anlage für OFE wird die randomisierte Anlage mit vollständigen Blocks sein.

Gibt es darüber hinaus quantitative Störgrößen, die die Ausprägung des Prüfmerkmals beeinflussen, aber nicht ursächlich mit der Wirkung der Prüfglieder zusammenhängen, so können diese bei der Auswertung als Kovariable berücksichtigt werden. Das betrifft vor allem die Prüfung der Notwendigkeit einer Beachtung von Bodentrends im Auswertungsmodell. Ob diese Bodentrends mit Hilfe von Blocks oder Zeilen/Spalten schon ausreichend gut abgebildet werden können oder eine Berücksichtigung über Funktionen mit den räumlichen Koordinaten als Kovariable erfolgen muss, ist wiederum im Rahmen der Modellbildung zu entscheiden (siehe Abschnitt 3).

Ebenso wie in Parzellenversuchen ist auch in OFE eine detaillierte Verlaufsdokumentation und Erfassung relevanter Zusatzmerkmale von großer Bedeutung. Die Zusatzmerkmale können bei der inhaltlichen Interpretation der Versuchsergebnisse herangezogen werden oder eventuell auch als Kovariablen in Betracht kommen. Dazu gehören: Merkmale zur Charakterisierung der Pflanzenentwicklung, des Witterungsablaufs, der Begleitflora sowie des Auftretens von Schaderregern und Krankheiten.

Meist sollten in einem Einzelversuch alle Wiederholungen auf einem Schlag angelegt werden. Werden die Wiederholungen (aus welchen Gründen auch immer) auf verschiedene Schläge verteilt, so wird es meist schwieriger sein, die Einheitlichkeit der Ausgangsbedingungen und die der Versuchsdurchführung für alle Schläge zu gewährleisten (Nichteinhaltung des ceteris-paribus-Prinzips). Um eine Einheitlichkeit jedoch bestmöglich zu gewährleisten, wird man in dieser Situation die verschiedenen Schläge bewusst auswählen. Dabei können zwei Situationen unterschieden werden:

1. *Wenn ein Schlag einem (vollständigen oder unvollständigen) Block entspricht*, dann entsprechen Blockunterschiede den Unterschieden zwischen den Schlägen. Eventuell auftretende Wechselwirkungen zwischen Schlägen (Blocks) und Prüfgliedern sind mit den zufallsbedingten Effekten vermengt. Eine Trennung der beiden Komponenten ist nicht möglich.
2. *Wenn je Schlag mehrere Wiederholungen angelegt werden*, dann können Wechselwirkungen zwischen Schlägen und Prüfgliedern geprüft werden, was hinsichtlich der unter Umständen nicht völlig einheitlichen Ausgangsbedingungen von Interesse sein kann (aber evtl. auch die Interpretation erschwert).

Wenn also der Versuch auf mehreren Schlägen angelegt wird, dann ist die zweite Situation vorzuziehen. In jedem Fall sollte nicht ein Schlag mit einem Prüfglied identifiziert und dieses nur auf diesem einen Schlag angelegt werden, da dann Prüfgliedunterschiede mit Schlagunterschieden untrennbar vermengt sind.

In Versuchsserien gelten i. A. die zuvor getroffenen Aussagen für jeden einzelnen in der Serie enthaltenen Standort und jedes Jahr. Wird jedoch die Serie so konzipiert, dass die Schläge eine

Zufallsstichprobe aus einer Region repräsentieren, dann kann in speziellen Situationen entsprechend den Ausführungen in Abschnitt 2.3 verfahren werden. Randomisationen sind für jeden Einzelversuch der Serie erneut auszuführen.

2.3 Versuchsanlagen für OFE

Aus Abschnitt 2.2 geht bereits hervor, dass wir als Versuchsanlagen für OFE randomisierte Blockanlagen mit vollständigen Blocks, u. U. mit unvollständigen Blocks und in Spezialfällen Zeilen-Spalten-Anlagen empfehlen. Der Unterschied gegenüber den gleichfalls für Kleinparzellenversuche empfohlenen Anlagen ergibt sich aus den in 2.2 (a) und (b) genannten Besonderheiten.

Der von Versuchsanstellern/Praktikern häufig genutzte Begriff *Streifenversuch* (THOMAS, 2006) oder *Streifenanlage* (STEIN-BACHINGER et al., 2000), wird auch in der internationalen Literatur (strip trial) nicht einheitlich verwendet. Allgemein wird darauf fokussiert, dass die Prüfglieder auf Parzellen, die lange Streifen über den Schlag bilden, angelegt sind (vgl. 2.2 (a)). Es gibt keine einheitliche Aussage darüber, ob Wiederholungen der Prüfglieder vorliegen, ob die Parzellen in Blocks angelegt sind und ob innerhalb der Parzellen mehrere Beobachtungen vorliegen. In STEIN-BACHINGER et al. (2000) wird teilweise von einer Blockstruktur, teilweise von nicht vorhandenen Wiederholungen ausgegangen. Neben dieser nicht einheitlichen Definition kann es aber auch mit dem sonst im Versuchswesen üblichen Begriff der Streifenanlagen (split-block design/strip-plot design) für zweifaktorielle Versuche mit Wiederholungen zu Verwechslungen kommen. Über die konkrete Anlage sagt der Begriff nichts aus. Nach unserem Verständnis kann darauf verzichtet werden.

Gleichfalls kritisch zu bewerten ist der so genannte *Fensterversuch* (STEIN-BACHINGER et al., 2000) oder die „Fenstermethode“ (THOMAS, 2006). Bei diesem Versuchstyp sind in eine einheitlich bewirtschaftete Fläche Parzellen (Fenster) eingefügt, auf denen die Prüfglieder appliziert sind. Entscheidend ist hierbei, wie diese Fenster gesetzt werden. Erfüllt das Anlagenschema die oben genannten Grundsätze, so kann dieser Versuch als Einzelversuch im Rahmen eines OFE, ansonsten nur als Demonstrationsversuch bewertet werden. Im letzteren Fall sollten die Ergebnisse nur mit Methoden der beschreibenden Statistik ausgewertet werden.

Unter den in 2.2 formulierten Grundprinzipien sind gleichfalls die Langparzellenanlage und der ein- oder zweifaktorielle kontrollierte Anbauvergleich (für ein oder zwei Prüffaktoren mit jeweils zwei Stufen) kritisch zu bewerten. Das Prinzip der Randomisation ist verletzt und es liegen keine echten Wiederholungen vor (THOMAS, 2006). Anlagen, welche die Grundprinzipien aus 2.2 verletzen, dienen im Wesentlichen der Demonstration.

Wird ein OFE als Versuchsserie geplant, so ergeben sich einige Besonderheiten in Abhängigkeit von dem geplanten Aussagebereich entsprechend Tabelle 1. Dabei wird zunächst von vollständigen Blocks ausgegangen. Während ein Einzelversuch mit einem vollständigen Block (auch bei mehreren Einzelmessungen je Parzelle) keine verallgemeinerungsfähige Aussagekraft hat und bestenfalls als Demonstrationsversuch gewertet werden kann, können bei der Einbeziehung mehrerer Orte eines Anbaugesbietes mit je einem Schlag und je einem vollständigen Block (Schlag=Block) auch ohne Vorliegen wiederholter Messungen durchaus Informationen über die Prüfgliedwirkung gewonnen werden. Dieses setzt jedoch eine ausreichende Anzahl von Orten mit möglichst gleichen Produktions- und Umweltbedingungen voraus, da Wechselwirkungen zwischen Prüfgliedern und Orten nicht geprüft werden können. SVÁB (1957) fordert für diesen Fall 10 bis 20 Einzelversuche (siehe THOMAS, 2006). Da aber oft gerade die Wechselwirkungen der Prüfglieder mit den anderen Faktoren von Interesse sind, müssen die Einzelversuche so geplant werden, dass die interessierenden Wechselwirkungen

schätzbar sind. Am Beispiel einer mehrortigen Serie ohne wiederholte Messungen sind vier verschiedene Situationen in Tabelle 2 zusammengefasst.

Tabelle 2: Mehrortige Versuchsserie eines OFE

Situation	Anzahl der Schläge/Ort	Anzahl vollständiger Blocks je Schlag	mögliche zu testende Wirkung
a	1	1	Prüfglied und Ort
b	1	>1	Prüfglied, Ort, Block in Ort und WW Prüfglied x Ort
c	>1	1	Prüfglied, Ort, Schlag in Ort, WW Prüfglied x Ort
d	>1	>1	Prüfglied, Ort, Schlag in Ort, Block in (Schlag x Ort), WW Prüfglied x Ort, WW Prüfglied x Schlag in Ort

Werden die Orte entsprechend einer hierarchischen Struktur aus mehreren Anbaugebieten ausgewählt (Anbaugebiet/Ort) und je Ort liegt ein Schlag mit einem vollständigen Block vor (Tabelle 2 Situation (a) erweitert auf mehrere Anbaugebiete), so ist bei der Serienauswertung die Hauptwirkung der Anbaugebiete und die Wechselwirkung Prüfglied x Anbaugebiet schätzbar, jedoch wiederum nicht die Wechselwirkung Prüfglied x Ort, weil sie auch hier mit dem Versuchsfehler vermengt ist.

Das britische Großforschungsprojekt „Farm Scale Evaluation“ (FSE) zur Analyse der Effekte des Anbaus von verschiedenen herbizidresistenten GMO auf die floristische und faunistische Biodiversität wurde 1998 bis 2002 durchgeführt. PERRY et al. (2003) diskutieren, ob es für diese Fragestellung (Prüfung von GMO gegen konventionelle Variante) günstiger ist, die beiden Varianten an einem Ort auf einem Schlag je zur Hälfte (half-fields) oder auf zwei verschiedenen Schlägen (paired fields) anzulegen. Trotz einiger Nachteile wird die half-fields-Anlage gewählt (PERRY et al., 2003; ROTHERY et al., 2003). Je Jahr und GMO wurde der Versuch auf 60 jährlich wechselnden Schlägen angelegt. Je Jahr entspricht die Versuchsstruktur der geschilderten Hierarchie Anbaugebiet/Ort, wobei hier zusätzlich wiederholte Messungen je Parzelle vorlagen.

Es ist erkennbar, dass bezüglich der Versuchsstruktur sehr viele spezielle Situationen vorliegen können, wozu hier nur einige Beispiele aufgeführt wurden.

Werden in einem OFE mehr als zwei Varianten geprüft, so sind auch Anlagen mit unvollständigen Blocks denkbar. Damit in der Anlage mit unvollständigen Blocks noch alle Prüfglieddifferezenzen auf der Basis der Intra-Block-Information schätzbar sind, muss es sich um einen verbundenen Plan handeln. Die Prinzipien für Versuchsanlagen mit unvollständigen Block (JOHN and WILLIAMS, 1995) gelten auch für OFE.

2.4 Modellbildung für OFE

Die Festlegung der neben den Prüffaktoren in das Modell aufzunehmenden festen und zufälligen Effekte ist bei OFE wegen einer schlechteren Möglichkeit der Kontrolle von Einflussfaktoren besonders bedeutsam. Die Vernachlässigung dieses Gesichtspunktes kann zu einem Verlust an Treffgenauigkeit und Verlust an Präzision führen, wobei der erstgenannte Aspekt besonders bedeutsam ist. Im Zusammenhang mit einer Serienauswertung ist zu beachten, dass

in den einzelnen Umwelten verschiedene Modelle zur Berücksichtigung des Randomisierungsplanes sowie von räumlichen Trends möglich sind. Das kann beispielsweise der Fall sein, wenn Bodentrends in den einzelnen Umwelten verschieden bedeutsam sind. Das kann zu komplexen Modellen führen, deren Entwicklung einen gesonderten Analyseschritt darstellt.

Oft bezieht sich die Modellselektion sowohl auf die Auswahl der Erwartungswertstruktur für die festen Effekte als auch auf die der Kovarianzstruktur für die zufälligen Effekte. In Abhängigkeit vom zu bearbeitenden Problem kann die Anzahl der jeweils möglichen Ansätze sowohl für die Erwartungswert- als auch Kovarianzstruktur und damit die Zahl der möglichen Kombinationen groß sein. Aus diesem Grund wird neben einer gemeinsamen Optimierung von Erwartungswert- und Kovarianzstruktur (NGO and BRAND, 1997) auch eine mehrstufige Vorgehensweise vorgeschlagen (WOLFINGER, 1993).

Dabei wird auch aus Sicht der Praktikabilität die folgende Schrittfolge empfohlen:

1. Schritt: Auswahl der Erwartungswertstruktur bei Unterstellung von unabhängigen Resteffekten mit homogener Varianz, geschätzt gemäß Gewöhnlicher Kleinsten Quadrate – Methode (Ordinary Least Squares, OLS).

2. Schritt: Auswahl der Kovarianzstruktur bei Verwendung der Erwartungswertstruktur gemäß Ergebnis des ersten Arbeitsschrittes.

Die vorgeschlagene Vorgehensweise erlaubt in Schritt 1 die Verwendung der OLS-Residuen für jedes untersuchte Modell zur Kontrolle auf systematische Veränderungen der Residuen, was auf die Nichtbeachtung bedeutsamer fester Effekte im Auswertungsmodell hinweist. Das ist bei Beobachtungen in verschiedenen Umwelten und Teilflächen besonders anschaulich möglich. Die Festlegung der Erwartungswertstruktur wird sich im vorliegenden Zusammenhang vor allem auf die Berücksichtigung von Bodentrends beziehen. Die Verteilung der OLS-Residuen bspw. in verschiedenen Umwelten gibt weiterhin einen Hinweis auf eine mögliche Varianzheterogenität. Schließlich kann die für die Residuen beobachtete Kovarianzstruktur mit der im Ergebnis von Schritt 2 geschätzten Struktur bei Nutzung der Schätzwerte verglichen werden. Dabei sollte die geschätzte Kovarianzstruktur im Wesentlichen die Verteilung der OLS-Residuen abbilden. Um diese Vorgehensweise realisieren zu können, wird eine umgekehrte Schrittfolge gegenüber dem Vorschlag von WOLFINGER (1993) empfohlen, der zunächst eine Optimierung der Kovarianzstruktur vornimmt.

Für einen Modellvergleich stehen analytische Kriterien zur Verfügung. Häufig benutzte Informationskriterien sind das Kriterium von AKAIKE (1974) und seine Modifikation von HURVICH and TSAI (1989) sowie das Kriterium von SCHWARZ (1978). Der Vergleich verschiedener Erwartungswertstrukturen sollte auf der Grundlage der Maximum-Likelihood-Methode und den auf dieser Basis berechneten Kriterien erfolgen. Verschiedene Kovarianzstrukturen bei identischer Erwartungswertstruktur können anschließend mittels der restringierten Maximum-Likelihood-Methode (REML-Methode) und darauf basierenden Kriterien bewertet werden. Die Informationskriterien erlauben eine Rangierung der Modelle, wobei je nach Berechnungsvorschrift das Modell mit dem kleinsten bzw. größten Wert zu bevorzugen ist. Eine Untersuchung der Wirksamkeit dieser Vorgehensweisen für Parzellenversuche ist bei SPILKE et al. (2010) zu finden.

2.5 Auswertung von OFE ohne Georeferenzierung

Die Auswertung von OFE erfolgt für quantitative Zielgrößen mit annähernder Normalverteilung der zufälligen Effekte mit einem Varianzanalysemodell, bei zusätzlicher Berücksichtigung von Kovariablen mit einem Kovarianzanalysemodell entsprechend den Ergebnissen der Modellbildung nach Abschnitt 2.4. In diesem Abschnitt soll zunächst die Auswertung von Daten ohne Georeferenzierung behandelt werden; auf den Fall mit Georeferenzierung wird im Abschnitt 3 eingegangen.

Falls von der Normalverteilung abweichende Verteilungen vorliegen oder keine bekannte Verteilung angepasst werden kann, können generalisierte lineare Modelle bzw. Rangverfahren verwendet werden. Auf die mit diesen Auswertungsverfahren verbundene Spezifik kann hier nicht eingegangen werden.

Häufig ist eine Betrachtung der erhaltenen Prüfglieddiffenzen und der zugehörigen Konfidenzintervalle einem reinen Signifikanztest und einer Fokussierung auf Signifikanzen vorzuziehen. Die Testentscheidung kann hierbei einfach aus dem Konfidenzintervall einer Differenz hergeleitet werden, wobei bei einem Konfidenzniveau von $1-\alpha$ das Signifikanzniveau α beträgt. Enthält das Intervall die Null, so ist die Differenz nicht signifikant von Null verschieden. Ist sie dagegen signifikant, enthält das Intervall die Null nicht. In diesem Fall ist es besonders hilfreich und verbessert die Interpretation, sich den überdeckten Bereich des Konfidenzintervalls anzuschauen. Es kann z. B. sein, dass dieses zwar die Null nicht enthält, aber sehr schmal ist und zudem so nahe an der Null liegt, dass die Differenzen als praktisch irrelevant zu betrachten sind. Gleiches gilt für Kleinparzellenversuche.

In den Abschnitten 2.5.1 und 2.5.2 werden Auswertungsmodelle für OFE ohne georeferenzierte Daten beschrieben, wobei angenommen wird, dass eine Anlage mit Blocks gewählt wurde. Diese Modelle werden später für georeferenzierte Daten erweitert. Die der Auswertung zugrunde liegenden Modelle werden für den Fall der Existenz von Kovariablen beschrieben (Kovarianzanalysemodell). Beispielhaft seien $kv1$ und $kv2$ zwei Kovariablen. Sie können linear, quadratisch oder in beliebiger polynomialer Form in das Modell eingehen; auch Wechselwirkungen sind möglich, etwa in der Form $kv1*kv2$.

Müssen nach Abschnitt 2.4 keine Kovariablen berücksichtigt werden, so entfallen alle Terme, die diese enthalten, und das Modell reduziert sich auf ein Varianzanalysemodell.

2.5.1 Auswertung von Einzelversuchen ohne Georeferenzierung

Es werden nur einfache Modelle angeführt, um das Prinzip zu zeigen. Für Erweiterungen muss auf die Anwendungsbeispiele des Workshops verwiesen werden. Die Modelle werden im Folgenden in der SAS-Notation beschrieben, da sie leichter als andere Schreibweisen (welche mehr mathematische Symbolik verwenden) verständlich ist.

- ohne wiederholte Messungen je Parzelle (1 Parzelle je Block und Prüfglied)

```
CLASS variante block
MODEL merkmal=variante block kv1 kv2
```

- mit wiederholten Messungen je Parzelle (1 Parzelle je Block und Prüfglied)

```
CLASS variante block parzelle
MODEL merkmal=variante block kv1 kv2
RANDOM parzelle
```

Diese Modellierung setzt eine konstante Korrelation der Messfehler innerhalb der Parzelle voraus, was den tatsächlichen Gegebenheiten widersprechen kann, wenn räumliche Abhängigkeiten bestehen. In diesen Fällen ist eine geostatistische Modellierung vorzuziehen, was eine Georeferenzierung der Messwiederholungen voraussetzt (siehe Abschnitt 3).

Werden die wiederholten Messungen zu einem Parzellenmittelwert zusammengefasst, so ist wie im Fall ohne wiederholte Messungen zu verfahren.

2.5.2 Auswertung von Versuchsserien

Für diesen Abschnitt gilt ebenso, dass nur vergleichsweise einfache Modelle kurz beschrieben werden können. Gerade für die Auswertung von Versuchsserien gilt, dass eine sehr sorgfältige Modellbildung vorzunehmen ist und die resultierenden Modelle sehr spezifisch sein können. Deshalb muss auch hier auf die Beispielbibliothek verwiesen werden.

Die Spezifik und Notwendigkeit einer differenzierten Betrachtung ergibt sich auch aus dem Sachverhalt, wonach im Allgemeinen davon ausgegangen werden kann, dass die betriebsübliche Variante in allen Jahren/Orten unterschiedlich gestaltet wird. So ist neben der ohnehin schon oft auftretenden Wechselwirkung Variante x Umwelt (Umwelt entspricht Jahr x Ort) mit deren zusätzlicher Erhöhung durch die betriebsübliche Variante zu rechnen. Im Rahmen der Modellbildung muss überprüft werden, ob beispielsweise alle Varianten die gleiche Fehler-Varianz und die gleiche Interaktions-Varianz haben (evtl. hat *betriebsüblich* eine größere Varianz) und insbesondere bei mehrjährigen Versuchsserien sollte in Betracht gezogen werden, dass häufig bezüglich der Jahre Varianzheterogenität besteht.

Die nachfolgenden Fälle beziehen sich auf die Situation, dass je Variante nur eine Beobachtung je Block vorliegt - *ohne wiederholte Messungen*

mehrere fixe Orte (eine Kovariable, ortsspezifisch; Varianzhomogenität der Fehlereffekte) und Situation b aus Tabelle 2

```
CLASS variante ort block
MODEL merkmal=variante ort variante*ort block(ort) kv1*ort
```

mehrere zufällige Orte (Varianzheterogenität der Fehlereffekte) und Situation b aus Tabelle 2

```
CLASS variante ort block
MODEL merkmal=variante
RANDOM INT variante block/SUB=ort
REPEATED / SUB=ort GROUP=ort
```

Das obige Modell geht von einer einstufigen Auswertung aus. Dabei ist zu beachten, dass schon für der Fall heterogener Restvarianzen in Abhängigkeit der Datenstruktur (Anzahl Orte, Blocks innerhalb Ort etc.) lange Rechenzeiten auftreten können. Weitere ggf. notwendige Modellverfeinerungen wie heterogene Blockvarianzen zwischen den Orten oder unterschiedliche Versuchspläne von Ort zu Ort zwingen oft zu einer zweistufigen Auswertung. Dabei wird in der ersten Stufe der Einzelversuch ausgewertet. Die in der ersten Stufe berechneten Mittelwerte gehen dann in eine Serienauswertung ein, wobei die Standardfehler der Mittelwerte als Gewichte berücksichtigt werden (MÖHRING und PIEPHO, 2009).

3 Besonderheiten von OFE mit georeferenzierten Daten

Versuche mit georeferenzierten Daten stellen eine spezifische Klasse innerhalb von OFE dar. Sie weisen mehrere besonders zu berücksichtigende Sachverhalte auf und werden daher gesondert besprochen. Für OFE mit georeferenzierten Daten besteht auf Grund der relativ neuen Problemstellung ein besonderer Erklärungsbedarf bezüglich der Gestaltung der Phasen Versuchsplanung, -auswertung und -interpretation.

3.1 Versuchsfrage und daraus abzuleitende Anforderung an die Versuchsfläche

- (a) Oft besteht das Ziel eines OFE mit georeferenzierten Daten in dem Vergleich von Prüfgliedern mit und ohne teilflächenspezifische Applikation/Bearbeitung (kurz: *teilfl/einh*). Diese als Precision-Farming-Experimente (PFE) bezeichneten Versuche (s. Abbildung 1) erfordern zur Prüfung teilflächenspezifischer Arbeitsweisen heterogene Flächen bzw. Bestände. Die Variante *einh* entspricht meist der betriebsüblichen Variante.
- (b) OFE mit georeferenzierten Daten, bei denen kein Prüfglied mit teilflächenspezifischer Applikation existiert, sollten zur Gewährleistung des ceteris-paribus-Prinzips auf ausgewählten, möglichst homogenen Flächen angelegt werden.
- (c) Pro Parzelle liegen im Allgemeinen mehrere, oft sehr viele, georeferenzierte Einzelmessungen vor, die eine geostatistische Auswertung ermöglichen.

3.2 Grundsätze der Versuchsplanung und einige Konsequenzen für die Auswertung bei georeferenzierten Daten

Bei diesen Versuchen gelten die gleichen Prinzipien wie zuvor allgemein für OFE beschrieben. Die Möglichkeit geostatistischer Auswertungen bedeutet nicht, dass diese elementaren Grundsätze der Versuchsplanung verzichtbar oder außer Kraft zu setzen wären (COX, 2009).

Randomisation

Die Anwendung der Randomisation erlaubt zunächst eine Auswertung von Parzellenmittelwerten mittels eines Modells mit unabhängigen Resteffekten. Diese erfolgt je nach Randomisation in der Regel entsprechend einer Block- bzw. Zeilen/Spaltenanlage. Die Randomisationstheorie stellt sicher, dass die hieraus erzielten Schlüsse valide sind. Diese klassische Auswertung kann dann als Vergleichsmaßstab für weitergehende geostatistische Auswertungen unter Einbeziehung von Kovariablen sowie von Messwiederholungen je Parzelle herangezogen werden, welche keine randomisations-theoretische Grundlage haben (siehe Abschnitt 2.4). Die Auswertung auf Basis der Mittelwerte ist gerechtfertigt, wenn zu Versuchsbeginn die in einem Block liegenden Parzellen auf „gleichmäßig heterogenen“ bzw. homogenen Flächen liegen. Von homogenen Flächen wird in den Versuchen nach 3.1 (b) ausgegangen. Bei Versuchen nach 3.1 (a) kann „gleichmäßige Heterogenität“ angenommen werden, wenn sich die Parzellen eines Blocks vor Versuchsbeginn/vor der Applikation sowohl im Mittelwert als auch der Varianz von geeigneten zu wählenden Boden- und/oder Pflanzenbestandsmerkmalen nicht unterscheiden. Dieses kann vor der Versuchsdurchführung verifiziert werden, wenn nach dem Kartenansatz appliziert wird (siehe unten) oder kann *ex-post* bei einer Applikation nach Sensorsansatz bewertet werden. Kann diese gleichmäßige Heterogenität nicht gewährleistet werden, so siehe Absatz „Blockbildung und weitere Gruppierung“ in diesem Abschnitt.

Wiederholung

Grundsätzlich ist das Prinzip des Wiederholens auch hier anzuwenden. Die geostatistische Referenzierung der Beobachtungen, insbesondere auch bei Messwiederholungen je Parzelle, ermöglicht weiterführende Auswertungen durch geostatistische Ansätze (siehe Abschnitte 3.4 und 3.5), ersetzt aber nicht die Notwendigkeit echter Wiederholungen.

Blockbildung und weitere Gruppierung

Auch bei OFE mit georeferenzierten Daten sollten räumlich benachbarte Parzellen in Blocks angelegt werden. In Versuchen nach 3.1 (b) mit gezielt ausgewählten homogenen Flächen ist es relativ wahrscheinlich, dass die räumliche Nähe für eine Gruppierung genutzt werden kann, was eine Gewährleistung des *ceteris-paribus*-Prinzips ermöglicht.

Bei den unter 3.1 (a) besprochenen PFE müssen drei verschiedene Herangehensweisen bei der Applikation der Prüfglieder unterschieden werden: Kartenansatz (Offline-Ansatz), Sensoransatz (Online-Ansatz) und eine Kombination beider.

Beim *Kartenansatz* (z. B. auf Grundlage von Ertragskarten vorangegangener Jahre oder vorliegender EC₂₅-Karten) können vor Versuchsbeginn gezielt Teilflächen ausgewählt werden, die gleiche Bedingungen für den Vergleich der Prüfglieder erwarten lassen. In diesen sog. Managementzonen können die Prüfglieder in einfacher oder mehrfacher Wiederholung in Blocks angelegt werden. Der Vorteil einer mehrfachen Wiederholung liegt in der möglichen Bewertung je Managementzone. Weiterhin kann bei der Auswertung über alle Zonen (Betrachtung von „Managementzone“ als zusätzlichen Prüffaktor) die Wechselwirkung zwischen Prüfgliedern und Managementzonen untersucht werden. Die Bildung von Managementzonen entspricht einer vor Versuchsbeginn erfolgten weiteren Gruppierung der Versuchsfläche.

Beim *Sensoransatz* wird erst bei der Applikation der Prüfglieder die Heterogenität der Fläche/des Bestands quantifiziert. Die Sensormessung wird beim PG *teilfl* direkt in eine Applikationsempfehlung umgesetzt, beim PG *ein* zur Kontrolle der Ausgangssituation genutzt. Hier wird das *ceteris-paribus*-Prinzip häufig nicht bzw. nicht ausschließlich über die räumliche Nähe gewährleistet sein. In diesem Fall kann *ex-post* (zusätzlich zu den *ex-ante* gebildeten Blocks) eine Gruppierung durch Diskretisierung der Sensormessung definiert werden. Bei dieser Diskretisierung sind verschiedene Vorgehensweisen denkbar. Wichtig ist in jedem Fall, dass je Gruppierungseinheit ein ausreichender Stichprobenumfang existiert. Die zu einer Gruppierungseinheit gehörenden Beobachtungswerte können räumlich weit auseinander liegen und sind oft nicht an die Blockstruktur gekoppelt. Bei der Auswertung stellen diese zusätzlichen Gruppen, analog zum Kartenansatz, die Stufen eines weiteren Faktors dar. In der Regel wird ein Prüfgliedvergleich je *Ex-post*-Gruppe sinnvoll sein. Ob dabei die Berücksichtigung von Blockeffekten möglich und im Sinne einer besseren Modellanpassung sinnvoll ist, muss im Einzelfall entschieden werden. Der Vergleich je Gruppe ermöglicht eine inhaltliche Bewertung des Applikationsalgorithmus. Da die Sensormessung ursächlich das PG *teilfl* bestimmt (u. U. nicht entsprechend einer linearen Funktion), jedoch das PG *ein* davon unabhängig ist, darf sie nicht als Kovariable in eine Kovarianzanalyse zwecks Ausschaltung unterschiedlicher Ausgangsbedingungen eingehen.

Werden *beide Ansätze* miteinander *kombiniert*, kann zunächst analog zum Kartenansatz der Versuch angelegt werden. Ergebnisse aus den Sensormessungen können evtl. mit der Karte verschnitten werden, so dass dadurch detailliertere Straten definiert werden. Zusätzliche Kovariablen können u. U. die Treffgenauigkeit sichern und Präzision erhöhen.

3.3 Datenaufbereitung

In Versuchen mit georeferenzierten Daten ist eine kritische Bewertung aller erhobenen Merkmalsdaten (u. a. Applikations- und Ertragsdaten) unerlässlich. Filter (automatische und/oder visuelle) sollten zur Plausibilitätskontrolle eingesetzt werden, um unmögliche/unplausible Daten und Daten aus dem Vorgewende auszuschließen. Messungen, deren Qualität durch die Fahrgeschwindigkeit beeinflusst ist, sollten bei extremen Abweichungen von der durchschnittlichen Fahrgeschwindigkeit gleichfalls aus den Datensätzen entfernt werden. Mögliche Nachkalibrierungen auf der Basis der je Parzelle ausgebrachten Gesamtapplikationsmengen und –ertragswerte sind sinnvoll.

3.4 Aggregation und Zusammenführung von Merkmalsdaten

Häufig liegen Applikations- und Ertragsdaten sowie Daten von Kovariablen nicht von exakt gleichen Positionen bzw. in unterschiedlich erfasster Intensität vor. Eine räumliche Zuordnung entsprechend der Parzellenstruktur sollte jedoch gewährleistet sein. In einem solchen Fall müssen die Daten teilweise verdichtet und auf gleiche Positionen bezogen werden.

Zwei gebräuchliche Möglichkeiten sind:

(a) *Rasterbezug*

Ein einheitliches Raster wird über die Gesamtversuchsfläche gelegt, so dass in jeder Rasterzelle Daten von jedem Merkmal vertreten sind. Die in einer Zelle vertretenen Daten eines Merkmals werden gemittelt und auf den räumlichen Zellenmittelpunkt bezogen.

(b) *Kriging*

Hierbei wird zunächst ein Merkmal („Basismerkmale“) mit seinen Positionen ausgewählt (in der Regel das Merkmal mit der niedrigsten Erfassungsintensität). Für alle anderen Merkmale werden deren Werte auf den durch das Basismerkmale festgelegten Positionen vorhergesagt. Das Punkt-Kriging ist ein Verfahren der Vorhersage von Merkmalsdaten an nicht beprobten Positionen, während das Block-Kriging zur Vorhersage für nicht beprobte Flächeneinheiten geeignet ist. Grundlage beider Kriging-Verfahren ist die Anpassung eines räumlichen Modells an das empirische Semivariogramm der Daten des vorherzusagenden Merkmals. Dabei sollte aus dem Datensatz des vorherzusagenden Merkmals ein Teildatensatz so ausgewählt werden, dass die vorherzusagenden Positionen möglichst nahe an den beobachteten Positionen liegen (im Allgemeinen sollte der Teildatensatz aus der jeweiligen Parzelle genutzt werden). Die geschätzten Vorhersagewerte sind beste lineare erwartungstreue Vorhersagen (BLUP). Für jede Position kann der Vorhersagefehler durch die Kriging-Varianz beschrieben werden. Wird bei der anschließenden Auswertung mit den gekrigten Werten gearbeitet, so kann dieses mit oder ohne Wichtung durch die Kriging-Varianz erfolgen. Für den Fall, dass sich die Kriging-Varianzen (Vorhersagefehler) deutlich unterscheiden, sollte mit Wichtung gearbeitet werden. Die gekrigten Werte sind in der Regel sowohl innerhalb der Parzellen, aber u. U. auch zwischen den Parzellen korreliert. Deshalb muss bei deren Analyse neben der eventuellen Wichtung eine erneute Anpassung an ein geostatistisches Modell erfolgen. Welcher Ansatz für das geostatistische Modell gewählt wird, ergibt sich aus den Überlegungen im Abschnitt 3.5 (a).

Ein weiterer Grund für eine notwendige Aggregation kann bei sehr hoch aufgelösten georeferenzierten Messungen in dem großen Datenumfang und den damit verbundenen numerischen Problemen bezüglich der erforderlichen Rechentechnik und Rechenzeit bestehen. In solchen

Fällen ist eine Aggregation auf eine reduzierte Zahl von Teilflächen innerhalb der Parzelle sinnvoll. Die höchste mögliche Form der Aggregation besteht immer in der Berechnung von Parzellenmittelwerten, die dann klassisch ausgewertet werden können, sofern dieses nach den obigen Ausführungen zu Grundsätzen der Versuchsplanung gerechtfertigt ist (adäquate Randomisation etc.). Falls eine geostatistische Auswertung vorgenommen werden soll, ist wichtig, dass nach der Aggregation noch eine genügende Anzahl von Werten zur Verfügung steht, um geostatistische Kovarianzstrukturen verlässlich schätzen zu können.

3.5 Weitere Spezifika der Auswertung

Nach Datengewinnung sowie den Arbeitsschritten 3.3 und 3.4 kann die eigentliche Auswertung erfolgen, wozu in Abschnitt 3.2 schon einige Hinweise gegeben wurden.

(a) Hinweise zur möglichen geostatistischen Auswertung

Geostatistische Auswertungen auf Basis der Einzelmessungen je Parzelle bieten das Potential zu einer aussagekräftigeren Auswertung mit einem Gewinn an Teststärke und Präzision gegenüber einer klassischen Auswertung auf Basis von Parzellenmittelwerten. Die wesentliche Schwierigkeit liegt in der Identifikation von geostatistischen Modellen, welche die Daten adäquat abbilden, indem sie die räumliche Korrelation benachbarter Beobachtungen berücksichtigen und somit valide statistische Schlüsse erlauben. Dieses trifft gleichfalls auf die besprochene Modellanpassung zum Zwecke der Vorhersage an nicht beprobten Positionen bei 3.4 (b) zu. Eine Unsicherheit in der zu treffenden Modellselektion ist unausweichlich. Das ist die wesentliche Schwäche dieses Ansatzes. Es ist zu beachten, dass geostatistische Modellansätze keine Fundierung in einer Randomisationstheorie haben. Die Wahl eines inadäquaten geostatistischen Modells birgt daher immer die Gefahr einer verzerrten Schätzung von Behandlungseffekten, falls beispielsweise ein falscher Trend einbezogen wird, und von invaliden statistischen Schlüssen (Nichteinhaltung des Signifikanzniveaus etc.), falls die falsche Kovarianzstruktur verwendet wird. Aus diesem Grunde muss die Modellselektion mit besonderer Sorgfalt erfolgen.

Bei geostatistischen Auswertungen, die auf der Nutzung der Korrelation der Beobachtungen in Abhängigkeit von ihrer Distanz beruhen, sind verschiedene Ansätze denkbar:

1. Ausschließliche Berücksichtigung der Korrelation innerhalb der Parzellen,
2. Berücksichtigung der Korrelation innerhalb der Parzellen und Blocks,
3. Berücksichtigung der Korrelation über die gesamte Versuchsfläche.

Die Ansätze 1 und 2 stellen in gewisser Weise konservative Ansätze dar. Beim Ansatz 2 wird eine räumliche Kovarianz nur zwischen Messpunkten innerhalb von vollständigen Wiederholungen bzw. innerhalb von Blockeinheiten, innerhalb derer eine Randomisation stattgefunden hat, nicht jedoch zwischen solchen Einheiten modelliert. Noch weitergehender ist die Beschränkung der räumlichen Korrelation auf Messpunkte derselben Parzelle (Ansatz 1). Sind die wiederholten Messungen innerhalb einer Parzelle voneinander abhängig, so ist die effektive Wiederholungsanzahl je Parzelle umso niedriger, je höher deren Korrelation ist. Im Extremfall bedeutet das bei einer Korrelation von 1, dass die effektive Wiederholungsanzahl für dieses Prüfglied in dieser Parzelle nur 1 ist. Dann haben die Einzelmessungen von einer Parzelle in ihrer Gesamtheit keinen höheren Informationsgehalt als eine einzige Messung von der Parzelle. Die mit der Verwendung von Parzellenmittelwerten einhergehende Annahme der Unabhängigkeit zwischen den Versuchseinheiten steht im Einklang mit der Randomisationsstruktur des Versuches. Er bedingt zwar einerseits gegenüber Ansatz 3 potentiell einen

Verlust an Präzision der Auswertung, bietet aber andererseits einen gewissen Schutz gegen Verzerrungen durch Überanpassung (over-fitting). Die Annahme der Unabhängigkeit zwischen Versuchseinheiten reduziert zudem den Rechenaufwand für die Anpassung geostatistischer Modelle.

Selbst dann, wenn eine geostatistische Auswertung basierend auf Einzelmessungen vorgenommen wird, sollte die Randomisationsstruktur des Experiments im Modellansatz abgebildet sein. Dies bedeutet, dass das Modell unabhängige Effekte für Block- und Randomisationseinheiten (Parzellen) berücksichtigen sollte. Selbst wenn von einer homogenen Versuchsfläche ausgegangen werden kann, so ist es aufgrund der Versuchstechnik immer wahrscheinlich und keinesfalls auszuschließen, dass es an den Parzellen- oder Blockgrenzen Diskontinuitäten gibt. Im Rahmen einer Modellselktion (siehe Abschnitt 2.4) ist dann zu prüfen, ob solche Effekte im konkreten Fall in das Auswertungsmodell aufzunehmen sind oder nicht. Ohne eine solche kritische Prüfung sollten solche Designeffekte jedoch keinesfalls ignoriert und aus dem Modell genommen werden. Die generelle Berücksichtigung solcher Effekte bietet eine gewisse (aber keine vollständige) Versicherung gegen statistische Fehlschlüsse aufgrund eines inadäquaten statistischen Modells.

Voraussetzung für die Anwendung geostatistischer Verfahren ist, dass keine Trends auf der Fläche vorliegen und damit die sog. „Stationarität“ vorliegt. Dieses ist *a priori* in der Realität häufig nicht zu garantieren. So darf bei Ansatz 1 kein Trend innerhalb der Parzellen, bei Ansatz 2 nicht innerhalb der Blocks und bei Ansatz 3 keiner über die gesamte Versuchsfläche vorliegen. Oft ist die Einhaltung dieser Voraussetzung nicht gegeben. In diesen Fällen ist diese Voraussetzung durch die Aufnahme bspw. polynomialer oder Spline-basierter Ansätze für eine Regression auf die räumlichen Koordinaten der Beobachtungen zu sichern. Insbesondere im Ansatz 3 kann bereits durch die Berücksichtigung der Blocks in der Auswertung u. U. ein eventuell vorhandener Trend in dieser Richtung erfasst werden. Die Entscheidung über die Beachtung eines Trends ist ebenfalls Aufgabe der Modellbildung (siehe Abschnitt 2.4).

Beispielhaft wird nachfolgend angeführt, wie der zweite Fall in 2.5.1 bei den drei Kovarianz-Ansätzen erweitert werden kann. Für *modell* ist ein spezielles räumliches Modell einzusetzen, z. B. "sph" für das sphärische oder "pow" für das Power-Modell.

```

CLASS variante block parzelle
MODEL merkmals=variante block kv1 kv2
Ansatz 1: REPEATED/TYP=sp(modell) (x y) SUBJECT=parzelle
Ansatz 2: REPEATED/TYP=sp(modell) (x y) SUBJECT=block
Ansatz 3: REPEATED/TYP=sp(modell) (x y) SUBJECT=intercept

```

Auf die Bedeutung der Programmanweisungen und unterschiedlichen Vorgehensweisen der Freiheitsgradapproximation muss auf die Beispielbibliothek verwiesen werden. Eine ausführliche Beschreibung der Vorgehensweise zur Darstellung des statistischen Modells ist bei PIEPHO et al. (2003) zu finden.

(b) Konkretisierung der Zielstellung und Auswertung

Abhängig davon, welches konkrete Ziel im PFE mit dem Vergleich der Prüfglieder verbunden ist, sind neben den vorherigen Erläuterungen zur Auswertung weitere Besonderheiten zu beachten, die unmittelbar mit dem Aussagebereich verknüpft sind.

Soll im Sinne eines Demonstrationsversuchs eine Aussage für den konkreten Schlag unter Berücksichtigung seiner Flächenanteile erfolgen, so besteht das Hauptinteresse des Landwirts

u. U. an einer Gesamtaussage für den Schlag. Bei einer Applikation nach Kartenansatz gelangt man zu einer Gesamtaussage, wenn die Stufen des Faktors „Managementzone“ mit den entsprechenden Flächenanteilen gewichtet werden. Beim Sensoreinsatz und einer entsprechenden *Ex-post*-Gruppierung ist diese Zusammenfassung nur möglich, wenn die in den einzelnen Gruppen erhaltenen Stichprobenumfänge proportional zu den entsprechenden Flächenanteilen im Feld sind. Ob diese Proportionalität wirklich gegeben ist, wird oft ungeklärt bleiben.

Besteht das primäre Ziel in der Bewertung des Applikationsalgorithmus auf der Fläche (OFE im eigentlichen Sinne), so wird in der Regel die Bewertung je Managementzone bzw. je *Ex-post*-Gruppe von Interesse sein. Im Fall einer Bewertung je Managementzone sollte die Versuchsplanung anstreben, dass gleiche Stichprobenumfänge vorliegen, um die Effizienz zu optimieren. Im Fall einer Bewertung je *Ex-post*-Gruppe sind ungleiche Anzahlen unvermeidlich, was zu einem Effizienzverlust im Vergleich zu einem balancierten Design führt.

3.6 Hinweise zur Interpretation

Gerade bei über die Prozesstechnik erfassten georeferenzierten Daten liegt häufig ein sehr großer Datenumfang vor. Damit besteht die Gefahr, dass Prüfglieddiffenzen als signifikant ausgewiesen werden, die kleiner als die praktisch interessierende Mindestdifferenz sind. Gerade auch in einem solchen Fall muss das Testergebnis fachlich kritisch bewertet werden. Eine vorab durchgeführte Stichprobenumfangplanung wäre zwar hilfreich, stößt aber meist auf das Problem der fehlenden Vorinformation. In jedem Fall ist eine Berechnung von Vertrauensintervallen für erhaltene Prüfglieddiffenzen sinnvoll (siehe 2.3).

Literatur (zitierte Literatur und Auswahl methodisch orientierter Bücher)

- AKAIKE, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.
- BÄTZ, G.; DÖRFEL, H.; FUCHS, A.; THOMAS, E. (1982): Einführung in die Methodik des Feldversuchs. Landwirtschaftsverlag, Berlin.
- COX, D. R. (2009): Randomization in the design of experiments. *International Statistical Review* 77: 415-429.
- CRESSIE, N. A. C. (1993): *Statistics for spatial data*. Wiley & Sons, New York.
- HURVICH, C. M.; TSAI, C. L. (1989): Regression and time series model selection in small samples. *Biometrika* 76: 297-397.
- JOHN, J.A.; WILLIAMS, E.R. (1995): *Cyclic and computer-generated designs*. Chapman & Hall, London.
- MÖHRING, J.; PIEPHO, H. P. (2009): A comparison of weighting methods in two-stage analyses of series of experiments. *Crop Science* 49: 1977-1988.
- NGO, L.; BRAND, R. (1997): Model Selection in Linear Mixed Effects Models Using SAS Proc Mixed. SAS Users Group International (22) San Diego, California March: 16-19.
- NIELSEN, D. R.; WENDROTH, O. (2003): *Spatial and temporal statistics: sampling field soils and their vegetation*. Catena-Verlag, Reiskirchen.
- PERRY, J. N.; ROTHERY, P.; CLARK, S. J.; HEARD, M. S.; HAWES, C. (2003): Design, analysis and statistical power of the farm-scale evaluations of genetically modified herbicide-tolerant crops. *Journal of Applied Ecology* 40: 17-31.
- PIEPHO, H. P., BÜCHSE, A., EMRICH, K. (2003): A hitchhiker's guide to the mixed model analysis of randomized experiments. *Journal of Agronomy and Crop Science* 189: 310-322.

-
- PIEPHO, H.-P.; RICHTER, C.; SPILKE, J.; HARTUNG, K.; KUNICK, A. THÖLE, H. (2011): Statistical aspects of on-farm experimentation. *Crop & Pasture Science* 62: 721–735.
- ROTHERY, P.; CLARK, S. J.; PERRY, J. N. (2003): Design of the farm-scale evaluations of genetically modified herbicide-tolerant crops. *Environmetrics* 14: 711-717.
- SCHABENBERGER O.; PIERCE, F. J. (2002): Contemporary statistical models for the plant and soil sciences. Boca Raton, FL, CRC Press.
- SCHABENBERGER, O.; GOTWAY, C. A. (2005): Statistical methods for spatial data analysis. Boca Raton, FL, CRC Press.
- SCHWARZ, G. (1978): Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.
- SPECHT, G.; MÜLLER, K.-H. (1960): Probleme bei der Durchführung von Großflächenstreuversuchen. *Zeitschrift für landwirtschaftliches Versuchs- und Untersuchungswesen* 6: 148-163.
- SPECHT, G.; SCHULZE, J. (1963): Untersuchungen über die Anlage und Durchführung von Großflächenstreuversuchen. *Zeitschrift für landwirtschaftliches Versuchs- und Untersuchungswesen* 9: 405-426.
- SPILKE, J.; RICHTER, C.; PIEPHO, H. P. (2010): Model selection and its consequences for different split-plot designs with spatial covariance and trend. *Plant Breeding* 129: 590-598.
- STEIN-BACHINGER, K.; BACHINGER, J.; VÖGEL, R.; WERNER, A. (2000): Feldversuche: Leitfaden für Landwirte zur Durchführung produktionsbezogener Experimente. Rationalisierungs-Kuratorium für Landwirtschaft (RKL): 1072-1117.
- STROUP, W. W.; BAENZIGER, P. S.; MULITZE, D. K. (1994): Removing spatial variation from wheat yield trials: a comparison of models. *Crop Science* 86: 62-66.
- SVAB, J. (1957): Überlegungen zu methodischen Fragen der Großflächenversuche. *Zeitschrift für landwirtschaftliches Versuchs- und Untersuchungswesen* 3: 268-278.
- THOMAS, E. (2006): *Feldversuchswesen*. Ulmer, Stuttgart.
- TRAXLER, A.; HEISSENBERGER, A.; FRANK, G.; LETHMAYER, C.; GAUGITSCH, H. (2000): Ökologisches Monitoring von gentechnisch veränderten Organismen. Studie im Auftrag des Bundesministeriums für Umwelt, Jugend und Familie, Wien: 259 pp.
- WOLFINGER, R. D. (1993): Covariance structure selection in general mixed models. *Communications in Statistics – Simulation and Computation* 22: 1079-1106.

Thesen für Landwirte

Der heutige Entwicklungsstand der Agrar- und Ernährungswirtschaft ist auch Ergebnis einer mehr als hundertjährigen sorgfältigen Versuchstätigkeit. Dabei geht es schließlich immer um die Prüfung einer Produktionstauglichkeit von Forschungsergebnissen. So ist es verständlich, wenn Versuche unter Produktionsbedingungen das besondere Interesse der landwirtschaftlichen Praxis haben. Mit On-Farm-Experiment (OFE) wird ein solcher Versuchstyp bezeichnet. Hier werden Versuchsfragen unter Produktionsbedingungen mit betriebsüblicher Produktionstechnik bearbeitet, beispielsweise die Prüfung neuer Anbauverfahren, um Saatstärken, Saattermine, Bodenbearbeitung oder Pflanzenschutzmaßnahmen zu optimieren. Die Nähe zu den landwirtschaftlichen Produktionsbedingungen sichert grundsätzlich eine bessere Übertragbarkeit gegenüber Parzellenversuchen. Bei diesen Vorteilen darf aber nicht außer Acht gelassen werden, dass die Nähe zu den Produktionsbedingungen nur ein Aspekt ist. So sind Übertragbarkeit und Aussagekraft der Versuchsergebnisse nur dann zu garantieren, wenn auch bei OFE die Prinzipien zur Durchführung von Versuchen wie bei jedem anderen Versuchstyp auch eingehalten werden. Wegen der oft schwieriger zu gestaltenden Einheitlichkeit der Versuchsbedingungen gilt das bei OFE sogar in besonderem Maße. Auch OFE erfordern Aufwand und damit verbundene Kosten. Eine Rechtfertigung für diese Kosten ist nur gegeben, wenn durch Planung, Durchführung und Auswertung von OFE auch gesichert ist, dass aussagefähige Ergebnisse geliefert werden.

Folgende 10 Grundsätze fassen die aus Sicht des Landwirts wichtigsten Eckpunkte zusammen:

1. In OFE werden Forschungsergebnisse unter Praxisbedingungen bewertet.
2. OFE erfolgen unter Praxisbedingungen. Ihre Ergebnisse sind deshalb auf einen größeren Anwendungsbereich übertragbar als Ergebnisse aus Parzellenversuchen. Die Einheitlichkeit der Versuche ist bei OFE jedoch schwieriger zu sichern.
3. Die Ergebnisse von einem Schlag in einem Jahr sind nicht verallgemeinerungsfähig, da sie stark durch die Spezifik des Schlages und Jahres beeinflusst werden. Daher sind nur Ergebnisse mehrerer Schläge und Jahre verallgemeinerungsfähig. Die somit erforderlichen Versuchsserien bedingen eine Planung, die mehrere Betriebe und/oder Schläge berücksichtigen muss.
4. Für OFE gelten die selben Grundprinzipien zur Planung, Durchführung und Auswertung wie für jeden anderen Versuchstyp auch.
5. OFE bedingen spezifische Versuchsanlagen, die jedoch der jeweiligen Situation und Aufgabenstellung angepasst werden müssen.
6. Die statistischen Verfahren zur Auswertung erhobener Daten müssen bereits bei der Planung und Durchführung von OFE berücksichtigt werden.
7. Besondere Sorgfalt ist bei der Auswertung von OFE notwendig. Hier muss ein statistisches Auswertungsmodell gewählt werden, dass vor allem auch die spezifischen Störgrößen des jeweiligen Experiments berücksichtigt.
8. Bevor Messgeräte bei OFE zur Datenerfassung eingesetzt werden, sind Qualität und Einstellungen der Geräte von qualifiziertem Personal zu prüfen.
9. Bei Erfassung raumbezogener Daten muss ein einheitlicher Rasterbezug gesichert sein.
10. OFE verfügen über eine Spezifik, die im Landwirtschaftsbetrieb im Allgemeinen nicht beherrscht werden kann. An OFE interessierte Betriebe müssen daher von externen Experten unterstützt werden.

Thesen für Versuchsdurchführer

1. OFE stellen neben anderen Versuchstypen einen definierten Versuchstyp dar, bei dem es um die Bewertung von Forschungsergebnissen unter Praxisbedingungen wie beispielsweise von Precision Farming Technologien oder die Prüfung von Fragestellungen geht, die nur unter Praxisbedingungen untersucht werden können.
2. Der Vorteil von OFE gegenüber Parzellenfeldversuchen in Versuchsstationen besteht in der Durchführung unter Praxisbedingungen und einem damit verbundenen erweiterten Aussagebereich, der Nachteil in der schwereren Sicherung der Einheitlichkeit der Versuchsflächen und Gewährleistung des ceteris-paribus-Prinzips.
3. Ein OFE im Sinne eines Einzelversuchs hat ebenso wie ein einzelner Kleinparzellenversuch nur eine sehr beschränkte Aussagekraft. Eine verallgemeinerungsfähige Aussagekraft von OFE ist nur bei Planung und Auswertung im Sinne einer Versuchsserie möglich. Das stellt hohe Anforderungen an die Definition der Prüfglieder und deren Einhaltung im Experiment, insbesondere falls die Versuchsserie in verschiedenen Betrieben durchgeführt wird.
4. Die Grundprinzipien der Versuchsplanung (Randomisation, Wiederholung, Blockbildung) besitzen für OFE eine uneingeschränkte Gültigkeit. Die Möglichkeit einer geostatistischen Auswertung bei Vorliegen georeferenzierter Daten bedeutet nicht, dass diese Prinzipien außer Kraft gesetzt sind und damit ignoriert werden können.
5. OFE erfordern, wie jeder andere Versuchstyp auch, die Anwendung von Versuchsanlagen, wobei jedoch der spezifischen Aufgabenstellung angepasste Anlagen zu verwenden sind.
6. OFE müssen so geplant und durchgeführt werden, dass die zur Beantwortung der Versuchsfrage notwendigen statistischen Verfahren angewendet werden können (je nach Fragestellung stehen dabei Intervallschätzungen oder Signifikanzaussagen im Vordergrund).
7. Bei der Auswertung von OFE besteht die besondere Notwendigkeit einer statistischen Modellbildung und Überprüfung (Nutzung analytischer Kriterien zur Modellselektion, Residuenanalyse, Vergleich der bei Nutzung des Auswertungsmodells geschätzten Varianz und der Varianz der OLS-Residuen (ordinary least square)).
8. Eine qualifizierte Versuchsplanung, -durchführung und -auswertung erfordert im Allgemeinen eine externe Unterstützung der durchführenden Betriebe.
9. Die Sicherung einer hohen Datenqualität, insbesondere bei sensorgestützter Messwerterfassung, erfordert qualifiziertes Personal und besonderes Augenmerk auf die Kalibrierung der Sensoren.
10. Für georeferenzierte Daten, wie sie beispielsweise bei Fragestellungen im Precision-Farming anfallen, ist durch räumliche Interpolation ein einheitlicher Rasterbezug zu sichern. Das ist unverzichtbar, falls in unterschiedlichen Rastern erfasste Beobachtungen vorliegen (das kann sowohl für verschiedene Merkmale in einem Versuch als auch für das selbe oder verschiedene Merkmale bei einer Serienauswertung zutreffen).

* Die Thesen sollen möglichst knapp alle Forderungen und einzuhaltenden Kriterien nennen, die Ausformung der Thesen erfolgt im Leitfaden! Somit sind die Thesen stets im Zusammenhang mit dem Leitfaden zu nutzen.

Thesen für Entscheider

Die Qualität der Agrarforschung ist für die Zukunft der Agrar- und Ernährungswirtschaft von entscheidender Bedeutung. Wissenschaftliche Versuche nehmen dabei eine Schlüsselstellung im Erkenntnisprozess ein. Bei der Wahl des geeigneten Versuchstyps geht es immer um die Verbindung von möglichst praxisnahen Bedingungen bei gleichzeitiger Sicherstellung einer hohen Genauigkeit der Versuchsdurchführung. Der Parzellenversuch stellt hier einen akzeptierten Kompromiss dar, was seine hohe Verbreitung und Bedeutung sowohl heute als auch für die Zukunft begründet. Technische Entwicklungen ermöglichen aber stärker als in der Vergangenheit eine Versuchsdurchführung auch unter Produktionsbedingungen. Dieser als On-Farm-Experiment (OFE) bezeichnete Versuchstyp ermöglicht grundsätzlich auch Experimente auf praxisüblichen Ackerflächen mit einer ähnlichen Maschinenausstattung wie in der landwirtschaftlichen Primärproduktion durchzuführen. Der mit der Nähe zu praktischen Produktionsbedingungen verbundene Vorteil gegenüber Parzellenversuchen ist aber nur nutzbar, wenn man auch OFE als einen Versuchstyp begreift, für den die gleichen biometrischen Grundsätze gelten wie für jeden anderen Versuchstyp auch. Nur bei deren Einhaltung ist eine verallgemeinerungsfähige Interpretation der Ergebnisse und Rechtfertigung der auch mit OFE verbundenen Kosten zu erwarten.

In Thesen zusammengefasst bedeutet das:

1. OFE ist ein Versuchstyp, bei dem Forschungsergebnisse unter Praxisbedingungen bewertet werden. Damit rücken OFE zunehmend in das Interesse.
2. Da OFE unter Praxisbedingungen erfolgen, bieten OFE-Ergebnisse das Potenzial der Übertragbarkeit auf einen größeren Anwendungsbereich als die Ergebnisse aus Parzellenversuchen. Einheitlichkeit der Versuche ist bei OFE dagegen schwieriger zu sichern.
3. Ein OFE (ein Jahr, ein Schlag) bietet nur eine begrenzte Aussage - nur Ergebnisse aus einer Versuchsserie sind verallgemeinerungsfähig. Eine Versuchsserie ist sorgfältig über mehrere Betriebe und/oder Schläge zu planen.
4. Die Prinzipien zur Durchführung von Versuchen (Randomisation, Wiederholung, Blockbildung) gelten auch für OFE. Nur deren Einhaltung sichert Auswertbarkeit und Interpretierbarkeit der Ergebnisse.
5. Die Spezifik von OFE muss sich in den verwendeten Versuchsanlagen widerspiegeln, wobei zusätzlich die jeweilige Aufgabenstellung zu beachten ist.
6. Die anzuwendenden statistischen Schlüsse zur Auswertung der erhobenen Daten müssen bereits bei der Planung und Durchführung von OFE beachtet werden.
7. Wegen der oft schwieriger zu gewährleistenden Einheitlichkeit der Versuchsbedingungen und entsprechend zu beachtenden Störgrößen ist bei der Auswertung von OFE-Daten besondere Sorgfalt notwendig. Hier muss ein statistisches Modell gewählt und auf seine Eignung geprüft werden.
8. Die gerade bei OFE zu beachtende Komplexität bei Versuchsplanung, -durchführung, und -auswertung erfordert eine Unterstützung der durchführenden Betriebe durch externe Experten.
9. Die Sicherung einer hohen Qualität der Messwerte erfordert eine ständige Überprüfung der Messgeräte durch qualifiziertes Personal.
10. OFE sind nicht zwingend mit geringeren Kosten gegenüber anderen Versuchstypen verbunden. In jedem Fall werden jedoch die aufzuwendenden Kosten nur dann zu rechtfertigen sein, wenn die Prinzipien zur Planung, Durchführung und Auswertung von Versuchen auch bei OFE berücksichtigt werden. Nur dann kann von OFE der gewünschte Informationszuwachs geliefert werden.

Fehlerquellen bei der Erfassung von teilflächenspezifischen Ertragsdaten und Korrekturmöglichkeiten

HANS-GEORG SCHÖN

Zusammenfassung

Die Qualität von teilflächenspezifischen Ertragsdaten wird in starkem Maß von den verwendeten Ertragsmesssystemen und deren Wartung und Bedienung beeinflusst. Hier kann mit entsprechender Sorgfalt in einem gewissen Rahmen gegengesteuert werden.

Die Filterung der Rohdaten ermöglicht die Qualität im Nachgang der Generierung zu heben. Hierzu stehen diverse Möglichkeiten zu Verfügung. Filterung kann aber nur bedingt Prozessfehler beseitigen. Deshalb muss großer Wert auf durchgehende Sorgfältigkeit gelegt werden. Dies gilt in besonderem Maß für das Versuchswesen.

1 Einleitung

Die teilflächenspezifische Erfassung von Ertragsdaten ist eine technische Errungenschaft, die eine Vielzahl von Möglichkeiten eröffnet. Sie stellt eine wichtige Basis für die teilflächenspezifische Bewirtschaftung bereit, ermöglicht aber auch im Versuchswesen ganz neue Perspektiven. Der Einsatz von Ertragskartierungssystemen kann hier zu großer Kostenersparnis führen. Zudem wird die Durchführung von Praxisversuchen durch den Einsatz dieser Technologie wesentlich vereinfacht. In jedem Fall, insbesondere im letztgenannten, ist hohe Datenpräzision, Treffgenauigkeit und Zuverlässigkeit erforderlich, um zu belastbaren Aussagen kommen zu können. Die existierenden Ertragsfassungssysteme sind komplexe Gebilde. Im Prozess der Generierung der teilflächenspezifischen Erträge schlummern diverse Fehlerquellen. Diese können bis zu einem gewissen Grad minimiert werden. Teilweise besteht auch die Möglichkeit fehlerbehaftete Daten im Nachhinein zu identifizieren und zu filtern. Wichtige potentielle Fehlerquellen und Filtermöglichkeiten sollen im folgenden Text kurz beleuchtet werden.

2 Wichtige Fehlerquellen und Gegenmaßnahmen

Ertragskarten auf der Basis von Mähdescherrohdaten weisen in der Regel typische Muster auf, die auf Verzerrungen hindeuten. So werden z. B. Bereiche ohne bzw. mit unrealistisch hohem Ertrag ausgewiesen und innerhalb und zwischen Fahrspuren sind mehr oder minder abrupte Ertragssprünge zu verzeichnen. Auch Zwickel und Vorgewende sind regelmäßig deutlich zu erkennen (siehe Abbildung 1 und Abbildung 2).

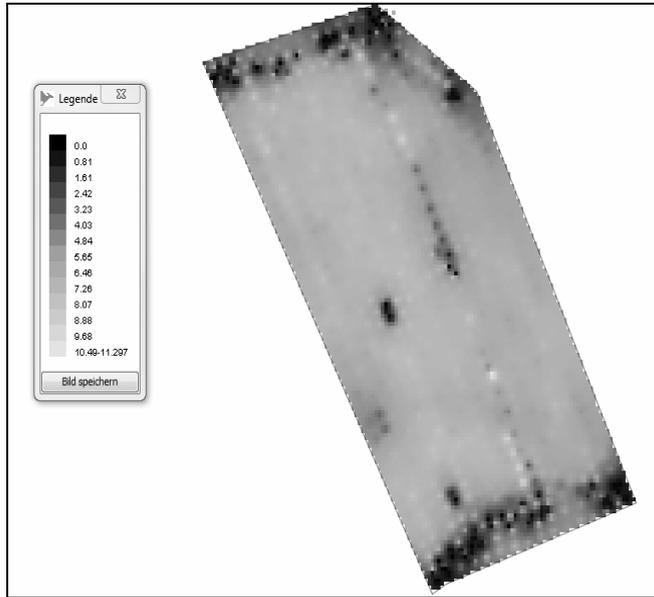


Abbildung 1: Ertragsinterpolation auf Basis von Rohdaten, Winterweizen in $t\ ha^{-1}$, Feldschlag Rusterberg 2002

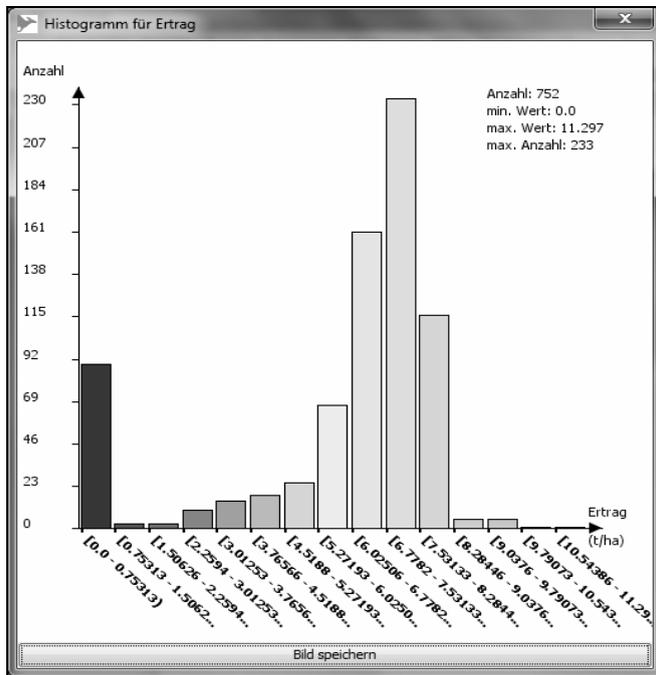


Abbildung 2: Häufigkeitsverteilung der unkorrigierten Ertragswerte, Feldschlag Rusterberg 2002

Es gibt eine Vielzahl von Fehlerquellen, die Mähdrescherrohdaten verzerren können.

Hierzu zählen unter anderem:

- unvollständige Kartierung,
- Positionierungsfehler,
- mangelhafte Kalibrierung,
- Ungenauigkeit des Kornsensors,
- fehlerhafte Wegmessung,
- Schnittbreite,
- Zeitversatz,
- Füllphasenfehler nach Start,
- Entleerungsphasenfehler nach Stopp,
- Fehler infolge von Geschwindigkeitsänderungen.

Positionierungsfehler können folgende Ursachen haben: Ausfall des Korrektursignals, Empfang von zu wenig Satelliten, schlechte Position der Satelliten, zu schwache Signale, Abschattung (Bäume, Bunkerklappen!) oder systematischer Fehler durch exzentrische Antennenposition.

Unzureichende Kalibrierung führt ebenfalls zu Fehlern. Volumetrische Sensorsysteme sind besonders anspruchsvoll. Das Hektolitergewicht muss korrekt ermittelt werden. Generell sollten Eichungen mehrfach am Tag durchgeführt werden.

Die Ertragsmesssysteme an sich sind mehr oder minder fehlerbehaftet (NOACK, 2005). Verzerrungen können zudem durch Neigungswinkel entstehen.

Auch die Wegmessung ist nicht fehlerfrei. Erfolgt die Wegmessung GPS-basiert, so können hierbei Positionsfehler entstehen. Bei Wegmessung über einen Radsensor kann Schlupf verzerren. Die Ertragsermittlung erfolgt über folgende Formel:

$$\text{Ertrag} = \frac{\text{Korndurchflussrate} * \text{Umrechnungsfaktor}}{\text{Schnittbreite} * \text{Strecke} / \Delta \text{Zeit}}$$

Eine fehlerbehaftete Streckenermittlung wirkt sich hierüber direkt auf die ausgewiesenen Erträge aus.

Dieser Formel ist ebenfalls zu entnehmen, dass die korrekte Schnittbreite natürlich auch bedeutend für eine unverzerrte Ertragsermittlung ist (siehe auch Abbildung 3).

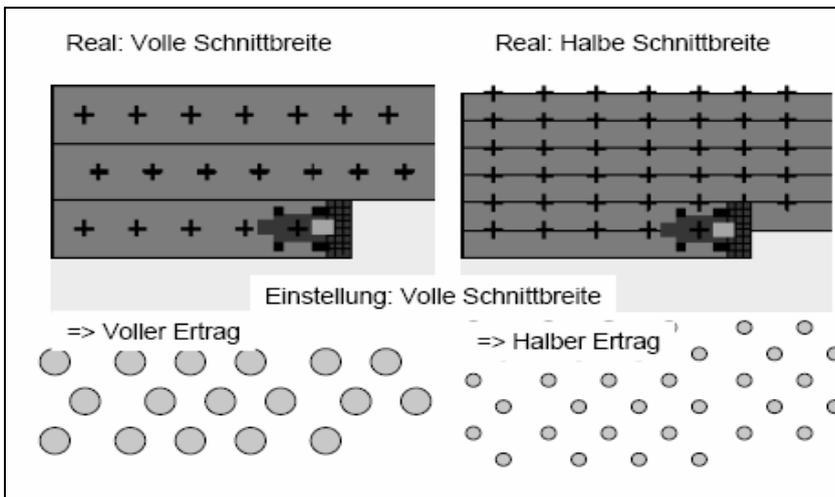


Abbildung 3: Auswirkung fehlerhafter Schnittbreite auf den Ertrag (BLACKMOORE, 2007; GEBBERS, 2008)

Zeitversatz ist auch eine Ursache für fehlerhafte Ertragsermittlung (Abbildung 4). Dies betrifft sowohl die unterschiedlichen Transportstrecken des Erntegutes als auch den zeitlichen Versatz zwischen Positions- und Durchflussmessung.

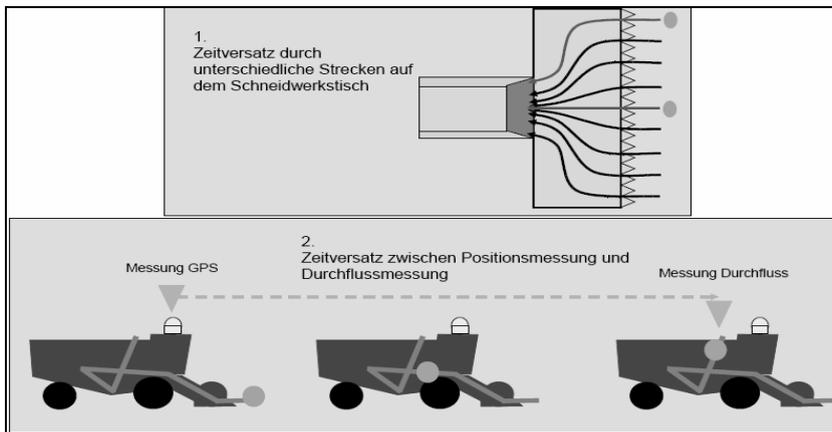


Abbildung 4: Auswirkung des Zeitversatzes auf die Ertragsermittlung (GEBBERS, 2008)

In den Rohdaten sind auch deutlich Füllphasenfehler zu beobachten. Dies ist in der Tatsache begründet, dass eine gewisse Druschstrecke benötigt wird, bis das System den vollen Durchsatz erlangt hat (Abbildung 5). Im Gegenzug führt der Druschvorgang bei der Ausfahrt aus dem Bestand zum Entleerungsphasenfehler.

Verminderungen der Geschwindigkeit bei zunächst gleichbleibender Durchflussrate erhöhen den Ertrag. Dies ist in der Trägheit des Druschsystems begründet.

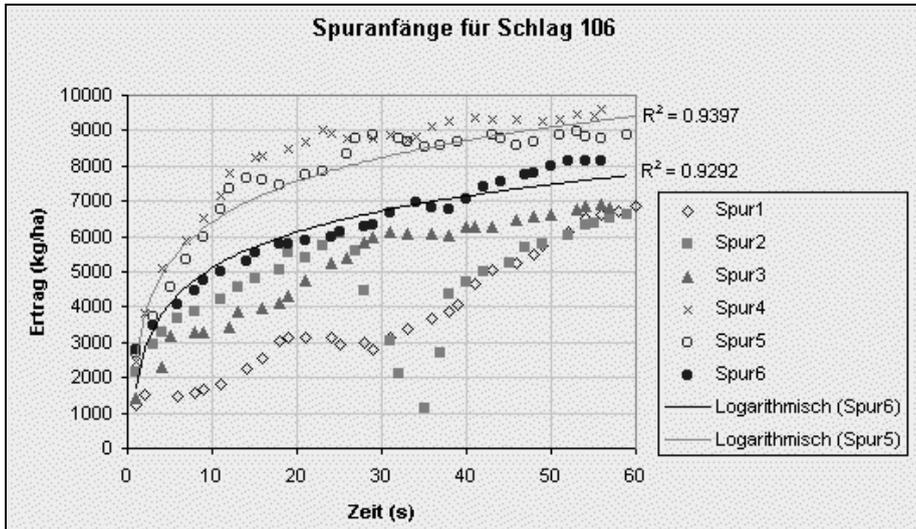


Abbildung 5: Füllphasenfehler unterschiedlicher Druschspuren (EBERT, 1999)

Abgeleitet aus den diversen Fehlerquellen ergibt sich folgender Maßnahmenkatalog:

- Wartung der Mähdrescher, Schulung der Fahrer, Speichermedien prüfen, einheitlich und eindeutig dokumentieren,
- mindestens einmal pro Schlag kalibrieren,
- gleichmäßig fahren,
- Schnittbreiteneinstellung konstant halten,
- Schneidwerk auf abgerenteten Flächen anheben,
- Lenkhilfen nutzen,
- zeitnahe Auswertung der Daten,
- Speicherung der Rohdaten.

Im Versuchswesen ist besondere Sorgfalt geboten. Hier ist zu überlegen, ob Parzellenerträge nicht zur Kontrolle separat mit einer Radlastwaage gegengewogen werden können.

3 Datenbereinigung

Datenbereinigung und –aggregation werden bereits im Mähdrescher vorgenommen. Die Vorgehensweise ist nicht immer offengelegt. Bedingt durch die im vorstehenden Text angesprochenen Fehlerquellen müssen Rohdaten demzufolge mehr oder minder stark bereinigt und gefiltert werden. Hierzu werden Globalfilter und Lokalfilter genutzt. Beim globalen Filtern wird die Verteilung der Rohdaten genutzt, um Unter- und Obergrenzen abzuleiten, beispielsweise Mittelwert ± 2 s. Letzteres setzt Normalverteilung voraus, die nicht immer gegeben ist. Bei großer Streuung der Werte führt überdies ein solches globales Vorgehen zum Verbleib von Ertragswerten, die immer noch ungewöhnlich klein oder groß sind und bei denen das Gefühl verbleibt, dass Verzerrungen vorliegen.

Lokale Filter vergleichen Messwerte mit den Werten in ihrer näheren Umgebung. Es unterstellt die Regel, dass nah beieinander liegende Werte ähnlicher sind als weiter entfernt liegende. Hierzu wurden diverse Methoden entwickelt. Eine ist die sogenannte H-Methode von NOACK (2005).

Das Ertragsfilterungsprogramm Expertfilter (CIS GmbH) nutzt als weitere Möglichkeit zur Datenfilterung eine Vielzahl von Informationen (Geschwindigkeit, Messpunktabstand in und zwischen der Reihe usw.), um die Daten in Segmente zu unterteilen. Messpunkte am Anfang und Ende der Segmente werden verworfen (Füll- und Entleerungsphasenfehler). Segmente mit weiten oder sehr engen Punktabständen werden entfernt. Viele der Filterkriterien können vom Benutzer eingestellt werden, was wiederum gewissen Spielraum für Datenmanipulation lässt.

4 Abschließende Bemerkungen

Der Einsatz genauer GPS-Systeme unter Verwendung präziser Korrektursignale ermöglicht heutzutage eine Georeferenzierung, deren Fehlerbehaftung vernachlässigbar ist. Die zu Verfügung stehenden Ertragsmesssysteme führen nicht zu deckungsgleichen Ertragswerten. Sie erfordern zudem eine intensive Wartung und sachkundige Bedienung. Anderenfalls kann die Qualität von teilflächenspezifischen Ertragsdaten in bedeutendem Umfang beeinträchtigt werden. Schulung des Personals und regelmäßige Kalibrierung wirken dem entgegen. Ziel muss es sein, die Rohdatenqualität zu erhöhen.

Durch Filterung der Rohdaten kann die Qualität im Nachgang der Generierung verbessert werden. Hierzu stehen diverse Methoden und Programme zu Verfügung. Die Filterung kann in beschränktem Maß die Fehlerbehaftung reduzieren. Systematische Fehler können allerdings durch die vorhandenen Filter nicht entdeckt werden. Deshalb muss durchgehend mit größtmöglicher Sorgfalt gearbeitet werden. Dies gilt verständlicherweise in besonderem Maß für das Versuchswesen.

Arbeitsschritte und wichtige Unterpunkte der Ertragsdatengenerierung und -verarbeitung sind nachfolgend zusammengestellt:

Ernte:

- Regelmäßige Wartung der Ertragsmesssysteme (einschließlich Feuchtemessung)
- Fahrer schulen
- Genaues Korrektursignal beziehen
- Regelmäßige Kalibrierung
- Konstante Geschwindigkeit beim Dreschen
- Wenn der Einsatz mehrerer Mähdrescher nicht umgangen werden kann, möglichst baugleiche Systeme einsetzen
- Gleichbleibende Schnittbreite durch präzise Lenkhilfe
- Schneidwerk auf abgeernteten Flächen anheben



Datenhaltung:

- Rohdaten mit Metainformationen (Fahrername, Mähdrescher-ID, Kommentare, Erntebedingungen, besondere Vorkommnisse, Feuchtekorrektur) in einem DBMS speichern
- Datensätze auf Vollständigkeit prüfen und gegebenenfalls kennzeichnen



Filterung der Ertragsrohdaten:

- Verzerrte Ertragswerte durch geeignete Filter identifizieren und kennzeichnen
- Plausible Daten in neue Datei überführen (Filtrat)



Datenauswertung/Interpolation:

- Filtrat nutzen
- Eventuell Transformation der Koordinaten
- Statistische/geostatistische Analyse
- Interpolation
- Kartierung

Literatur

BLACKMORE, S. (2007): Instrumentation for Precision Agriculture. Short Course. 6 European Conference on Precision Agriculture (6. ECPA). Skiatohos, Greece.

EBERT, C. (1999): Ertragskartierung in Mecklenburg-Vorpommern. Stand, Probleme sowie Korrektur- und Auswertungsmöglichkeiten von Ertragskarten. Rostock: Universität Rostock, Fachbereich Agrarökologie, unveröff. Diplomarbeit.

GEBBERS, R. (2008): Ertragskartierung. Vortrag an der FH Osnabrück am 2. Juni 2008.

NOACK, P. O. (2005): Entwicklung fahrspurbasierter Algorithmen zur Korrektur von Ertragsdaten im Precision Farming, Dissertation, Technische Universität München.

Zu den Grundsätzen der Versuchsplanung (insbesondere Randomisation)

HANS-PETER PIEPHO, JENS MÖHRING und EMLYN R. WILLIAMS

Zusammenfassung

Die drei elementaren Grundsätze der Versuchsplanung sind: Wiederholung, Randomisation und Blockbildung. Diese Grundsätze gelten insbesondere auch bei On-Farm-Experimenten (OFE). Dieser Beitrag rekapituliert zunächst kurz diese drei Grundsätze und illustriert dann im Detail die Bedeutung der Randomisation mit Hilfe von zwei hypothetischen Datensätzen. Ein Datensatz weist einen deutlichen systematischen Trend auf, während der andere ein eher erraticus Muster zeigt. Die Schlussfolgerungen aus diesen beiden Beispielen werden unterstützt durch Analysen eines Blindversuches sowie durch eine kleine Simulationsstudie. Die Ergebnisse zeigen, dass sowohl eine Modell-basierte geostatistische Analyse als auch eine Randomisations-basierte Analyse unter Annahme unabhängiger Fehler valide sind, sofern randomisiert wurde. Dagegen werden beide ungültig, wenn nicht randomisiert wird. Es ergibt sich hieraus die Schlussfolgerung, dass eine Randomisation eine Absicherung gegen verschiedene Arten von räumlichem Trend bietet. Daher sollten landwirtschaftliche Versuche nach Möglichkeit immer randomisiert werden.

1 Einleitung

Jeder landwirtschaftliche Versuch sollte die drei Grundsätze der Versuchsplanung berücksichtigen: (1) Wiederholung, (2) Randomisation und (3) Blockbildung.

(1) *Wiederholungen*: Ohne Wiederholungen kann die Fehlervarianz nicht geschätzt werden. Somit besteht ohne Wiederholungen keine Basis zur Durchführung von Tests (Varianzanalyse, t-Tests) oder der Berechnung von Vertrauensintervallen.

(2) *Randomisation*: Um Verzerrungen zu vermeiden und eine „gerechte“ Verteilung der Behandlungen auf die Versuchseinheiten (Parzellen) zu gewährleisten ist es erforderlich, dass diese Verteilung nach dem Zufallsprinzip erfolgt.

(3) *Blockbildung*: Während (1) und (2) Sollbestimmungen sind, ist die Blockbildung eine Kann-Bestimmung. Durch Blockbildung, also die Gruppierung der Parzellen zu relativ homogenen Einheiten (Blöcken), kann die Präzision eines Versuches erheblich erhöht werden.

Diese Grundsätze gelten auch in OFE, insbesondere auch dann, wenn man pro Parzelle mehrere Messwerte erhalten kann. Vordergründig mögen diese Messwiederholungen als echte Wiederholungen erscheinen. Da diese Messungen aber jeweils auf derselben Randomisationseinheit (Parzelle) erfolgen, handelt es sich lediglich um Pseudowiederholungen. Eine elementare Möglichkeit der Auswertung besteht darin, die Messungen je Randomisationseinheit zu mitteln. Eine verfeinerte Auswertung modelliert die Einzelmessungen je Parzelle und ermöglicht es so, den Einfluss von Kovariablen zu erfassen, die wiederholt pro Parzelle gemessen wurden (EC38, etc.). Da diese Messungen innerhalb einer Parzelle aber nicht der Randomisation unterliegen, muss die Auswertung die räumliche Korrelation berücksichtigen. Sofern die Messungen georeferenziert sind, können hierzu geostatistische Methoden verwendet werden. Falls

keine Georeferenzierung besteht, kann man die Autokorrelation durch Aufnahme eines zufälligen Parzelleneffektes in das Modell berücksichtigen (PIEPHO et al., 2011). Die Auswertung von Parzellenmittelwerten stellt aber den Anfangspunkt einer Auswertung dar und sollte in jedem Fall erfolgen. Ist eine solche Basisauswertung nicht möglich, so ist dies ein untrügliches Zeichen dafür, dass der Versuch nicht richtig randomisiert wurde. Insofern stellt die Basisauswertung einen praktischen Lackmustest dar für die Frage, ob der Versuch richtig randomisiert wurde.

Bei der geostatistischen Modellierung von Einzelmessungen je Parzelle bestehen grundsätzlich drei Möglichkeiten, abhängig davon, mit welcher „Reichweite“ die räumliche Korrelation modelliert wird: (i) Nur die Korrelation von Messpunkten innerhalb derselben Parzelle. (ii) Nur die Korrelation von Messpunkten innerhalb desselben Blocks. (iii) Die Korrelation von allen Messpunkten. Die Modellierung der Korrelation über die Parzelle (Randomisationseinheit) hinaus erhöht den Aufwand der Auswertung (Modellselektion etc.). Designeffekte für Randomisationseinheiten (Parzellen, Blöcke, etc.) sollten auf jeden Fall im Modell bleiben. Auf Details zu diesen Auswertungsmöglichkeiten gehen die Beiträge von J. SPILKE und C. RICHTER in diesem Band ein.

Im Folgenden wird detailliert auf die Bedeutung der Randomisation eingegangen. Hierbei wird insbesondere herausgearbeitet, dass eine geostatistische Auswertung eine fehlende Randomisation keinesfalls ersetzen kann. Der Einfachheit halber beschränkt sich die Betrachtung auf vollständig randomisierte Versuche mit einem Analysewert pro Parzelle. Die generellen Schlussfolgerungen gelten aber auch für geblockte Versuche und für Versuche mit mehreren Messwerten pro Parzelle.

2 Die Bedeutung der Randomisation

Der Zweck eines wissenschaftlichen Experimentes besteht gewöhnlich darin, eine Verallgemeinerung auf ähnliche Situationen in der Zukunft zu erlauben, und nicht nur darin, irgendwelche numerischen Versuchsergebnisse zu erzielen. Im Hinblick auf diese Zielstellung sollte ein einzelnes Experiment lediglich als seine einzelne Stichprobe aus einer potentiell sehr großen Grundgesamtheit von ähnlichen Experimenten aufgefasst werden, die man hätte durchführen können oder die durchgeführt werden könnten. Statistische Auswertungsverfahren erlauben valide Schlussfolgerungen, die über das konkrete Experiment hinaus verallgemeinert werden können, vorausgesetzt das Experiment erfüllt einige elementare statistische Bedingungen. In geplanten Versuchen besteht die wesentliche Voraussetzung in einer adäquaten Randomisation.

Obwohl die Bedeutung der Randomisation als elementares Prinzip des experimentellen wissenschaftlichen Arbeitens seit der Arbeit von FISHER (1925) bekannt ist, scheint seine Bedeutung nicht immer angemessen berücksichtigt zu werden. Nach unseren Erfahrungen aus der statistischen Beratung und der Begutachtung von Manuskripten werden nach wie vor zahlreiche Experimente gar nicht oder nicht richtig randomisiert. Dieser Beitrag zeigt anhand mehrerer Beispiele, dass die Randomisation eine zentrale Voraussetzung zur Erzielung valider statistischer Schlüsse ist. Um das Problem zu illustrieren und wichtige Grundideen zu erläutern, beginnen wir mit einem kleinen hypothetischen Sortenversuch mit fünf Sorten und 20 Parzellen. Danach präsentieren wir Ergebnisse basierend auf einem Blindversuch, mit welchem verschiedene Versuchsanlagen untersucht werden, sowie eine Simulationsstudie.

3 Drei Beispiele

3.1 Ein kleiner hypothetischer Sortenversuch

Um die Wichtigkeit der Randomisation zu illustrieren, nehmen wir an, dass ein Sortenversuch auf 20 Parzellen durchgeführt werden soll, die in einem Gitter aus vier Zeilen und fünf Spalten angeordnet sind. Des Weiteren sei angenommen, dass fünf verschiedene Weizensorten (A bis E) auf jeweils vier Parzellen geprüft werden sollen. Durch ein Versehen enthalten jedoch alle Saatgutsäcke nur Saatgut der Sorte A, obwohl die Säcke mit den Sortennamen A bis E bezeichnet sind. Was wir also für fünf verschiedene Sorten halten, ist in Wirklichkeit ein und dieselbe Sorte, d. h. Sorte A wird auf allen 20 Parzellen geprüft. Abbildung 1 zeigt ein hypothetisches Ergebnis eines solchen Versuches. Die Daten zeigen einen systematischen räumlichen Trend, wobei die Erträge von unten links nach oben rechts absinken. Dieser Trend könnte beispielsweise durch Unterschiede in der Bodenart und des Wasserhaltevermögens verursacht sein, welche sich auf den Ertrag auswirken. Die Parzelle im unteren linken Teil des Plans könnte also deswegen einen erhöhten Ertrag aufweisen, weil dort das Wasserhaltevermögen am besten ist.

7.6	7.4	7.2	7.0
8.1	7.6	7.3	7.0
8.3	8.1	7.9	7.3
8.8	8.5	8.0	7.7
9.0	8.6	8.4	8.0

Abbildung 1: Erträge (in $t\ ha^{-1}$) für einen hypothetischen Versuch mit 20 Parzellen, angeordnet in einem Gitter aus fünf Zeilen und vier Spalten, in dem fünf Sorten mit den Bezeichnungen A bis E geprüft werden

Dieses hypothetische Experiment entspricht einem Blindversuch. Solche Versuche, in welchen auf jeder Parzelle dieselbe Behandlung geprüft wird, werden durchgeführt, um alternative Versuchsanlagen und Parzellengrößen zu vergleichen (SMITH, 1938; WILLIAMS and LUCKETT, 1988). Betrachten wir nun die Zuordnung der Sortenbezeichnungen zu den Parzellen in einem randomisierten Experiment. Die einfachste Versuchsanlage ist die vollständig randomisierte Anlage, bei der die Sortenbezeichnungen den Parzellen völlig zufällig zugeordnet werden. Im vorliegenden Beitrag konzentrieren wir uns der Einfachheit halber ausschließlich auf die vollständig randomisierte Anlage. Es sei aber betont, dass gute Versuchsanlagen zur effizienten Fehlerkontrolle in der Regel eine Blockbildung umfassen (vollständige oder unvollständige Blöcke, ein-dimensionale Blockbildung oder Zeilen-Spalten-Pläne) (JOHN and WILLIAMS, 1995). Obschon wir uns auf die vollständig randomisierte Anlage konzentrieren, gelten unsere Aussagen hinsichtlich der Vorzüge der Randomisation in gleichem Maße auch für jede Art von geblockter Versuchsanlage.

Eine mögliche Randomisation für eine vollständige Anlage wird in Abbildung 2(a) gezeigt. Es sei nochmals darauf hingewiesen, dass wir trotz der Zuordnung verschiedener Sortenbezeichnungen annehmen, dass in Wahrheit auf allen Parzellen die Sorte A geprüft wurde. Eine klassische Varianzanalyse, welche einen F -Test der globalen Nullhypothese der Abwesenheit von Behandlungseffekten liefert, fußt auf einer zufallsmäßigen Zuordnung der Behandlungen zu den Parzellen wie in Abbildung 2(a) veranschaulicht. Insbesondere nimmt die Varianzanalyse an, dass jede der möglichen Zuordnungen die gleiche Wahrscheinlichkeit hat. Mit dieser An-

nahme kann die Verteilung der F -Statistik unter der Nullhypothese, kurz “die Nullverteilung”, entweder analytisch oder mit Hilfe der Randomisationsverteilung hergeleitet werden (PITMAN, 1938; GOOD, 2000). Wir werden in Kürze die Randomisationsverteilung etwas näher betrachten und später zeigen, dass beide Nullverteilungen im Wesentlichen identisch sind.

(a) Randomisierter Plan				(b) Systematischer Plan			
D	B	D	E	A	A	A	A
C	E	A	B	B	B	B	B
A	C	E	A	C	C	C	C
D	D	A	B	D	D	D	D
B	C	E	C	E	E	E	E

Abbildung 2: Zwei mögliche Zuordnungen der Behandlungsbezeichnungen zu den Parzellen für den Feldplan in Abbildung 1

(a) Eine mögliche randomisierte Zuordnung; (b) Systematische Zuordnung

Zum Vergleich betrachten wir nun eine systematische Zuordnung der Sorten zu den Parzellen wie in Abbildung 2(b) gezeigt. Systematische Versuchsanlagen wie diese werden manchmal zum Beispiel in Versuchen zum Präzisionspflanzenbau verwendet (PIEPHO et al., 2011). Man beachte, dass diese Anlage auch als vollständige Blockanlage mit Blöcken in Spalten sowie einer systematischen Anordnung der Sorten in den Spalten betrachtet werden kann. Prinzipiell könnte eine solche Anordnung rein zufällig bei einer vollständigen Randomisation auftreten, aber die Wahrscheinlichkeit dafür ist außerordentlich gering. Wenn diese systematische Anordnung verwendet wird, dann hat die F -Statistik der Varianzanalyse den Wert $F = 5.63$. Aber in Wirklichkeit bestehen keine Unterschiede zwischen den Behandlungen (jede Behandlung ist dieselbe Sorte A), so dass in diesem Fall der große F -Wert ausschließlich durch den ausgeprägten systematischen räumlichen Trend (siehe Abbildung 1) entlang der Spalten verursacht wird. Die daraus resultierende Verzerrung wird auch deutlich, wenn die Behandlungsmittelwerte für die systematische Anlage mit denen der randomisierten Anlage verglichen werden (Tabelle 1).

Tabelle 1: Mittelwerte (in $t\ ha^{-1}$) für Daten aus Abbildung (systematischer räumlicher Trend) basierend auf der Randomisation in Abbildung(a) und auf dem systematischen Versuchsplan in Abbildung(b)

Sorten	Randomisiert (Abb. 2a)	Systematisch (Abb. 2b)
A	7.73	7.30
B	7.78	7.50
C	8.20	7.90
D	8.03	8.25
E	7.73	8.50
Spannbreite	0.47	1.20
F -Wert	0.46	5.63

Die in Abbildung 1 gezeigten Daten weisen einen starken systematischen Trend auf. Um zu zeigen, dass eine Randomisation auch dann sinnvoll ist, wenn der Trend eher erratischen Charakter hat, betrachten wir die Daten in Abbildung 3. Die Erträge in Abbildung 3 sind genau dieselben wie in Abbildung 1, aber ihre Positionen sind gegenüber denen in Abbildung 1 ver-

ändert. Die Daten wurden aus Abbildung 1 erzeugt, indem zuerst die Zeilen permutiert wurden und dann die Parzellen innerhalb der Zeilen, so dass die Zusammensetzung der Zeilen unverändert ist. Dies hat zur Folge, dass der F -Wert für die systematische Anlage in Abbildung 2(b) $F = 5.63$ beträgt, was identisch ist zum F -Wert für die Daten in Abbildung 1, ebenso wie die Sortenmittelwerte in Tabelle 2.

Tabelle 2: Mittelwerte (in $t\ ha^{-1}$) für Daten aus Abbildung 3 (erratischer räumlicher Trend) basierend auf der Randomisation in Abbildung 2(a) und auf dem systematischen Versuchsplan in Abbildung 2(b)

Behandlung	Randomisiert (Abb. 2a)	Systematisch (Abb. 2b)
A	7.85	8.25
B	7.85	7.30
C	7.58	7.90
D	8.45	8.50
E	7.73	7.50
Spannbreite	0.87	1.20
F -Wert	1.36	5.63

Für die randomisierte Anlage in Abbildung 2(a) beträgt der F -Wert nur $F = 1.36$. Die beiden Beispiele zeigen, dass eine unterlassene Randomisation ein Problem sein kann, sowohl mit als auch ohne klar erkennbaren systematischen räumlichen Trend.

8.8	7.7	8.0	8.5
7.6	7.2	7.4	7.0
7.3	8.1	7.9	8.3
8.0	9.0	8.4	8.6
8.1	7.6	7.3	7.0

Abbildung 3: Erträge (in $t\ ha^{-1}$) für einen hypothetischen Versuch mit 20 Parzellen, angeordnet in einem Gitter aus fünf Zeilen und vier Spalten, in dem fünf Sorten mit den Bezeichnungen A bis E geprüft werden

Um einen näheren Einblick zu erhalten ist es hilfreich, die Nullverteilung der F -Statistik für die randomisierte Anlage etwas näher zu betrachten. Der Bezug zu einer Nullverteilung ist deswegen erforderlich, weil die in einem konkreten Versuch verwendete Zuordnung der Behandlungen zu den Parzellen nur eine mögliche Zuordnung einer in der Regel sehr großen Zahl möglicher Zuordnungen ist, so dass all diese möglichen Zuordnungen bei der Beurteilung eines einzelnen F -wertes betrachtet werden müssen. Wie oben erwähnt, wird als Nullverteilung des F -Tests meist die analytische F -Verteilung mit geeigneten Freiheitsgraden für Zähler und Nenner verwendet (MEAD et al., 2002). Diese Nullverteilung nimmt an, dass beim verwendeten Randomisationsschema jede der möglichen Zuordnungen gleich wahrscheinlich ist. Alternativ kann die Nullverteilung auch empirisch durch Randomisation wie folgt erhalten werden: Generiere eine große Anzahl (wenn möglich alle Möglichkeiten) von randomisierten Zuordnungen der Sortenbezeichnungen zu den Parzellen und berechne für jede Zuordnung den Wert der F -Statistik. Dies liefert eine empirische Version der Nullverteilung der F -Statistik (PITMAN, 1938; GOOD, 2000). Der Grundgedanke dieser Randomisationsverteilung ist wie folgt: Wenn es keine Behandlungsunterschiede gibt, wenn also jede Behandlung denselben Erwartungswert hat, dann sollte es ohne Folge für den erwarteten Ausgang des Experimentes

sein, wenn wir die Zuordnung der Behandlungen zu den Parzellen durcheinanderwürfeln. Dies bedeutet, dass wir beim Test der Nullhypothese den für eine bestimmte Randomisation beobachteten F -Wert als einen Vertreter einer Vielzahl von möglichen F -Werten betrachten, der bei Gültigkeit der Nullhypothese zu erwarten ist. Abbildung 4 zeigt die Randomisationsverteilung von 100.000 Randomisationen der Daten in Abbildung 1. Man beachte, dass diese Randomisationsverteilung identisch ist mit derjenigen, die für die Daten in Abbildung 3 gilt, ebenso wie der beobachtete F -Wert für das systematische Design (Abbildung 2b), so dass die folgenden Überlegungen in gleichem Maße für beide Beispiele gelten (mit oder ohne systematischen Trend). Es dürfte klar sein, dass die Randomisationsverteilung in Abbildung 4 tatsächlich eine Nullverteilung ist, weil wir hier annehmen, dass aus Versehen auf allen Parzellen dieselbe Sorte A geprüft wurde. Es sei allerdings an dieser Stelle betont, dass der Randomisationsansatz auch dann gültig wäre, wenn die Sorten A bis E nicht durcheinander gebracht worden wären, weil unter der Nullhypothese die Behandlungsbezeichnungen immer austauschbar sind (GOOD, 2000).

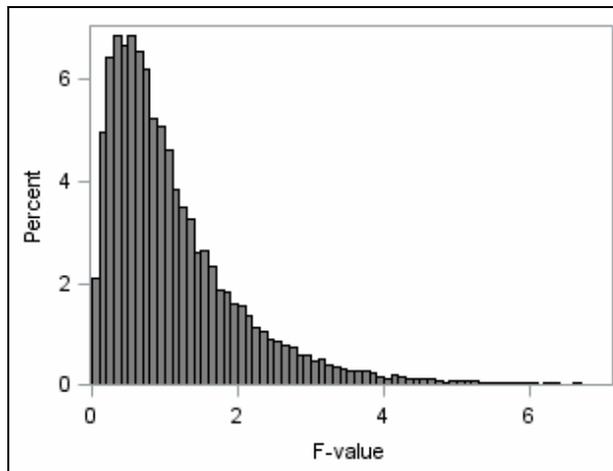


Abbildung 4: Randomisationsverteilung der F -Statistik für eine klassische Varianzanalyse mit den Daten aus Abbildung 1 (identisch mit den Daten aus Abbildung 3) (Anzahl Randomisationen = 10^5)

Die Grundidee eines Signifikanztests besteht darin, die Nullhypothese immer dann zu verwerfen, wenn die Teststatistik einen für die Nullverteilung untypischen Wert annimmt. Um zu entscheiden, wie typisch oder untypisch ein beobachteter F -Wert ist, können wir den Anteil der Fälle berechnen, in denen der F -Wert der Randomisationsverteilung den F -Wert des Versuchs überschreitet. Dieser Anteil ist als der P -Wert bekannt. Für die systematische Anlage in Abbildung(b) ist der Randomisations- P -Wert gleich $P = 0.0071$, was anzeigt, dass der beobachtete Wert $F = 5.63$ ein sehr untypischer Wert im Vergleich zu der Verteilung des F -Wertes ist, die unter der Nullhypothese zu erwarten ist (Abbildung 4): Nur für 0.71 % der F -Werte ist zu erwarten, dass sie so groß sind wie der für die systematische Anlage beobachtete Wert, oder größer. Wenn der beobachtete F -Wert untypisch unter Annahme der Nullverteilung ist, dann ist dies Anlass diese Nullhypothese zu verwerfen. Das übliche Vorgehen lehnt die Nullhypothese ab, wenn der P -Wert kleiner als $\alpha = 0.05$ ist. Der Randomisations- P -Wert ist sehr nahe bei demjenigen P -Wert, den man bei Verwendung einer analytischen F -Verteilung mit 4 Zähler- und 15 Nennerfreiheitsgraden erhält ($P = 0.0057$), die analytische Nullverteilung für dieses Beispiel. Im Prinzip könnten wir F -Tests immer mittels der Rando-

misationsverteilung durchführen, aber es ist weniger rechenzeitaufwändig, das analytische Ergebnis zu verwenden; die Ergebnisse in beiden Fällen werden in der Regel fast identisch sein.

In den beiden Beispielen führt die systematische Versuchsanlage zu einem sehr großen F -Wert, einem F -Wert, der als untypisch im Vergleich zur Randomisationsverteilung in Abbildung 4 zu werten ist. Daher würde die Nullhypothese für diese unrandomisierte Anlage fälschlicherweise verworfen, was die Problematik einer unterlassenen Randomisation unterstreicht. Der entscheidende Punkt, der hier nochmals hervorgehoben wird, ist, dass die eben betrachtete Nullverteilung nur dann die angemessene Referenzverteilung für den F -Test ist, wenn der Versuch randomisiert wurde. Denn nur dann ist jede Randomisation gleich wahrscheinlich und nur dann kann die Nullverteilung aus einem klassischen Randomisationsargument heraus hergeleitet werden (CALINSKI and KAGEYAMA, 2003). Man kann hier sogar argumentieren, dass die Randomisationsverteilung in Abbildung 4 nicht die adäquate Referenzverteilung für den systematischen Versuchsplan ist, weil dieser nicht als ein zufällig ausgewählter Repräsentant aller möglichen vollständigen Randomisationen gelten kann. Wenn der Versuchsansteller sich für eine systematische Versuchsanlage entscheidet mit Behandlungen in Zeilen mit jeweils vier Parzellen wie in Abbildung 2(b), dann besteht die einzig verbleibende Wahlmöglichkeit in einer Permutation der Behandlungen zwischen den Zeilen. Aber jede solche Permutation liefert exakt denselben F -Wert, so dass die Referenzverteilung auf einen einzigen möglichen F -Wert zusammenfällt! Diese Überlegungen zeigen, dass mit einem systematischen Versuchsplan kein sinnvoller Test auf Basis eines Randomisationsargumentes möglich ist.

Es ist auch hilfreich, die Verteilung der analytischen P -Werte der für die Randomisationsverteilung berechneten F -Statistiken zu betrachten. Es ist eine wichtige Tatsache, dass für den F -Test (wie für jeden anderen Signifikanztest) die Randomisationsverteilung der P -Werte einer Gleichverteilung folgen muss, damit der Test gültig ist. Daher bietet die Betrachtung der Randomisationsverteilung der analytischen P -Werte eine besonders anschauliche Möglichkeit, die Gültigkeit eines Signifikanztests zu beurteilen. Für das Beispiel sind die P -Werte schön gleichmäßig verteilt (Abbildung 5), so dass der Test Gültigkeit besitzt.

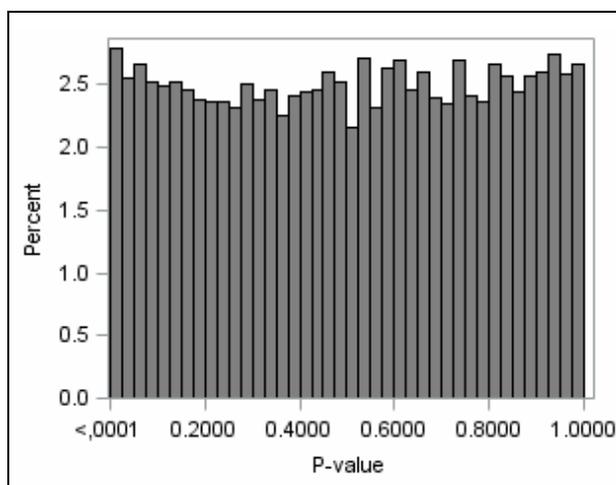


Abbildung 5: Gleichverteilung der P -Werte unter der Randomisationsverteilung basierend auf den Daten in Abbildung (Anzahl Randomisationen = 10^5)

Des Weiteren sind die geschätzten Behandlungsdifferenzen unverzerrt im Vergleich zur Randomisationsverteilung, wie für den Fall der Mittelwertdifferenz der Sortenbezeichnungen A und E in Abbildung 6 für die Daten aus Abbildung 1 gezeigt wird (die Resultate in diesem Absatz sind identisch für die Daten in Abbildung 3, wenn wir dort den Vergleich der Sorten A und B betrachten; Tabelle 2). Die Verteilung ist bei Null zentriert, was korrekt die Tatsache widerspiegelt, dass beide Sortenbezeichnungen in Wirklichkeit von Sorte A stammen, so dass es keine Behandlungsdifferenz gibt. Dagegen finden wir für die systematische Anlage eine beobachtete Differenz von $D = -1.20$ (Tabelle 1), was viel weiter vom Zentrum der Randomisationsverteilung entfernt und daher als signifikant einzustufen ist ($P = 0.0011$). Dies zeigt, dass erhebliche Verzerrungen entstehen können, wenn nicht randomisiert wird. Des Weiteren ist die Randomisationsverteilung annähernd normal, so dass wir problemlos den t -Test für Mittelwertvergleiche verwenden können. Der P -Wert kann wieder analytisch berechnet werden durch Verwendung einer t -Verteilung oder durch Bestimmung der Randomisationsverteilung, welche durch Permutation der beiden Behandlungsbezeichnungen erhalten werden kann.

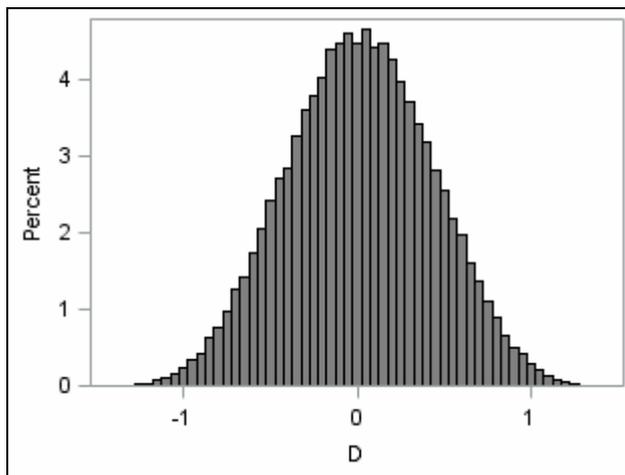


Abbildung 6: Randomisationsverteilung der Differenzen D (in $t \text{ ha}^{-1}$) der Mittelwerte für die Sortenbezeichnungen A und E basierend auf den Daten in Abbildung 1 (Anzahl Randomisationen = 10^5 ; Ergebnisse in dieser Abbildung gelten gleichermaßen auch für die Daten in Abbildung 3, wenn der Vergleich der Sorten A und B betrachtet wird)

3.2 Ein Blindversuch

Um weiter die Begrenzungen in den Analysemöglichkeiten bei fehlender Randomisation aufzuzeigen, haben wir einen Blindversuch mit Winterweizen analysiert, der auf dem Ihingerhof der Universität Hohenheim im Jahr 2007 auf einem Parzellengitter mit 36 Zeilen und 30 Spalten durchgeführt wurde. Auf allen 1080 Parzellen wurde dieselbe Sorte angesät und alle Parzellen wurden pflanzenbaulich einheitlich geführt, so dass es nur eine einzige Behandlung gab. Um die Wirkung eines systematischen Versuchsplans zu untersuchen, haben wir denselben systematischen Versuchsplan wie im vorangegangenen Abschnitt (Abbildung 2b) verwendet, d. h. wir haben ein gleitendes Fenster von 5 Zeilen x 4 Spalten verwendet, um kleine Versuche mit 20 Parzellen zu simulieren. Für jedes 5 Zeilen x 4 Spalten Gitter wurden die Behandlungen systematisch den Zeilen zugeordnet, d. h. Sorte A wurde jeweils den vier Par-

zellen in der ersten Zeile zugeordnet, Behandlung B wurde der zweiten Zeile zugeordnet, usw. Weil die Pläne den Blindversuchsdaten "übergestülpt" wurden, simulieren die resultierenden Versuche die globale Nullhypothese abwesender Behandlungseffekte (siehe auch RICHTER und KROSCHEWSKI, 2012).

Bei richtiger Randomisation sollten die analytischen P -Werte für die F -Tests einer Gleichverteilung folgen. Abbildung 7 zeigt, dass dies näherungsweise der Fall ist bei Randomisation, obwohl das Histogramm wegen der kleinen Zahl von Tests ($n = 864$) nicht völlig gleichmäßig verteilt sein kann. Ohne Randomisation sind kleine P -Werte klar überrepräsentiert, was anzeigt, dass der F -Test nicht gültig ist. Wir haben außerdem untersucht, ob sich das Problem durch eine geostatistische Analyse beheben lässt. Hierzu wurde ein separierbares AR1 x AR1 Modell (GILMOUR et al., 1997) verwendet. Um in fast allen Fällen Konvergenz des räumlichen Modells zu erzielen, wurden die Schätzwerte der Korrelationen auf unter 0.99 begrenzt. Die Signifikanz der F -Werte wurde mit der „first-order“ Methode von KENWARD und ROGER (2009) approximiert. Simulationsläufe, die nicht zur Konvergenz des REML Algorithmus führten, wurden verworfen.

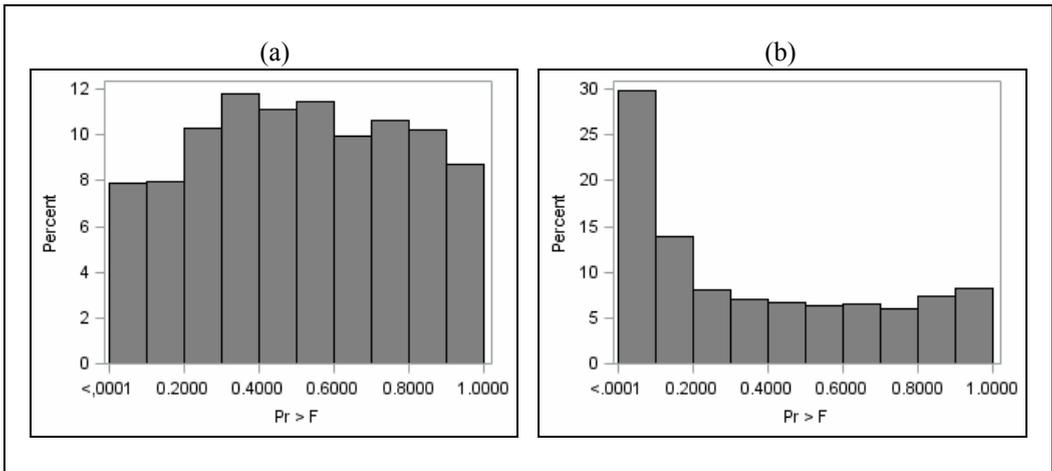


Abbildung 7: Histogramm von $n = 864$ analytischen P -Werten eines F -Tests für die Analyse der Blindversuchsdaten bei Annahme unabhängiger Fehler
 (a) Randomisierter Plan (b) Systematischer Plan

Abbildung 8 zeigt etwas Verbesserung, aber kleine P -Werte sind weiterhin überrepräsentiert und die Verteilung ist weit davon entfernt, gleichverteilt zu sein, wenn nicht randomisiert wurde. Im Fall einer Randomisation war die geostatistische Analyse leicht liberal. Selbstverständlich behaupten wir hier nicht, das jeweils am Besten passende räumliche Modell verwendet zu haben, so dass es andere Modelle geben mag, die hier zu einer etwas gleichmäßigeren Verteilung der P -Werte geführt hätten. Das Beispiel zeigt aber, dass von einer räumlichen Analyse nicht generell erwartet werden kann, dass sie ein unrandomisiertes Experiment „rettet“ (BAILEY et al., 1995; COX, 2009).

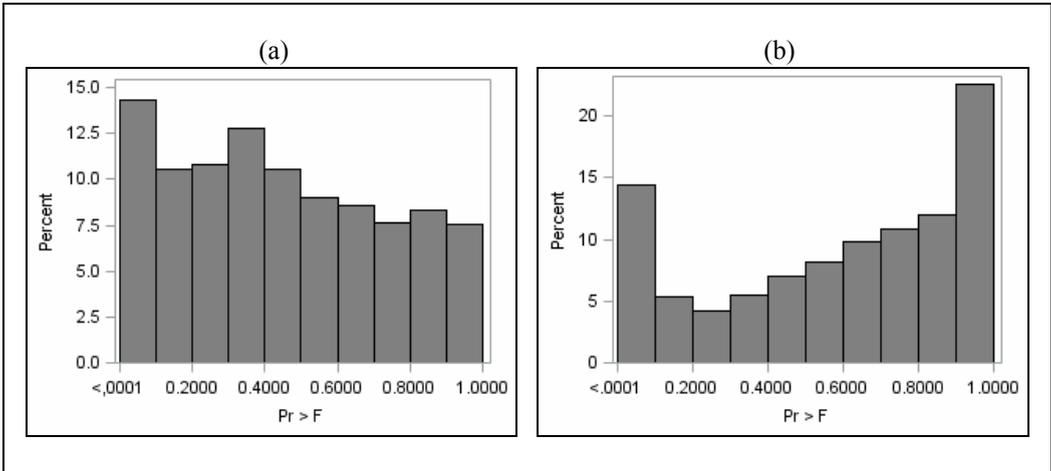


Abbildung 8: Histogramm von $n = 864$ analytischen P -Werten eines F -Tests für die Analyse der Blindversuchsdaten bei Annahme räumlich korrelierter Fehler ($AR1 \times AR1$)
 (a) Randomisierter Plan (b) Systematischer Plan

Die aus dem Blindversuch abgeleiteten Befunde sind mit etwas Vorsicht zu genießen wegen der relative kleinen Zahl von Tests und weil die mittels des gleitenden Fensters erhaltenen Tests und die zugehörigen P -Werte wegen der teilweisen Überlappung nicht statistisch unabhängig sind. Wir hätten hier versuchen können, das Problem zu umgehen, indem wir nur nicht-überlappende Fenster zugelassen hätten, aber dies hätte die Zahl der Tests übergebührlich reduziert, so dass die resultierenden Histogramme wenig aussagefähig geworden wären. Daher haben wir überlappende und folglich korrelierte Fenster von je 20 Parzellen bevorzugt. Im nächsten Abschnitt werden wir Simulationen verwenden, um diese Probleme zu umgehen.

3.3 Eine kleine Simulationsstudie

Um weiter zu untersuchen, welchen Effekt eine fehlende Randomisation hat und in welchem Ausmaß geostatistische Verfahren diesen Mangel ausgleichen können, haben wir Feldversuchsdaten nach einem geostatistischen Modell simuliert. Der Vorteil einer Simulation besteht darin, dass das datengenerierende Modell bekannt ist, so dass eine falsche Modellwahl ausgeschlossen werden kann. Im vorliegenden Fall wurden Parzellenfehler nach einem separierbaren $AR1 \times AR1$ Modell mit verschiedenen Werten für die Autokorrelationsparameter ρ_r und ρ_c für die Korrelation über die Zeilen und Spalten und räumlicher Varianz $\sigma_s^2 = 1$ simuliert. In einer weiteren Simulation wurde das räumliche Modell mit einem zufälligen Zeileneffekt kombiniert, welcher die Varianz $\sigma_r^2 = 0.5$ hat. In diesem Fall wurde die räumliche Varianz auf den Wert $\sigma_s^2 = 0.5$ reduziert. Die Zahl der Behandlungen (t) wurde gleich der Zahl der Zeilen (r) gesetzt, während die Zahl der Wiederholungen pro Behandlungen gleich der Zahl der Spalten war (c). Wir betrachten die Fälle $c = 4$ und $r = t = 5, 10$.

Hinsichtlich der Zuordnung der Behandlungen zu den Parzellen wurden zwei Szenarios untersucht:

- (1) Systematische Zuordnung der Behandlungen zu den Parzellen einer ganzen Zeile, d. h. alle Wiederholungen der Behandlung liegen in derselben Zeile.
- (2) Vollständig randomisierte Zuordnung.

Die simulierten Erträge hatten keine Behandlungseffekte, so dass die Daten die globale Nullhypothese der Abwesenheit von Behandlungseffekten repräsentieren.

Drei verschiedene Analyseverfahren hinsichtlich der Modellierung der Fehlerverteilung wurden verglichen:

- (i) Klassische Varianzanalyse unter der Annahme unabhängiger Fehler.
- (ii) Analyse unter der Annahme desselben $AR(1) \times AR(1)$ Modells (mit oder ohne zufälligem Zeileneffekt), welches zur Simulation der Daten verwendet wurde, wobei die Parameter des Modells bei den entsprechenden in der Simulation verwendeten Werten fixiert wurden (ein Wald F -Test wurde zum Test der globalen Nullhypothese verwendet; die Behandlungsdifferenzen wurden mit der verallgemeinerten Kleinst-Quadrat-Methode geschätzt).
- (iii) Gleiche Analyse wie (ii), aber mit Schätzung der Parameter des Fehlermodells aus den Daten. Für die Schätzung des geostatistischen Modells und für die Bestimmung der Signifikanz des F -Tests wurden dieselben Spezifikationen verwendet wie beim Blindversuch.

Es wurde die empirische Wahrscheinlichkeit des Fehlers 1. Art bei einem Signifikanzniveau von $\alpha = 5\%$ ermittelt. Außerdem haben wir den mittleren quadratischen Fehler (MSE) der Mittelwertdifferenzschätzungen erfasst.

Falls die Versuche randomisiert wurden, waren die klassische Varianzanalyse und die räumliche Analyse mit bekannter Varianz-Kovarianz-Struktur valide (Tabelle 3), wie von der Theorie her zu erwarten ist. Eine räumliche Analyse, welche das korrekte Modell verwendet, aber die Parameter aus den Daten schätzt, war approximativ gültig. Vor allem bei kleinen Versuchsumfängen kann die beobachtete Fehlerrate etwas vom nominalen Niveau abweichen. In unseren Simulationen tendierten die Tests dazu, etwas liberal zu sein, d. h. die Tests verwarfen die Nullhypothese etwas zu oft. Ohne Randomisation war die klassische Varianzanalyse eindeutig ungültig, weil die mit den Zeilen assoziierten räumlichen Effekte nicht vom Behandlungseffekt getrennt werden konnten. Wenn das korrekte räumliche Modell verwendet wurde und die Varianzparameter bei ihren wahren Werten fixiert wurden, waren die Signifikanztests ebenfalls noch gültig. Wurden die Parameter dagegen geschätzt, waren die Tests nicht mehr generell gültig. Insbesondere war das empirische Signifikanzniveau im Fall eines zufälligen Zeileneffekts liberal, weil dieser Zeileneffekt mit dem Behandlungseffekt vermischt ist.

Tabelle 3: Empirische Wahrscheinlichkeit des Fehlers 1. Art bei $\alpha = 5\%$ in Simulation (10.000 Simulationsläufe); $r =$ Zahl der Zeilen; $t =$ Zahl der Behandlungen; $c =$ Zahl der Spalten; $\rho_r, \rho_c =$ serielle Korrelationen für Zeilen und Spalten; $\sigma_r^2 =$ Varianz der Zeileneffekte. Die räumliche Varianz ist $\sigma_s^2 = 0,5$, wenn außerdem ein Zeileneffekt vorliegt, andernfalls ist $\sigma_s^2 = 1,0$; ANOVA = klassische Varianzanalyse mit unabhängigen Fehlern

$r=t$	c	ρ_r	ρ_c	σ_r^2	Nicht randomisiert			Randomisiert		
					ANOVA	ARxAR fixiert	ARxAR geschätzt	ANOVA	ARxAR fixiert	ARxAR geschätzt
5	4	0.2	0.8	-	0.852	0.049	0.086	0.053	0.054	0.067
		0.5	0.5	-	0.304	0.052	0.094	0.054	0.051	0.069
		0.8	0.2	-	0.032	0.050	0.054	0.054	0.052	0.063
10	4	0.2	0.8	-	0.984	0.051	0.061	0.050	0.054	0.052
		0.5	0.5	-	0.561	0.052	0.071	0.047	0.053	0.054
		0.8	0.2	-	0.079	0.051	0.062	0.050	0.054	0.051
5	4	0.2	0.8	0.5	0.964	0.049	0.105	0.055	0.055	0.078
		0.5	0.5	0.5	0.831	0.049	0.178	0.055	0.053	0.070
		0.8	0.2	0.5	0.680	0.051	0.376	0.054	0.052	0.061
10	4	0.2	0.8	0.5	1.000	0.052	0.103	0.051	0.055	0.060
		0.5	0.5	0.5	0.977	0.051	0.216	0.051	0.053	0.056
		0.8	0.2	0.5	0.894	0.051	0.576	0.049	0.053	0.047

Ohne Randomisation war der mittlere quadratische Fehler (MSE) der Differenzen etwa gleich für klassische Varianzanalyse und räumliche Analyse (Tabelle 4). Für randomisierte Versuche war der MSE in den meisten Fällen deutlich reduziert, vor allem bei einer räumlichen Analyse. Dies deutet an, dass eine Randomisation die Präzision erhöhen kann und dass eine räumliche Analyse effizienter sein kann als eine klassische Varianzanalyse, sofern richtig randomisiert wurde.

Aus der Simulationsstudie ergeben sich die folgenden Schlussfolgerungen:

- (i) Eine klassische Varianzanalyse kann deutlich ungültig werden, wenn nicht randomisiert wurde.
- (ii) Vorausgesetzt es wurde richtig randomisiert, ist eine klassische Varianzanalyse auch dann gültig, wenn die Parzellenfehler einem räumlichen Modell folgen. Dies liegt darin begründet, dass die Randomisation jegliche räumliche Korrelation bricht.
- (iii) Eine räumliche Analyse kann eine fehlende Randomisation nicht ersetzen, selbst wenn das Modell korrekt spezifiziert ist, weil immer die Gefahr einer Vermengung von Behandlungseffekten mit dem räumlichen Trend besteht.
- (iv) Fehlende Randomisation kann einen erheblichen Präzisionsverlust der Schätzungen von Mittelwertdifferenzen zur Folge haben.
- (v) Mit adäquater Randomisation ist eine räumliche Analyse approximativ gültig. Darüber hinaus kann eine räumliche Analyse die Präzision im Vergleich zu einer klassischen Varianzanalyse merklich verbessern, vorausgesetzt das Modell ist korrekt spezifiziert.

Tabelle 4: Mittlerer quadratischer Fehler für Mittelwertdifferenzen in Simulationen (10.000 Simulationsläufe); r = Zahl der Zeilen; t = Zahl der Behandlungen; c = Zahl der Spalten; ρ_r, ρ_c = serielle Korrelationen für Zeilen und Spalten; σ_r^2 = Varianz der Zeileneffekte; die räumliche Varianz ist $\sigma_s^2 = 0,5$, wenn außerdem ein Zeileneffekt vorliegt, andernfalls ist $\sigma_s^2 = 1,0$; ANOVA = klassische Varianzanalyse mit unabhängigen Fehlern

$r=t$	c	ρ_r	ρ_c	σ_r^2	Nicht randomisiert			Randomisiert		
					ANOVA	ARxAR fixiert	ARxAR geschätzt	ANOVA	ARxAR fixiert	ARxAR geschätzt
5	4	0.2	0.8	-	1.407	1.364	1.377	0.410	0.150	0.180
		0.5	0.5	-	0.718	0.696	0.716	0.405	0.247	0.308
		0.8	0.2	-	0.233	0.231	0.241	0.404	0.141	0.171
10	4	0.2	0.8	-	1.476	1.431	1.443	0.459	0.152	0.165
		0.5	0.5	-	0.848	0.822	0.841	0.446	0.245	0.270
		0.8	0.2	-	0.339	0.336	0.345	0.421	0.131	0.142
5	4	0.2	0.8	0.5	1.706	1.685	1.691	0.413	0.079	0.097
		0.5	0.5	0.5	1.362	1.352	1.361	0.410	0.143	0.175
		0.8	0.2	0.5	1.121	1.120	1.125	0.410	0.098	0.116
10	4	0.2	0.8	0.5	1.741	1.719	1.725	0.463	0.079	0.089
		0.5	0.5	0.5	1.428	1.415	1.425	0.457	0.141	0.158
		0.8	0.2	0.5	1.174	1.173	1.178	0.444	0.091	0.098

4 Abschließende Bemerkungen

Wir hoffen, dass dieser Artikel den Leser davon überzeugt hat, dass Randomisation ein elementarer Bestandteil eines gut geplanten Versuchs sein sollte. Der Einfachheit halber haben wir uns bei der Darstellung auf die vollständig randomisierte Anlage beschränkt. Es ist jedoch oft vorteilhaft, die Randomisation einzuschränken, um bekannte oder vermutete größere räumliche Trends zu berücksichtigen. Insbesondere stellt die Blockbildung, bei der die Zusammensetzung der Blöcke von der Randomisation nicht verändert wird, eine effiziente Möglichkeit dar, räumliche Trends zu berücksichtigen und den Versuchsfehler zu reduzieren (EDMONDSON, 1995). In manchen Fällen können effiziente systematische Versuchspläne ohne Blockbildung und Randomisation gefunden werden, wenn die Art des räumlichen Trends bekannt ist und eine einfache parametrische Form hat (EDMONDSON, 1993); aber die Gültigkeit der statistischen Analyse hängt stark von der Gültigkeit des angenommenen Modells ab. Falls der Trend nicht genau bekannt ist, was die Regel sein dürfte, ist es besser, eine Blockbildung in Kombination mit Randomisation vorzunehmen, weil diese Planungsstrategie keine spezifischen Annahmen über die Form des Trends macht und weil die Randomisation eine statistische Validität der Ergebnisse sicherstellt. In mehrfaktoriellen Experimenten besteht darüber hinaus oft die Notwendigkeit, die Randomisation für einen oder mehrere Faktoren aus technischen Gründen einzuschränken. Spaltanlagen sind ideal dafür geeignet, solchen Notwendigkeiten Rechnung zu tragen. Aber trotz der genannten Restriktionen sollte eine (beschränkte) Randomisation für diese Versuchsanlagen immer durchgeführt werden, um eine valide statistische Analyse zu gewährleisten.

Es gibt auch die Möglichkeit, das Versuchsdesign für eine spätere geostatistische Analyse zu optimieren. Die Präzision von Designs, die für eine rein geostatistische Analyse optimiert sind, kann jedoch gering sein, wenn das Modell falsch gewählt wird. Der große Vorteil einer Randomisation besteht darin, dass sie eine Versicherung bietet gegen jegliche Art von Trend (räumlicher Trend; Effekte, die durch Bearbeitungsgeräte verursacht werden wie Fahrspuren, zeitliche Effekte bei variierendem Erntetermin, etc.). Für die „Versicherung durch Randomisation“ zahlt man eventuell einen Preis in Form einer etwas reduzierten statistischen Genauigkeit, weil die Randomisation auch solche Pläne zulässt, die für eine bestimmte geostatistische Auswertung nicht optimal sind, aber dieser Preis ist in der Regel klein im Vergleich zum Vorteil einer verlässlichen und reproduzierbaren Fehlerschätzung. Der Preis kann außerdem durch eine effiziente Blockbildung minimiert werden (unvollständige Blöcke, Zeilen-Spalten-Pläne). Wenn eine geblockte Versuchsanlage für eine bestimmte geostatistische Auswertung optimiert werden soll, so muss die Randomisation eingeschränkt werden. Dies heißt, dass die Menge der möglichen Randomisationen, aus der dann für den geplanten Versuch eine Anordnung zufällig ausgewählt wird, reduziert ist gegenüber derjenigen Menge an Randomisationen, die möglich wäre bei vollständiger Randomisation (WILLIAMS et al., 2006). Die reduzierte Menge umfasst dann alle diejenigen Zuordnungen der Behandlungen zu den Parzellen, welche unter dem angenommenen räumlichen Modell die effizientesten Effektschätzungen erlauben. Obwohl es keine vollständige Randomisationstheorie gibt, wenn die Menge an Randomisationen für eine geostatistische Modellierung reduziert wird (WILLIAMS, 1986), so ist doch bei räumlicher statistischer Analyse mit einem Effizienzgewinn und mit einer validen Analyse zu rechnen, sofern das räumliche Modell richtig gewählt wurde. In der Praxis ist es allerdings schwierig, das am besten geeignete räumliche Modell sicher zu identifizieren. Eine konservative Strategie, die diesem Problem begegnet, besteht darin, die vollständige Menge aller Randomisationen für eine effiziente geblockte Versuchsanlage zu verwenden und dann bei der Analyse verschiedene geostatistische Zusatzkomponenten im Auswertungsmodell auszuprobieren, aber die rein randomisationsbasierte Analyse unter Annahme unabhängiger Restfehler als Rückfallposition zu behalten, falls eine räumliche Analyse zu keiner Verbesserung in der Modellanpassung führt (PIEPHO and WILLIAMS, 2010).

Es gibt Faktoren, die wissenschaftlich von Interesse sind, sich aber nicht randomisieren lassen, allen voran der Faktor Zeit bei Wachstumskurven. In diesem Fällen liegt eine Datenstruktur mit Messwiederholungen vor. Versuche mit Messwiederholungen erlauben eine valide statistische Auswertung, sofern die anderen Behandlungsfaktoren adäquat randomisiert wurden. Bei solchen Versuchen gibt es einen einfachen „Lackmustest“, ob eine valide Auswertung möglich ist. Dieser beruht auf der Frage: Ist eine Auswertung auch mit einem einzigen Zeitpunkt möglich? Ist die Antwort auf diese Frage „Ja“, dann ist der Versuch adäquat randomisiert, so dass Standardverfahren für die Auswertung von Versuchen mit Messwiederholungen verwendet werden können. In gemischten Modellen für Messwiederholungen werden die Beobachtungen von verschiedenen Beobachtungseinheiten (Parzelle, etc.) als unabhängig modelliert, während Beobachtungen zu verschiedenen Zeitpunkten bei derselben Beobachtungseinheit als seriell korreliert modelliert werden (PIEPHO et al., 2004; BRIEN and DEMETRIO, 2010). Ignorieren dieser Korrelationen führt in der Regel zu invaliden statistischen Analysen, welche vorgegebene Irrtumswahrscheinlichkeiten nicht einhalten.

Die in diesem Beitrag betrachteten Beispiele betrafen alle jeweils Feldversuche. Es soll aber betont werden, dass Randomisation natürlich ein universell anwendbares Prinzip der Versuchsplanung ist. Nach unserer Erfahrung werden systematische Versuchsanlagen zum Beispiel häufig bei Laboranalysen angewendet, indem biologische Replikate aus dem Feld, dem

Gewächshaus, dem Stall, usw. in systematischer Reihenfolge verarbeitet werden. Es werden also alle Replikate der Behandlung A zuerst verarbeitet, dann alle Replikate der Behandlung B, usw. Wenn es dann einen zeitlichen Trend gibt, zum Beispiel durch schleichende Veränderung der Einstellungen an einer Messapparatur, so ist mit derselben Art von Verzerrung zu rechnen wie in Feldversuchen. Solche Verzerrungen können ganz einfach durch eine Randomisation vermieden werden.

Danksagung

Wir bedanken uns bei Miriam Athmann, Wilhelm Claupein, Rüdiger Graß, Sabine Gruber und Hanspeter Thöni für konstruktive Kommentare zu einer früheren Version dieses Textes.

Literatur

- BAILEY, R. A.; AZAÏS, J.; MONOD, H. (1995): Are neighbour methods preferable to analysis of variance for completely systematic designs? 'Silly designs are silly!' *Biometrika* 82: 655-659.
- BRIEN, C. J.; DEMETRIO, C. G. B. (2010): Formulating mixed models for experiments, including longitudinal experiments. *Journal of Agricultural Biological and Environmental Statistics* 14: 253-280.
- CALINSKI, T.; KAGEYAMA, S. (2003): *Block Designs: A Randomization Approach. Volume I: Analysis*. Berlin: Springer.
- COX, D. R. (2009). Randomization in the design of experiments. *International Statistical Review* 77: 415-429.
- EDMONDSON, R. N. (1993): Systematic row-and-column designs balanced for low order polynomial interactions between rows and columns. *Journal of the Royal Statistical Society Series B* 55: 707-723.
- EDMONDSON, R. N. (1995): Past development and future opportunities in the design and analysis of crop experiments. *Journal of Agricultural Sciences, Cambridge* 143: 27-33.
- FISHER, R. A. (1925). *Statistical Methods for Research Workers*. Edinburgh: Oliver and Boyd.
- GILMOUR, A. R.; CULLIS, B. R.; VERBYLA, A. P. (1997): Accounting for natural and extraneous variation in the analysis of field experiments, *Journal of Agricultural, Biological and Environmental Statistics* 2: 269-293.
- GOOD, P. (2000): *Permutation Tests. A Practical Guide to Resampling Methods for Testing Hypotheses*. Berlin: Springer.
- JOHN, J. A.; WILLIAMS, E. R. (1995): *Cyclic and Computer Generated Designs*, 2nd ed. London: Chapman & Hall.
- KENWARD, M. G.; ROGER, J. H. (2009): An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis* 53: 2583-2595.
- MEAD, R.; CURNOW, R. N.; HASTED, A. M. (2002): *Statistical Methods in Agriculture and Experimental Biology*. Boca Raton: CRC Press.
- PIEPHO, H. P.; BÜCHSE, A.; RICHTER, C. (2004): A mixed modelling approach to randomized experiments with repeated measures. *Journal of Agronomy and Crop Science* 190: 230-247.
- PIEPHO, H. P.; RICHTER, V.; SPILKE, J.; HARTUNG, K.; KUNICK, A.; THÖLE, H. (2011): Statistical aspects of on-farm experimentation. *Crop and Pasture Science* 62: 721-735.
- PIEPHO, H. P.; WILLIAMS, E. R. (2010): Linear variance models for plant breeding trials. *Plant Breeding* 129: 1-8.

- PITMAN, E. J. G. (1938): Significance tests which may be applied to samples from any populations. III. The analysis of variance test. *Biometrika* 29: 322-335.
- RICHTER, C.; KROSCHESKI, B. (2012): Geostatistical models in agricultural field experiments: Investigations based on uniformity trials. *Agronomy Journal* 104: 91-105.
- SMITH, M. F. C. (1938): An empirical law describing heterogeneity in the yields of agricultural crops. *Journal of Agricultural Science, Cambridge* 28: 1-23.
- WILLIAMS, E. R. (1986): Neighbour analysis of uniformity data. *Australian Journal of Statistics* 28: 182-191.
- WILLIAMS, E. R.; LUCKETT, D. J. (1988): The use of uniformity data in the design and analysis of cotton and barley variety trials. *Australian Journal of Agricultural Research* 39: 339-350.
- WILLIAMS, E. R.; JOHN, J. A.; WHITAKER, D. (2006): Construction of resolvable row-column designs. *Biometrics* 62: 103-108.

Anagemethoden für On-Farm-Experimente

CHRISTEL RICHTER

Zusammenfassung

Im Folgenden sollen geeignete Anagemethoden für On-Farm-Experimente (OFE) beschrieben und Probleme der in der historischen Literatur empfohlenen Anlagen verdeutlicht werden. Das Ziel dieser Versuche soll der Vergleich mehrerer Prüfglieder sein, wobei deren Anzahl im Allgemeinen zwischen 2 und 4 liegt. Wahl der Anagemethode und Planung der Auswertung sind zwei Aspekte der Versuchsplanung, die nicht voneinander zu trennen sind. Aus diesem Grund wird stets die Verbindung zwischen Anlage und Auswertungsmodell angesprochen werden. Die Empfehlungen beziehen sich sowohl auf Einzelversuche als auch auf Versuchsreihen.

1 Überblick über geeignete und ungeeignete Anagemethoden

Einen Überblick über bekannte Anagemethoden gibt Abbildung 1. Obwohl seit FISHER (1935) die Bedeutung von Wiederholung, Randomisation und Blockbildung als Grundprinzipien der Versuchsplanung bekannt und im Allgemeinen akzeptiert sind, wurde und wird bis in die Gegenwart bei ihrer praktischen Durchsetzung ein Unterschied zwischen Parzellenversuchen in Versuchsstationen und Versuchen auf Praxisschlägen gemacht. Systematische Anlagen sind historisch die älteren Versuchsanlagen. Während sie für Versuche in Versuchsstationen schon lange nicht mehr akzeptiert werden, werden sie für OFE trotz ihrer Probleme teilweise noch heute genutzt. Dabei sind insbesondere die in Abbildung 1 aufgeführten zwei systematischen Anlagen anzutreffen. Ihre Nutzung resultiert aus der Annahme, dass Randomisation mit der in der Praxis üblichen Technik nicht möglich ist. Bei näherer Betrachtung stellt sich dabei jedoch oft heraus, dass das Zusammenspiel von Blockbildung, Randomisation und Randomisationseinheit einerseits und Versuchseinheit als Ebene der Merkmalerfassung andererseits und dem daraus resultierenden Auswertungsmodell offensichtlich fehlinterpretiert wird. Dieses wird anhand der zwei ausgewählten systematischen Anlagen erläutert. Im Folgenden werden Anagemethoden für Versuche mit einem Prüffaktor beschrieben.

1.1 Probleme der systematischen Anlagen

Die Bezeichnung ‚systematische Versuchsanlage‘ resultiert daraus, dass das Prinzip der Randomisation verletzt ist. Bereits 1949 wies MUDRA unter Bezugnahme auf die Arbeiten von FISHER auf die Problematik der bis zu diesem Zeitpunkt im deutschsprachigen Raum üblichen systematischen Anlagen hin. Die von ihm erwähnten Anlagen – Reihenanzordnung nach MITSCHERLICH, Schachbrettanordnung oder Standardmethode – zeichnen sich durch fehlende Randomisation, jedoch nicht durch das Fehlen von echten Wiederholungen aus. Die von ihm gleichfalls beschriebene Langparzellenanlage nach ZADE wird insbesondere unter dem Aspekt der versuchstechnischen Vereinfachung angesprochen. Abhängig davon, ob die Betrachtung aus der Sicht der tatsächlichen Versuchsanlage oder der gewählten Auswertung betrachtet wird, muss von fehlenden echten Wiederholungen bzw. fehlender Randomisation ausgegangen werden. Obwohl die Langparzellenanlage und gleichfalls der so genannte kontrollierte Anbauvergleich kritisch gesehen werden, werden beide Anlagen selbst heute noch nicht konsequent abgelehnt (THOMAS, 2006).

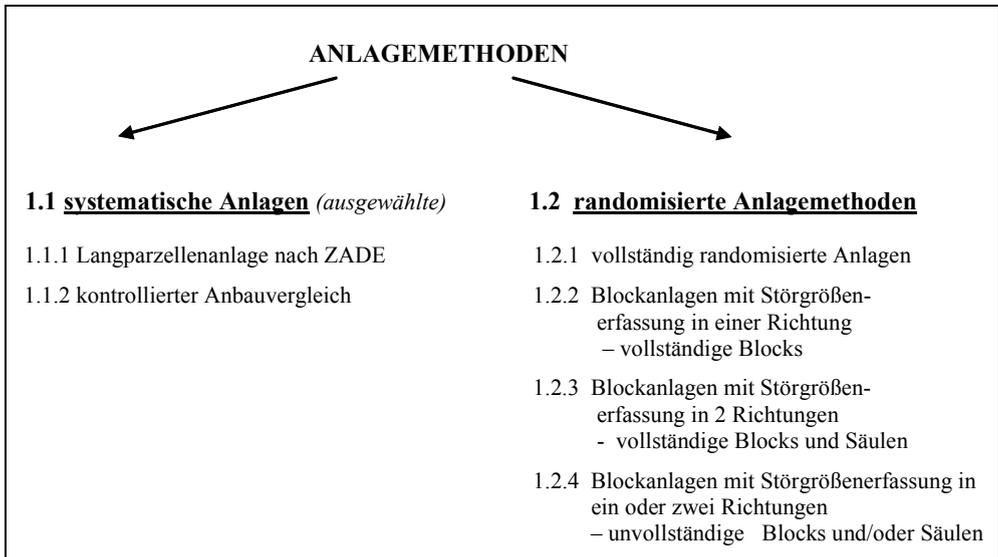


Abbildung 1: Überblick über bekannte Versuchsanlagen

1.1.1 Langparzellenanlage nach ZADE

Die Langparzellenanlage nach ZADE ist ein Spezialfall einer Versuchsanlage, die mit Standards arbeitet (Abbildung 2).

Die Prüfglieder werden je einmal in langen Streifen, in der Regel über die gesamte Schlaglänge, appliziert und nach zwei oder drei Prüfgliedern ein Standard dazwischen geschaltet. Der wiederholt in Streifen eingefügte Standard soll u. a. der Erfassung der Bodenheterogenität dienen. Quer zur Richtung der Streifen werden auf Ausschnittsparzellen, in Abbildung 2 zum Beispiel viermal (es können aber auch mehr sein), Merkmalerfassungen vorgenommen. Die in der Literatur beschriebene Auswertung besteht dann aus zwei Schritten. Im ersten Schritt wird unter Nutzung der Ausschnittsparzellen des Standards das ‚Ertragspotential‘ jeder Ausschnittsparzelle der Prüfglieder bestimmt. Dazu wird entweder zwischen den benachbarten Ausschnittsparzellen des Standards linear interpoliert (Differenzmethode; ZIMMERMANN, 1955) oder eine polynomiale Regressionsfunktion über alle Standardausschnittsparzellen geschätzt, wobei deren x- und y-Koordinaten die Regressoren sind (THOMAS und STRESSMANN, 1972). Auf dieser Grundlage erfolgt dann eine Korrektur der beobachteten Werte auf allen Ausschnittsparzellen der Prüfglieder. Nach dieser Ausschaltung der Bodenheterogenität wird im zweiten Schritt mit den korrigierten Daten eine Auswertung wie in einer randomisierten Anlage vorgenommen. ZIMMERMANN (1955) führt paarweise Prüfgliedvergleiche mit individuellen Standardabweichungen der Differenzen durch, ohne dabei zu berücksichtigen, dass es sich um Schätzwerte handelt; THOMAS und STRESSMANN (1972) arbeiten mit einer durchschnittlichen Standardabweichung, die im Ergebnis einer Varianzanalyse unter Berücksichtigung der Freiheitsgradreduktion durch die polynomiale Regressionsfunktion erhalten wird.

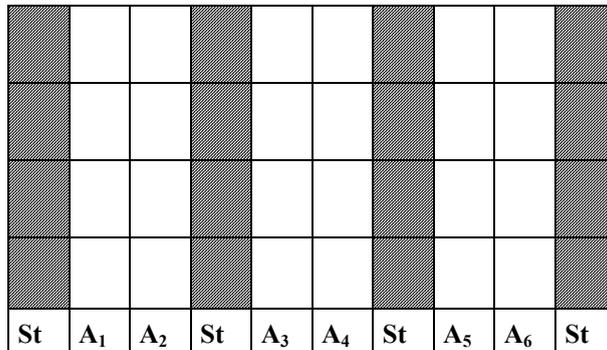


Abbildung 2: Langparzellenanlage mit 6 Prüfgliedern A1 ...A6 und dem Standardprüfglied St

Diese Versuchsanlage erfüllt nicht die Grundprinzipien der Versuchsplanung und die beschriebenen Auswertungen sind nicht adäquat:

- Für jedes Prüfglied existiert nur eine Randomisationseinheit (die Langparzelle), also eine echte Wiederholung. Damit ist das Grundprinzip des Wiederholens verletzt. Die Ausschnittsparzellen sind keine echten Wiederholungen, sie sind nur Pseudowiederholungen (wiederholte Messungen innerhalb der Parzelle).
- Die beschriebenen Auswertungen gehen davon aus, dass die Ausschnittsparzellen die Randomisationseinheiten sind. Auf die Ausschnittsparzellen bezogen ist das Prinzip der Randomisation verletzt.
- Behandelte man die Ausschnittsparzellen bei der Auswertung als wiederholte Messungen (wie es die durchgeführte Randomisation erforderte), so sind keine statistischen Schlüsse möglich, da nur eine echte Wiederholung vorliegt.

1.1.2 Kontrollierter Anbauvergleich

Abbildung 3 zeigt einen einfaktoriellen Anbauvergleich (auch kontrollierter Anbauvergleich genannt) mit zwei Stufen A1 und A2 eines Prüffaktors. A1 sei die Standardvariante und A2 die zu erprobende Variante. Dazu wird auf dem Schlag, dessen Hauptfläche der Standardvariante entspricht, ein Streifen (oder auch mehrere) mit der Variante A2 eingefügt. An den Trennlinien werden gegenüberliegende Ausschnittsparzellen beprobt (oft werden mehr als die in Abbildung 3 angedeuteten empfohlen). Diese gegenüberliegenden Ausschnitte werden als vollständige Blocks interpretiert und die Daten als randomisierte Blockanlage varianzanalytisch ausgewertet (Abbildung 3 zum Beispiel mit vier vollständigen Blocks). Dabei wird ignoriert, dass von A2 nur eine und von A1 zwei echte Wiederholungen vorliegen. Die Ausschnittsparzellen innerhalb eines Streifens sind wiederum nur Pseudowiederholungen. Auch wenn mehrere Streifen für A1 vorlägen, so sind in keinem Fall die Ausschnittsparzellen die Randomisationseinheiten.

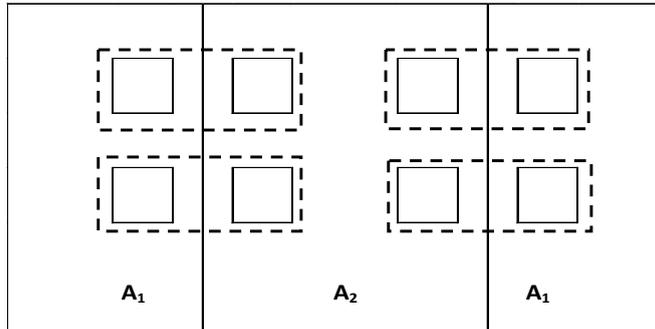


Abbildung 3: Einfaktorieller Anbauvergleich mit zwei Prüfgliedern;
 Interpretation als Block

1.1.3 Wertung der betrachteten systematischen Anlagen

Das Problem dieser Anlagen besteht nicht darin, dass überhaupt nicht randomisiert wird oder randomisiert werden kann. So können in der Langparzellenanlage die Prüfglieder den Streifen zufällig zugeordnet werden und in den Anbauvergleichen kann auch zufällig gewählt werden, wo die Streifen in den Schlag eingefügt werden. Damit sind aber die Streifen die Randomisationseinheiten und nicht die Ausschnittsparzellen – die Streifen sind echte Wiederholungen, die Ausschnittsparzellen sind nur wiederholte Messungen innerhalb der Streifen. Bei den beschriebenen Auswertungen werden aber die Ausschnittsparzellen wie Randomisationseinheiten behandelt.

Wenn bei Versuchen mit praxisüblichen Maschinen oft nicht kleinräumig randomisiert werden kann, so ist das auf Grund der technischen Gegebenheiten verständlich. Die anzutreffende Argumentation gegen randomisierte Anlagen in OFE resultiert aber offenbar aus dem Missverständnis, dass sich die Forderung nach Randomisation auf diese Ausschnittsparzellen beziehen würde. Sie bezieht sich aber auf die Streifen, was jedoch voraussetzt, dass sie mehrfach wiederholt werden. Wie den geforderten Grundprinzipien der Versuchsplanung nachgekommen werden kann, soll im Folgenden beschrieben werden. Zunächst werden die Grundsätze näher erläutert, in Abschnitt 2 und 3 wird deren Umsetzung in OFE beschrieben.

1.2 Randomisierte Anlagen

1.2.1 Vollständig randomisierte Anlagen

Vollständig randomisierte Anlagen gehen davon aus, dass die Prüfglieder den existierenden Versuchseinheiten ohne jede Einschränkung zufällig zugeordnet werden. Diese Versuchsanlage ist geeignet, wenn keine systematischen Störgrößen existieren bzw. zum Zeitpunkt der Versuchsplanung als solche erkannt werden.

Im Feldversuch muss jedoch davon ausgegangen werden, dass solche vorhanden sind. Klassische Begründung dafür ist die existierende Bodenheterogenität, ein Gefälle im Schlag oder vom Schlagrand ausgehende Störgrößen, die einen Gradienten auf der Fläche erzeugen, wie Wassergräben oder Baum- und Strauchbewuchs. Gibt es eine solche systematische Störgröße, so sollte eine randomisierte Blockanlage gewählt werden. Welche weiteren Störgrößen durch die Blockbildung erfasst werden können, wird in Abschnitt 1.2.5 erläutert.

1.2.2 Randomisierte Blockanlagen mit Störgrößenerfassung in einer Richtung – vollständige Blocks

Die Blocks werden quer zur Richtung der Störgröße angelegt und die Parzellen innerhalb eines Blocks sollten die in jedem Block existierende Heterogenität in möglichst gleicher Weise widerspiegeln. Abbildung 4 (links) zeigt, wie dieses umgesetzt werden kann; Abbildung 4 (rechts) wird dem nicht gerecht. Dabei kann die Störgröße in Form eines Trends (eines Gradienten von besseren zu schlechteren Bodeneigenschaften) wirken, könnte aber beispielsweise auch ein wellenförmiger Verlauf in der angegebenen Richtung sein. Entscheidend ist, dass innerhalb der Blocks weitgehend gleiche Bedingungen auf allen Parzellen und damit für alle darin liegenden Prüfglieder vorliegen. Wenn alle Prüfglieder in jedem Block genau einmal auftreten, handelt es sich um eine randomisierte Blockanlage mit vollständigen Blocks (Abbildung 4). Anlagen mit mehrfachem Auftreten der Prüfglieder in einem vollständigen Block würden die Blockgröße erhöhen und werden deshalb im Feldversuchswesen im Allgemeinen nicht gewählt. Ausnahmen liegen manchmal für Kontrollprüfglieder vor.

Die Prüfglieder werden den Parzellen innerhalb der Blocks zufällig zugeordnet. Die Parzellen sind hier also die Randomisationseinheiten. Die varianzanalytische Auswertung als randomisierte Blockanlage ist gerechtfertigt. Das Ziel der Blockbildung besteht darin, die durch die Störgröße hervorgerufene Variabilität im Prüfmerkmal zu erfassen und auszuschalten, so dass präzisere Schätzungen möglich sind. Darüber hinaus sollen durch die Parzellenanordnung in den Blocks mögliche Verzerrungen der Prüfgliedmittelwerte und -differenzen verhindert werden – alle Prüfglieder sind in gleicher Weise beispielsweise positiven wie negativen Bodenqualitäten ausgesetzt. Die Randomisation innerhalb der Blocks hat zweierlei Funktion: Zum einen gewährleistet sie, dass durch evtl. nicht erkannte Störgrößen innerhalb der Blocks kein Prüfglied im Mittel bevor- oder benachteiligt wird. Zum anderen bildet sie die Grundlage dafür, dass bei der Auswertung das klassische Modell einer randomisierten Blockanlage mit unabhängigen Resteffekten angesetzt werden darf - nur im Mittel aller Randomisationen kann von einer ‚Neutralisation‘ der (meist) existierenden Abhängigkeit ausgegangen werden.

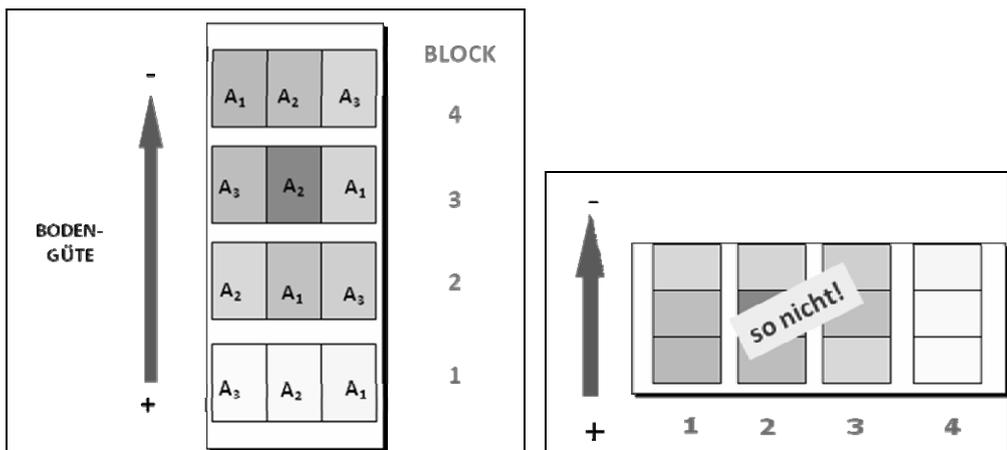


Abbildung 4: Randomisierte Blockanlage mit 3 Prüfgliedern und 4 vollständigen Blocks

An einem Beispiel soll demonstriert werden, welchen Einfluss die richtige Wahl der Parzellenanordnung im Block auf die Präzision eines Versuchs haben kann. Grundlage bildet ein Blindversuch, der mit verschiedenen Fruchtarten in fünf aufeinander folgenden Jahren auf

dem gleichen Schlag in Blumberg, einer Versuchsstation der Humboldt-Universität zu Berlin, durchgeführt wurde (RICHTER und KROSCHEWSKI, 2012). Die von Hand geernteten Teilstücke waren 2 m x 2 m groß. Abbildung 5 zeigt die Ertragskarte aus dem ersten Untersuchungsjahr mit der Fruchtart Hanf. Durch Zusammenfassung von fünf nebeneinander liegenden Teilstücken wurden Parzellen der Größe 10 m x 2 m gebildet, was einer üblichen Größe in Parzellenversuchen entspricht. Unter der Annahme, dass vier Prüfglieder getestet werden sollen, wurden drei verschiedene Anordnungen der Parzellen innerhalb eines Blocks untersucht, die zu folgenden Blockformen führen: 10 m x 8 m, 20 m x 4 m und 40 m x 2 m. Die gesamte Blindversuchsfläche wurde mit diesen Blocks ausgefüllt und anschließend als randomisierte Blockanlage ausgewertet. Abbildung 5 zeigt, dass in x-Richtung ein Gradient verläuft, der auch bei den anderen Fruchtarten festgestellt werden konnte.

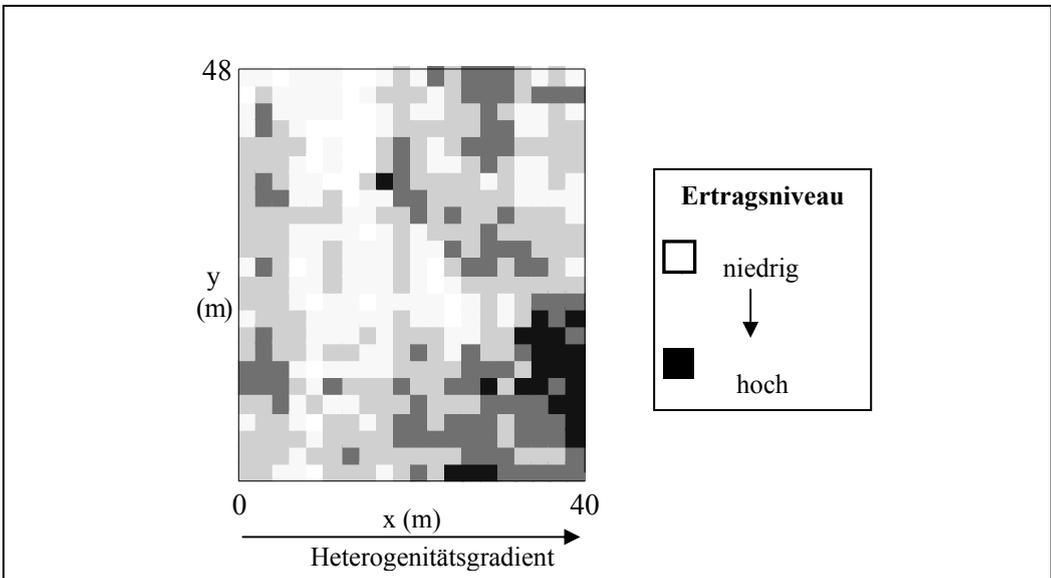
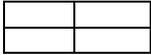


Abbildung 5: Klassifizierte Ertragskarte von Hanf (1. Jahr) für 2mx2m-Teilstücke

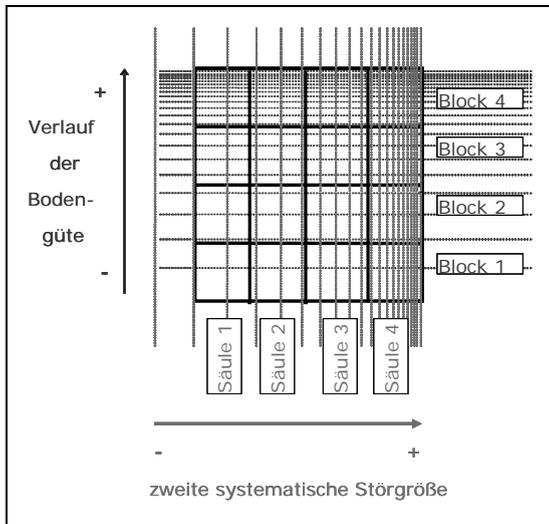
Demzufolge war nach den obigen allgemeinen Bemerkungen zu erwarten, dass die Blockform 10 m x 8 m am besten geeignet sein wird, die Variabilität zwischen den Parzellen innerhalb der Blocks zu reduzieren, gefolgt von 20 m x 4 m und von 40 m x 2 m. Diese Reihenfolge spiegelt sich auch in den extremen Unterschieden zwischen den $s\%(Rest)$ -Werten der Varianzanalyse wider (Tabelle 1). Dieses macht deutlich, wie wichtig die richtige Einschätzung der Heterogenität für präzise Schätzungen ist.

Tabelle 1: s% (Rest) in Abhängigkeit von der Anordnung der Parzellen im Block in einem Blindversuch mit vier Prüfgliedern und einer Parzellenform von 10 x 2 m

Blockform (m x m)	Anordnung der 4 Parzellen im Block	s %(Rest)				
		Hanf (1)	Roggen	Rüben	Hafer	Hanf (2)
10 x 8		9.2	3.5	7.0	12.6	7.4
20 x 4		12.0	5.4	12.6	27.6	8.0
40 x 2		18.5	11.8	15.6	35.3	10.8

1.2.3 Randomisierte Blockanlagen mit Störgrößenerfassung in zwei Richtungen – vollständige Blocks und Säulen

Existieren zwei Störgrößen, die nicht gleichgerichtet sind, so können in der einen Richtung Blocks und in der anderen Richtung Säulen gebildet werden. Enthält sowohl jeder Block als auch jede Säule jedes Prüfglied genau einmal, handelt es sich um das so genannte Lateinische Quadrat (Abbildung 6).



Block	Säule			
	(1)	(2)	(3)	(4)
(4)	A ₄	A ₂	A ₃	A ₁
(3)	A ₃	A ₁	A ₂	A ₄
(2)	A ₁	A ₃	A ₄	A ₂
(1)	A ₂	A ₄	A ₁	A ₃

Abbildung 6: Zwei nicht gleichgerichtete Störgrößen auf der Versuchsfläche und Lateinisches Quadrat mit vier Prüfgliedern

Im Lateinischen Quadrat werden die Prüfglieder den Prüfgliednummern eines Grundplanes zufällig zugeordnet und anschließend Blocks und Säulen jeweils separat zufällig vertauscht. Der in Abbildung 6 dargestellte Plan könnte bereits das Ergebnis dieser Randomisation sein, also ein möglicher Anlageplan sein, kann aber auch als *Grundplan* verwendet werden. Allgemein kennzeichnet die Bezeichnung Grundplan, dass es ein Plan ist, der die Charakteristika der jeweiligen Anlage erfüllt. Durch geeignete Randomisation, die der Spezifik der Anlage Rechnung tragen muss, wird daraus ein Anlageplan entwickelt.

1.2.4 Randomisierte Blockanlagen mit Störgrößenerfassung in ein oder zwei Richtungen – unvollständige Blocks und/oder Säulen

Die Notwendigkeit mit unvollständigen Blockanlagen zu arbeiten besteht dann, wenn auf Grund großer Teilstücke (Hauptargument in OFE) oder zu großer Prüfgliedanzahlen (möglichst zu vermeiden in OFE) die Blocks so groß werden würden, so dass innerhalb von vollständigen Blocks die angestrebten gleichen Bedingungen für alle Prüfglieder nicht mehr zu gewährleisten wären. In dem Fall werden unvollständige Blocks gewählt, wobei jeder dieser unvollständigen Blocks nicht mehr alle Prüfglieder enthält. Dann muss entschieden werden, welche Prüfglieder mit welchen anderen gemeinsam in den unvollständigen Blocks auftreten sollen. Die Zuordnung entscheidet darüber, welche Prüfgliedvergleiche mit welcher Präzision durchgeführt werden können. Hierfür seien einige ausgewählte Beispiele aufgeführt.

Blockanlagen – Störgrößen in einer Richtung – unvollständige Blocks

Es werden balancierte unvollständige Blockanlagen (BUB) und Anlagen mit unvollständigen Blocks, die nicht balanciert sind, unterschieden. Zur letzteren Gruppe gehören z. B. die teilweise balancierten unvollständigen Blockanlagen (TBUB). Die Bezeichnungen BUB und TBUB zielen vor allem auf die statistischen Eigenschaften eines Versuchsplanes ab. Die gleichfalls bekannten so genannten zyklischen Pläne und α -Pläne sind an bestimmte Konstruktionsmethoden gebunden und können im Spezialfall auch BUB oder TBUB sein. Abbildung 7 und Abbildung 8 zeigen Beispiele für zwei BUB und Abbildung 9 für einen TBUB. Die Graphen neben den Anlagen beschreiben die Struktur der Anlage: Die Knoten stellen die Prüfglieder dar; die Kanten kennzeichnen, welche Prüfglieder mit welchen anderen gemeinsam in einem unvollständigen Block auftreten. Besonders bekannt innerhalb der Gruppe der BUB sind die so genannten balancierten Gitter. In Abbildung 7 ist ein solches Gitter für vier Prüfglieder dargestellt. Allgemeines Kennzeichen der BUB ist, dass jedes Prüfglied mit jedem anderen gleich oft in den unvollständigen Blocks auftritt. Dadurch sind alle Prüfgliedvergleiche mit gleicher Präzision möglich - eine häufig erwünschte Eigenschaft. Damit ist ein BUB durch fünf Parameter, zwischen denen bestimmte Relationen bestehen müssen, vollständig charakterisiert: Anzahl der Prüfglieder, der Prüfglieder je unvollständigen Block, der Wiederholungen je Prüfglied und der unvollständigen Blocks sowie die Häufigkeit des gemeinsamen Auftretens jeweils zweier Prüfglieder im Gesamtversuch (in Abbildung 7 und Abbildung 8 kommt jedes Prüfglied mit jedem anderen genau einmal vor). Abhängig von der Prüfgliedanzahl und der Anzahl Prüfglieder je unvollständigen Block erfordert die Erreichung der Balanciertheit aber oft zu hohe Wiederholungsanzahlen (im Beispiel in Abbildung 8 beispielsweise fünf Wiederholungen bei sechs Prüfgliedern). Darüber hinaus existieren sie nicht für alle möglichen Parameterkonstellationen.

Wiederholung		
I	II	III
Block	Block	Block
(1) 1 2	(3) 1 3	(5) 1 4
(2) 3 4	(4) 2 4	(6) 2 3

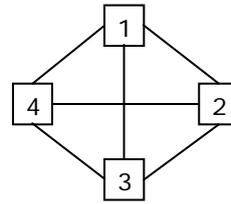


Abbildung 7: Balancierter unvollständiger Blockplan mit 4 Prüfgliedern, 2 Prüfgliedern je unvollständigen Block, 3 vollständigen Wiederholungen und 6 unvollständigen Blocks (die Zahlen entsprechen den Prüfgliednummern) – Grundplan vor Randomisation

Wiederholung				
I	II	III	IV	V
Block	Block	Block	Block	Block
(1) 1 2	(4) 1 3	(7) 1 4	(10) 1 5	(13) 1 6
(2) 3 4	(5) 2 5	(8) 2 6	(11) 2 4	(14) 2 3
(3) 5 6	(6) 4 6	(9) 3 5	(12) 3 6	(15) 4 5

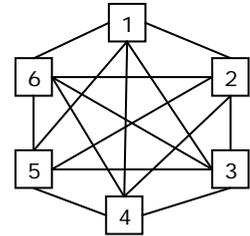


Abbildung 8: Balancierter unvollständiger Blockplan mit 6 Prüfgliedern, 2 Prüfgliedern je unvollständigen Block, 5 vollständigen Wiederholungen und 15 unvollständigen Blocks (die Zahlen entsprechen den Prüfgliednummern) – Grundplan vor Randomisation

Einen Ausweg können die TBUB darstellen. Hier ist die vollständige Balanciertheit nicht mehr gegeben, aber es lassen sich Gruppen (so genannte Assoziationsklassen) von Prüfgliedvergleichen bilden, die jeweils mit gleicher Präzision durchgeführt werden. Im Beispiel in Abbildung 9 gibt es drei solche Gruppen.

Wiederholung	
I	II
Block	Block
(1) 1 2	(4) 1 3
(2) 3 4	(5) 2 5
(3) 5 6	(6) 4 6

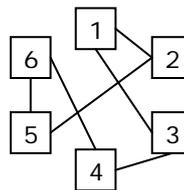


Abbildung 9: Teilweise balancierter unvollständiger Blockplan mit 6 Prüfgliedern, 2 Prüfgliedern je unvollständigen Block, 2 vollständigen Wiederholungen und 6 unvollständigen Blocks (die Zahlen entsprechen den Prüfgliednummern) – Grundplan vor Randomisation

Welcher Prüfgliedvergleich zu welcher Gruppe gehört, hängt davon ab, über welche Kanten des Graphen die beiden zu vergleichenden Prüfglieder erreichbar sind. Je kürzer die Pfade sind, umso größer ist die Präzision. Von welchem Prüfglied welches andere erreichbar ist, zeigt die Übersicht in Tabelle 2. Für die Präzision ist der jeweils kürzeste Pfad maßgeblich.

Damit haben die Prüfgliedvergleiche der ersten Assoziationsklasse die höchste Präzision, der zweiten eine niedrigere und der dritten die schlechteste.

Tabelle 2: Assoziationsklassen zum Versuch in Abbildung 9

Prüfgliedvergleich	“Erreichbarkeitspfade” im Graphen
1. Assoziationsklasse mit Pfadlängen 1 und 5	
1 - 2	1 - 2 und 1 - 3 - 4 - 6 - 5 - 2
1 - 3	1 - 3 und 1 - 2 - 5 - 6 - 4 - 3
2 - 5	2 - 5 und 2 - 1 - 3 - 4 - 6 - 5
3 - 4	3 - 4 und 3 - 1 - 2 - 5 - 6 - 4
4 - 6	4 - 6 und 4 - 3 - 1 - 2 - 5 - 6
5 - 6	5 - 6 und 5 - 2 - 1 - 3 - 4 - 6
2. Assoziationsklasse mit Pfadlängen 2 und 4	
1 - 4	1 - 3 - 4 und 1 - 2 - 5 - 6 - 4
1 - 5	1 - 2 - 5 und 1 - 3 - 4 - 6 - 5
2 - 3	2 - 1 - 3 und 2 - 5 - 6 - 4 - 3
2 - 6	2 - 5 - 6 und 2 - 1 - 3 - 4 - 6
3 - 6	3 - 4 - 6 und 3 - 1 - 2 - 5 - 6
4 - 5	4 - 6 - 5 und 4 - 3 - 1 - 2 - 5
3. Assoziationsklasse mit Pfadlängen zweimal 3	
1 - 6	1 - 2 - 5 - 6 und 1 - 3 - 4 - 6
2 - 4	2 - 5 - 6 - 4 und 2 - 1 - 3 - 4
3 - 5	3 - 1 - 2 - 5 und 3 - 4 - 6 - 5

In die Klasse der TBUB fallen zum Beispiel die bekannten Zwei- und Dreisatzgitter. Wie zuvor für die BUB beschrieben, existieren auch die TBUB nicht für alle Parameterkonstellationen, zumal zu den oben erwähnten fünf Parametern hier noch weitere für die Assoziationsklassen hinzukommen. Die oben erwähnten zyklischen Pläne oder α -Pläne stellen mögliche Alternativen dar.

Die Beispiele in den Abbildungen 7, 8 und 9 sind dadurch gekennzeichnet, dass jeweils zwei bzw. drei unvollständige Blocks zu vollständigen Wiederholungen gruppiert werden können. Anlagen mit unvollständigen Blocks, die diese Eigenschaft besitzen, werden *wiederholungsteilbare* Pläne genannt (englisch: resolvable designs). Dies kann bei Nutzung der REML-Methode zur Auswertung und Akzeptanz von Schätzwerten der Varianzkomponenten $> \text{Null}$ von Vorteil sein: Wenn die geschätzte Varianz zwischen den unvollständigen Blocks innerhalb der Wiederholungen Null ergibt, dann gehen die Ergebnisse automatisch in die einer vollständigen Blockanlage über (ein vollständiger Block entspricht dann einer Wiederholung). Bei der Planung eines Versuchs mit unvollständigen Blocks sollten nur so genannte verbundene Versuchspläne genutzt werden. Sie sind dadurch gekennzeichnet, dass jedes Prüfglied von jedem anderen über einen Pfad des Graphen erreichbar ist und alle Prüfglieddifferenzen mittels der Intra-Block-Analyse geschätzt werden können. Alle Pläne in Abbildung 7, 8 und 9 sind solche verbundenen Pläne. Abbildung 10 zeigt einen Plan, der weder wiederholungsteilbar noch verbunden ist. Die Prüfglieder zerfallen in zwei Gruppen (1, 2 und 3 einerseits und 4, 5 und 6 andererseits). Dieser Plan wäre nicht zu empfehlen.

Block	Block	Block
(1) 1 2	(3) 2 3	(5) 4 6
(2) 1 3	(4) 4 5	(6) 5 6

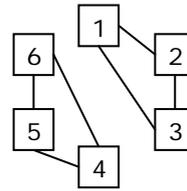


Abbildung 10: Unverbundener und nicht wiederholungsteilbarer unvollständiger Blockplan mit 6 Prüfgliedern, 2 Prüfgliedern je unvollständigen Block, 2 Wiederholungen je Prüfglied und 6 unvollständigen Blocks (die Zahlen entsprechen den Prüfgliednummern)

Wie auf der Basis eines existierenden Grundplans die Randomisation in Anlagen mit unvollständigen Blocks realisiert wird, ist an einem Beispiel demonstriert. Ausgehend von dem Grundplan aus Abbildung 7 wurden die Wiederholungen, die unvollständigen Blocks innerhalb der Wiederholungen und die Prüfglieder innerhalb der unvollständigen Blocks randomisiert, so dass der Anlageplan in Abbildung 13 (rechts) entstand.

Blockanlagen – Störgrößen in zwei Richtungen – unvollständige Blocks und vollständige Säulen

Eine der bekanntesten Anlagen mit Störgrößenerfassung in zwei Richtungen und unvollständigen Blocks, aber vollständigen Säulen ist die *Youden-Anlage*, auch unvollständiges Lateinisches Quadrat genannt (Abbildung 11). Sie entsteht aus dem Lateinischen Quadrat mit gleicher Prüfgliedanzahl durch Weglassen einer Säule. In den unvollständigen Blocks kommt jedes Prüfglied mit jedem anderen gleich oft – hier zweimal – vor.

Block	Säule			
	(1)	(2)	(3)	(4)
(4)	4	3	2	1
(3)	3	1	4	2
(2)	2	4	1	3
(1)	1	2	3	4

Abbildung 11: Youden-Anlage mit 4 Prüfgliedern, 3 Prüfgliedern je unvollständigen Block und 3 vollständigen Säulen (die Zahlen entsprechen den Prüfgliednummern) – Grundplan vor Randomisation

Allgemein werden Anlagen, welche die Möglichkeit der Störgrößenerfassung in zwei Richtungen erlauben – egal ob mit vollständigen oder unvollständigen Blocks und vollständigen oder unvollständigen Säulen – auch als *Zeilen-Spalten-Anlagen* bezeichnet. Zu den Anlagen mit unvollständigen Blocks und unvollständigen Säulen gehört zum Beispiel das so genannte *Gitterquadrat*, das als vollständig balanciertes, halbbalanciertes oder teilweise balanciertes Quadrat existiert. Neben den erwähnten Anlagen gibt es auch hier sehr viele Alternativen.

Die wenigen aufgeführten Beispiele dienen als Übersicht für besonders gebräuchliche Anlagemethoden, die auch für OFE von Interesse sind. Insbesondere ist bei den unter Abschnitt 1.2.4 erläuterten Anlagen vorstellbar, dass sie nicht einfach zu konstruieren sind. Dafür sollte Expertenwissen zu Rate gezogen werden. Das Programm CycDesignN (näheres siehe

<http://www.cycdesign.co.nz>) bietet eine Vielzahl von Möglichkeiten, Pläne mit Störgrößenerfassung in ein und zwei Richtungen zu konstruieren und anschließend zu randomisieren.

1.2.5 Abschließende Bemerkungen zur Blockbildung

Die Blockbildung – Vorgehen und Zielstellung – wurde bisher unter dem Aspekt von räumlich existierenden Störgrößen diskutiert. Daraus ergibt sich, dass der (vollständige oder unvollständige) Block eine räumlich homogene Einheit sein muss. Die verschiedenen Blocks eines Versuchs werden in der Regel auch kompakt auf einem Schlag liegen. Denkbar wäre aber auch eine Streulage auf dem Schlag, wenn anders gleiche Ausgangsbedingungen innerhalb der Blocks nicht gewährleistet werden können. Wichtig ist jedoch, dass die Bodenbedingungen für die einzelnen Blocks vergleichbar sind, da Wechselwirkungen zwischen Blocks und Prüfgliedern nicht geprüft werden können. Dieses resultiert daraus, dass jedes Prüfglied nur einmal in einem vollständigen Block bzw. einmal oder keinmal im unvollständigen Block vorkommt. Soll bei wiederholungsteilbaren unvollständigen Blockanlagen die oben erwähnte Option zur Auswertung als vollständige Blockanlage aufrecht erhalten werden, so müssen die unvollständigen Blocks, die zu einer Wiederholung gehören, auch kompakt auf einem Schlag angelegt werden. Wenn Blocks und Säulen gebildet werden, dann sollte in jedem Fall von einer kompakten Anordnung ausgegangen werden. Welche weiteren Überlegungen bei OFE in diesem Zusammenhang anzustellen sind, wird im Abschnitt 2 erläutert.

Blocks sollten aber nicht nur räumlich homogene Einheiten darstellen, sondern auch als homogene Organisationseinheit zur Ausschaltung eventueller personeller und zeitlicher Effekte genutzt werden. Arbeiten beispielsweise verschiedene Personen an dem Versuch oder erstreckt sich ein Arbeitsgang über einen längeren Zeitraum (unterschiedliche Witterungsbedingungen!), so sollte nicht prüfgliedweise, sondern stets blockweise vorgegangen werden. Kommen im OFE für einen Versuch mehrere Maschinen und Aggregate zum Einsatz (und diese sind nicht Stufen eines Prüffaktors), sollte immer die Einheitlichkeit je Block gewahrt werden. Wird so vorgegangen, dann manifestieren sich die aufgeführten Effekte als Blockeffekte, deren Einfluss eliminiert werden kann, und Verzerrungen von Prüfgliedeffekten werden somit verhindert.

2 Umsetzung in Einzelversuchen

Die Kontrolle der Versuchsbedingungen ist bei OFE meist sehr viel schwieriger als bei Parzellenversuchen in Versuchsstationen, da sehr viel mehr Störgrößen das Versuchsgeschehen und –ergebnis beeinflussen. Welche Störgrößen schon bei der Versuchsplanung berücksichtigt werden können, hängt zum einen vom vorhandenen Vorwissen über den Schlag ab, ist aber ebenso eine Frage der technisch-technologischen Durchführbarkeit des Versuchs. Sind bereits georeferenzierte Karten (z. B. Ertragskarten, Karten der elektrischen Leitfähigkeit) vorhanden, so können geeignete Teilflächen ausgewählt und Störgrößen bei der Versuchsplanung berücksichtigt werden. Andernfalls muss nach Augenschein und Erfahrung vorgegangen werden. Abhängig von den eingesetzten Maschinen und Aggregaten können die Parzellen, d. h. die Randomisationseinheiten, klein- oder großflächig angelegt und bearbeitet werden. Müssen aus technisch-technologischer Sicht Parzellen gebildet werden, die als lange Streifen über den gesamten Schlag verlaufen, so werden sowohl die Parzellen eines Blocks als auch die Blocks selbst häufig nebeneinander auf dem Schlag in Richtung der Fahrgassen angeordnet. Abbildung 12 (links) stellt eine Blockanlage dar, wo eher kleinräumig gearbeitet und randomisiert werden kann; Abbildung 12 (rechts) zeigt eine Blockanlage mit Parzellen, die über den ge-

samtent Schlag verlaufen, wo jeweils erst nach einer Durchfahrt (unter Umständen auch zwei Durchfahrten, siehe Abbildung 13 links) randomisiert werden kann.

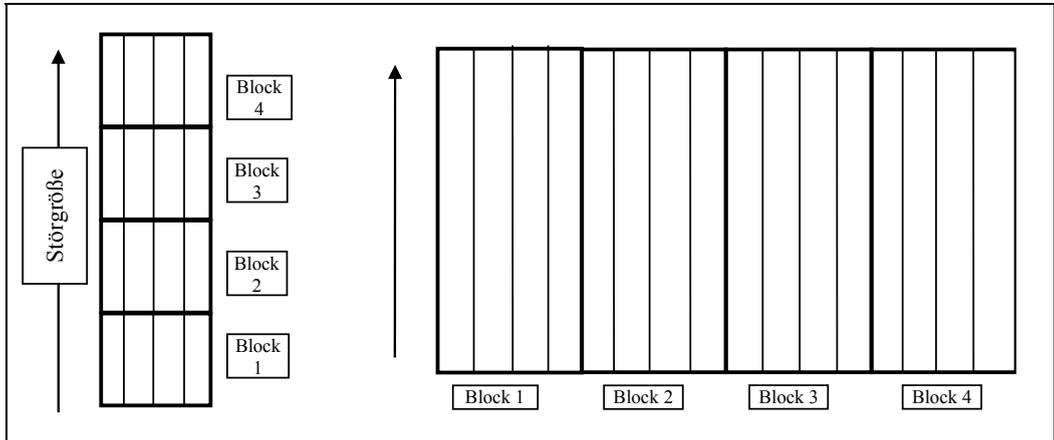


Abbildung 12: Zwei Lagepläne von Anlagen mit randomisierten Blocks für OFE mit 4 Prüfgliedern und 4 Blocks

Wichtig ist in jedem Fall, dass, wie in 1.2.2 beschrieben, je Block neu randomisiert wird (vgl. Abbildung 4), ansonsten ist keine valide Auswertung möglich. Diese Parzellen- und Blockanordnung gemäß Abbildung 12 ist auch nutzbar, wenn innerhalb der Parzellen mehrfache Messungen vorgenommen werden. Analog zu Abbildung 12 (rechts) zeigt Abbildung 13 (links) einen Versuch mit drei Prüfgliedern, realisiert als Anlage mit vollständigen Blocks und wiederholten Messungen. In Abbildung 13 (rechts) wurde ein Versuch mit vier Prüfgliedern in unvollständigen Blocks entsprechend dem Grundplan aus Abbildung 7 angelegt.

Verläuft die wesentliche Störgröße in der Hauptarbeitsrichtung, so ist bei einer Parzellenanordnung, wie in Abbildung 12 (rechts) und Abbildung 13 dargestellt, die dadurch verursachte Variation innerhalb jeder Parzelle eines Blocks zu finden. Blockunterschiede sind dann nicht mehr durch die Bearbeitungsrichtung, sondern durch andere (u. U. auch diskontinuierlich wirkende) Störgrößen und aus der Nutzung der Blocks als Organisationseinheiten zu erklären. Handelt es sich um ein Precision-Farming-Experiment, wo eine teilflächenspezifische Variante mit einer (oder mehreren) einheitlichen Applikationsvariante(n) verglichen werden soll, so erfordert die teilflächenspezifische Variante heterogene Flächen. Auch hier muss garantiert werden, dass alle Varianten vergleichbare Ausgangsbedingungen haben. Werden die Parzellen wie in Abbildung 12 (rechts) angelegt, liegt innerhalb jeder Parzelle vergleichbare Heterogenität vor.

Anlagen zum Störgrößenausgleich in zwei Richtungen stellen höhere Ansprüche an die Flexibilität bei der Randomisation. Mit Parzellen (Randomisationseinheiten), die über den ganzen Schlag verlaufen, sind sie nicht realisierbar. Sie kommen eher für kleinräumig arbeitende Aggregate in Frage.

Die bisherigen Erläuterungen machen deutlich, dass zur Beschreibung dieser Versuchsanlagen keine neuen Begriffsbildungen notwendig sind – es sind Anlagen mit vollständigen oder unvollständigen Blocks (unter Umständen auch Säulen) mit oder ohne wiederholten Messungen. Teilweise in Publikationen eingeführte Begriffe wie ‚Streifenversuch‘, ‚strip trial‘ oder ‚Streifenanlage‘ sind also völlig überflüssig, zumal oft unterschiedliche Auslegungen existieren. Die Tatsache, dass die Parzellen gegebenenfalls größer als im Parzellenversuch einer Ver-

suchsstation sind, hat nichts mit den Charakteristika der Anagemethode zu tun. Die Bezeichnung ‚Streifenanlage‘ sollte in keinem Fall genutzt werden, da diese eine andere wohldefinierte Versuchsanlage beschreibt (siehe z. B. THOMAS, 2006).

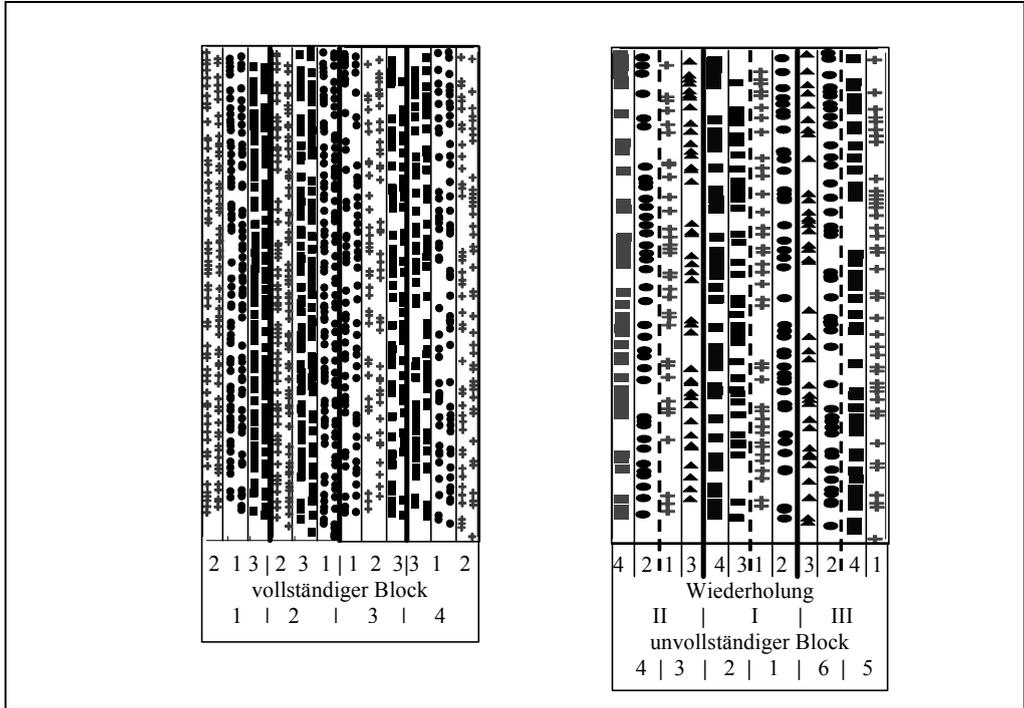


Abbildung 13: OFE mit wiederholten Messungen; links: 3 Prüfglieder, hier mit 2 Fahrspuren je Parzelle und 4 vollständigen Blocks; rechts: 4 Prüfglieder, hier mit einer Fahrspur je Parzelle und 6 unvollständigen Blocks (Plan aus Abbildung nach Randomisation).

Die Symbole geben die Positionen der wiederholten Messungen an.

Auch die Tatsache, dass wiederholte Messungen innerhalb von Parzellen vorliegen, ist keineswegs eine Spezifik der OFE. In Abhängigkeit von der Erfassungsart von Prüfmerkmalen kommt sie auch bei klassischen Parzellenfeldversuchen in Versuchsstationen vor. Wird beispielsweise ein Merkmal an mehreren Pflanzen oder Teilflächen jeder Parzelle bonitiert, gemessen oder ausgezählt, so liegt genau die gleiche Situation vor. Meist wird dann ein Mittelwert je Parzelle berechnet und die Mittelwerte werden entsprechend der Versuchsanlage varianzanalytisch ausgewertet. Das gleiche Vorgehen ist auch bei OFE nicht nur möglich, sondern sollte auch als Basisauswertung genutzt werden. Nur wenn der Versuch auf diese Weise ausgewertet werden kann, erfüllt er die Grundprinzipien der Versuchsplanung. Sollen weitere Analysen unter Nutzung der Einzelwerte vorgenommen werden, so dient das Ergebnis auf der Basis der Mittelwerte als Vergleichsmaßstab. Gleiches wäre auch im Parzellenfeldversuch möglich. Wie die wiederholten Messungen je Parzelle in die Auswertung der Versuchsanlage eingebunden werden kann, wird bei RICHTER und THÖLE in diesem Band erläutert.

Bisher wurde davon ausgegangen, dass alle Blocks, die zu einem Versuch gehören, auf einem Schlag liegen. Es sind aber auch Situationen denkbar, wo sie (aus welchen Gründen auch immer) auf mehrere Schläge eines Ortes verteilt sind. Abbildung 14 zeigt mögliche Situatio-

nen und deren Konsequenzen bei der Auswertung. Aus der Sicht der Einhaltung des ceteris-paribus-Prinzips wird im Allgemeinen die Anlage aller Blocks auf einem Schlag vorzuziehen sein. Werden die Blocks eines Versuchs jedoch auf mehrere Schläge aufgeteilt, so sollten auf diesen gleiche Konstantfaktoren der Versuchsdurchführung gewährleistet sein, d. h. sämtliche acker- und pflanzenbaulichen Maßnahmen, einschließlich der Vorfrucht, müssen einheitlich gestaltet werden. Im Beispiel eröffnet die Aufteilung auf die zwei Schläge die Möglichkeit, die Wechselwirkung Schlag x Prüfglied zu testen. Allerdings erschweren existierende Wechselwirkungen häufig die Interpretation. Bei der Aufteilung auf vier Schläge ist die Wechselwirkung nicht mehr schätzbar. Auf keinen Fall sollte die Anordnung in Abbildung 14 (unten) gewählt werden, hier ist überhaupt kein valider Vergleich zwischen den Prüfgliedern möglich.

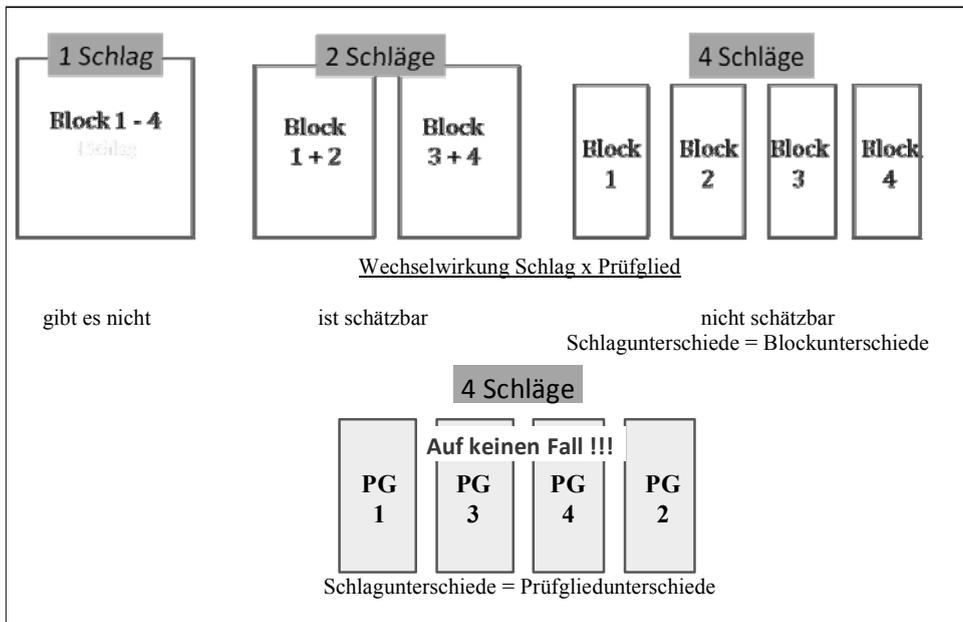


Abbildung 14: Einzelversuch mit 4 Prüfgliedern. Oben: 4 vollständige Blocks, deren mögliche Anlage auf einem oder mehreren Schlägen und deren Konsequenz. Unten: Anlage aller 4 Prüfglieder auf je einem Schlag.

3 OFE in Serie

Wie die Parzellenfeldversuche können auch OFE als mehrjährige, mehrortige oder mehrjährige und mehrortige Serie angelegt werden, um den Aussagebereich zu erweitern. Alle zuvor empfohlenen Anlagen können für jeden Einzelversuch der Serie genutzt werden. Zu beachten ist, dass je Jahr, Ort und Block neu zu randomisieren ist.

Die Art der Auswahl der Orte (Jahre) – zufällig oder fix – entscheidet über den Aussagebereich der Serie. Die Anzahl der Schläge und Blocks je Ort/Jahr und die Verteilung der Blocks auf die Schläge innerhalb der Orte und/oder Jahre entscheidet darüber, welche Wirkungen geprüft werden können. Abbildung 15 zeigt beispielhaft eine mehrortige Versuchsserie, in der die vier betrachteten Situationen aus Tabelle 2 des Leitfadens veranschaulicht sind. Im ersten

Fall sind die Ortseffekte mit den Blockeffekten vermengt, so dass Wechselwirkungen zwischen Orten und Prüfgliedern nicht schätzbar sind. Oft ist aber gerade dieses der interessante Aspekt einer Versuchsserie. Dieses kann durch die anderen Zuordnungen in Abbildung 15 erreicht werden. Hierbei wird von vollständigen Blocks ausgegangen. Welche weiteren Wirkungen jeweils schätzbar sind, kann dem Leitfaden (Tabelle 2) entnommen werden und hängt nicht von der Existenz wiederholter Messungen ab. Auch Anlagen mit unvollständigen Blocks sind möglich, dann hängt aber die Schätzbarkeit der einzelnen Effekte stark von den Eigenschaften des Plans und deren Aufteilung auf Schläge und Orte ab.

1 Anbaugebiet							
Variante	Ort 1	Ort 2	Ort 3	Ort ...	Ort 20		
1	1 Schlag = 1 Block	Wechselwirkung Prüfglied x Ort nicht schätzbar					
	1 Schlag > 1 Block / Schlag		Wechselwirkung Prüfglied x Ort schätzbar				
1 Schlag > 1 Block / Schlag							
2	1 Schlag > 1 Block / Schlag	Wechselwirkung Prüfglied x Ort schätzbar					
	1 Schlag > 1 Block / Schlag						
3	1 Schlag > 1 Block / Schlag	Wechselwirkung Prüfglied x Ort schätzbar					
	1 Schlag > 1 Block / Schlag						
4	1 Schlag > 1 Block / Schlag	Wechselwirkung Prüfglied x Ort schätzbar					
	1 Schlag > 1 Block / Schlag						

Abbildung 15: Mögliche Varianten einer mehrortigen Versuchsserie und Konsequenzen für die Wechselwirkung Prüfglied x Ort. Hier: Block = vollständiger Block

4 Abschließende Bemerkungen

Die Planung eines OFE sollte wie die eines Parzellenfeldversuchs entsprechend den Grundsätzen Randomisation, Wiederholung (gemeint sind echte Wiederholungen) und Blockbildung vorgenommen werden. Sie bilden die Basis dafür, dass eine valide Auswertung der Ergebnisse vorgenommen werden kann. Wiederholte Messungen (Pseudowiederholungen) des interessierenden Merkmals an sich bringen für den eigentlichen Prüfgliedvergleich zunächst keinen Gewinn, es sei denn, dass Varianzheterogenität der Prüfglieder vorliegt. Erst durch die eventuelle Einbeziehung von Kovariablen, die von den gleichen Positionen wie das interessierende Merkmal vorliegen, und/oder geostatistische Analysen können die Pseudowiederholungen für eine detailliertere Analyse nutzbar gemacht werden. Ob dadurch ein zusätzlicher Gewinn an Präzision ermöglicht wird, hängt sehr stark von den Gegebenheiten am Versuchsort und von den Versuchsergebnissen ab, ist also datengeleitet und im Prinzip nicht planbar. Dieses unterstreicht die Bedeutung der im Zuge der Versuchsplanung beeinflussbaren Aspekte: Wahl der Versuchsanlage und Wahl der Anzahl der echten Wiederholungen.

Versuchsanlagen, die dem Ziel einer lokalen Schätzung optimaler Applikationen dienen, wurden hier nicht angesprochen. Dazu sei auf PIEPHO et al. (2011) verwiesen.

Literatur

- FISHER, R. A. (1935): *The Design of Experiments*. Edinburgh, Oliver and Boyd.
- PIEPHO, H.-P.; RICHTER, C.; SPILKE, J.; HARTUNG, K.; KUNICK, A.; THÖLE, H. (2011): Statistical aspects of on-farm experimentation. *Crop & Pasture Science* 62: 721-735.
- RICHTER, C.; KROSCHESKI, B. (2012): Geostatistical models in agricultural field experiments – investigations based on uniformity trials. *Agron. J.* 104: 91-105.
- THOMAS, E. (2006): *Feldversuchswesen*. Stuttgart, Eugen Ulmer.
- THOMAS, E.; STRESSMANN, U. (1972): Auswertung von Standardversuchen mittels Ausgleichspolynoms. *Arch. Acker- u. Pflanzenbau u. Bodenkd.* 16: 655-663.
- ZIMMERMANN, K. (1955): *Technik des Versuchswesens und der Pflanzenzüchtung*. Leipzig, Hirzel Verlag.

Entwicklung des Auswertungsmodells

JOACHIM SPILKE

Zusammenfassung

OFE stellen eine besondere Anforderung an die Modellwahl, da neben den geplanten Prüf- und Designfaktoren oft weitere feste und zufällige Einflussgrößen wie beispielsweise räumlicher Trend, räumliche Kovarianzen oder weitere als Kovariable fungierende Merkmale im Auswertungsmodell zu berücksichtigen sind. Das gilt insbesondere, wenn die Analyse auf Einzelwerten innerhalb Parzelle basiert. Der Beitrag beschreibt eine Schrittfolge zur Modellwahl. Basierend auf den analytischen Kriterien AIC und BIC sowie einer Residuenanalyse zur Überprüfung der ermittelten Erwartungswertstruktur und Varianzfunktion zur Kontrolle der ermittelten Kovarianzstruktur wird gezeigt, wie schrittweise das Auswertungsmodell entwickelt und die eigene Modellwahl kontrolliert werden kann. Die Auswertungsergebnisse bei Nutzung der Parzellenmittel und Einzelwerte werden gegenüber gestellt.

1 Einleitung

Ein wesentlicher Vorteil von OFE besteht in der Nähe zu landwirtschaftlichen Produktionsbedingungen und der damit verbundenen guten Übertragbarkeit der Ergebnisse auf die landwirtschaftliche Praxis. Dieser Vorteil bedingt aber, neben den mit der Versuchsplanung festgelegten Prüf- und Designfaktoren weitere Einflussfaktoren im Sinne von Störgrößen zu berücksichtigen. Die Nichtbeachtung der Störgrößen kann zu einer Verletzung der Treffgenauigkeit und einer unkorrekten Schätzung der Wiederholungsgenauigkeit führen. Verzerrte Parameterschätzungen und/oder eine Nichteinhaltung der Fehlerwahrscheinlichkeiten von Inferenzaussagen sind die Folge. Daher ist die Festlegung eines den auszuwertenden Daten adäquaten Auswertungsmodells eine unverzichtbare Aufgabe im Zusammenhang mit der Auswertung von OFE. In diesem Beitrag wird dazu eine entsprechende Vorgehensweise vorgeschlagen und an einem Beispiel demonstriert.

2 Vorgehensweise bei der Modellwahl

Modellwahl bedeutet im vorliegenden Sachzusammenhang festzulegen, welche festen und zufälligen Faktoren über die Prüf- und Designfaktoren hinaus in das Auswertungsmodell aufzunehmen sind. Dabei beeinflussen feste Effekte die Erwartungswertstruktur, zufällige Effekte die Kovarianzstruktur. In Abhängigkeit des zu bearbeitenden Problems kann die Anzahl der jeweils möglichen Varianten sowohl für die Erwartungswert- als auch die Kovarianzstruktur und damit die Zahl der möglichen Kombinationen groß sein. Aus diesem Grund wird neben einer gemeinsamen Optimierung von Erwartungswert- und Kovarianzstruktur (NGO and BRAND, 1997) auch eine mehrstufige Vorgehensweise vorgeschlagen (WOLFINGER, 1993).

In diesem Beitrag wird die folgende Schrittfolge gegangen:

1. Schritt: Vorläufige Auswahl der Erwartungswertstruktur bei Unterstellung von Resteffekten mit diagonaler Struktur und homogener Varianz, geschätzt mit der ordinary least square (OLS) Methode (SEARLE, 1971);
2. Schritt: Auswahl der Kovarianzstruktur bei Verwendung der Erwartungswertstruktur im Ergebnis des ersten Arbeitsschrittes;

3. Schritt: Endgültige Festlegung der Erwartungswertstruktur durch Signifikanzprüfung von Störfaktoren unter Verwendung der optimierten Kovarianzstruktur.

Diese Vorgehensweise erlaubt in Schritt 1 die Verwendung der OLS-Residuen für jedes untersuchte Modell zur Kontrolle auf Verzerrungen. Das ist im vorliegenden Fall besonders bedeutsam, da damit überprüft werden kann, ob ein möglicherweise vorliegender räumlicher Trend im Auswertungsmodell richtig und vollständig berücksichtigt wurde. Weiterhin ergibt die Verteilung der Residuen in der Fläche einen Hinweis auf mögliche Varianzheterogenität, beispielsweise zwischen Blocks. Schließlich kann die für die Residuen beobachtete Kovarianzstruktur mit der im Ergebnis von Schritt 2 geschätzten Struktur mit Hilfe der Varianzfunktion verglichen werden. Dabei sollte die geschätzte Kovarianzstruktur im Wesentlichen die Verteilung der Residuen abbilden. Um diese Vorgehensweise realisieren zu können, wird gegenüber dem Vorschlag von WOLFINGER (1993) eine umgekehrte Schrittfolge gewählt.

Bei der Bestimmung der vorläufigen Erwartungswertstruktur sollten gegebenenfalls sogar überparametrisierte Modelle bevorzugt werden. Das heißt, neben allen Prüf- und Designfaktoren sollten alle möglichen Störgrößen im Modellansatz Berücksichtigung finden. Dadurch wird erreicht, dass die OLS-Methode konsistente Schätzer für die festen Effekte liefert. Liegt die optimierte Kovarianzstruktur vor, kann in einem 3. Schritt die Signifikanzprüfung von Stufen fester Störfaktoren oder von Regressionskoeffizienten, zugehörig zu Kovariablen (abgeleitet aus den stetigen Störgrößen), beispielsweise mit Hilfe des approximativen t- bzw. F-Tests erfolgen (GIESBRECHT and BURNS, 1985; FAI and CORNELIUS, 1996; KENWARD and ROGER, 1997, 2009; SPILKE et al., 2005). Für den denkbaren Fall, dass im ersten Schritt als bedeutsam identifizierte feste Störgrößen bei Nutzung der optimierten Kovarianzstruktur keine Signifikanz aufweisen, ist im 3. Schritt nochmals eine Analyse der Residuen durchzuführen. Dadurch soll der fälschliche Ausschluss wichtiger Störgrößen im Sinne eines Fehlers 2. Art vermieden werden. Das ist erforderlich, da Erwartungswert- und Kovarianzstruktur nicht voneinander unabhängig sind. Zusätzlich kann abschließend die Verteilung der standardisierten Residuen und möglicherweise weiterer vorhandener zufälliger Modelleffekte überprüft werden, beispielsweise mit Hilfe von Histogrammen. Durch den Vergleich der aus OLS-Residuen und den geschätzten Parametern ermittelten Varianzfunktion besteht eine weitere Möglichkeit einer anschaulichen Kontrolle und eines Vergleichs der untersuchten Modelle.

3 Methoden zur Modellselektion

Nachfolgend wird die Nutzung des Likelihood-Ratio Tests, analytischer Kriterien sowie der Residuenanalyse für die Modellselektion beschrieben.

3.1 Likelihood-Ratio Test (LRT)

Mit dem Likelihood-Ratio Test besteht die Möglichkeit des Vergleichs der Modellanpassung, sofern eines der Modelle dem anderen hierarchisch untergeordnet ist. Das ist der Fall, falls durch Fixierung bestimmter Modelleffekte ein Modell als Spezialfall s eines allgemeinen Modells g angesehen werden kann. Weiter seien $L(\hat{\theta}_s)$ und $L(\hat{\theta}_g)$ die geschätzte Likelihood bei Nutzung des Parametervektors des speziellen (s) bzw. des allgemeinen (g) Modells. Der LRT ergibt sich dann aus:

$$LRT = -2 \log \left(\frac{L(\hat{\theta}_s)}{L(\hat{\theta}_g)} \right) = -2 \log L(\hat{\theta}_s) + 2 \log L(\hat{\theta}_g) \sim \chi^2(FG).$$

Die Teststatistik des LRT folgt asymptotisch einer χ^2 -Verteilung, wobei sich die Freiheitsgrade (FG) aus der Anzahl Restriktionen ergeben die erforderlich sind, um das allgemeine Modell g in den Spezialfall s zu überführen (FAHRMEIER and HAMERLE, 1984; GREENE, 2003). Die Modellanpassung des allgemeinen Modells gegenüber dem speziellen Modell wird dann als besser angesehen, falls $LRT > \chi^2(1-\alpha, FG)$ ermittelt wird, wobei im Allgemeinen $\alpha \leq 0.05$ genutzt wird.

Falls zwei Modelle bezüglich der Erwartungswertstruktur verglichen werden sollen, sind die Likelihoodfunktionen nach der klassischen ML-Methode zu bestimmen und zu maximieren. Soll sich der Modellvergleich auf die Kovarianzstruktur bei konstanter Erwartungswertstruktur beziehen, sind die Likelihoodfunktionen nach der REML-Methode (Restricted Likelihoodfunktionen) zu nutzen (WOLFINGER, 1993; NGO and BRAND, 1997).

Der LRT basierend auf dem Quotienten von Restricted Likelihoodfunktionen wird in der Literatur auch mit RLRT bezeichnet. Zu den klassischen Voraussetzungen des LRT gehört, dass der Test basierend auf unabhängigen Zufallsgrößen y_1 bis y_n durchgeführt wird und dass die Parameter unter der Nullhypothese nicht auf dem Rand des Parameterraumes liegen. Diese Voraussetzungen sind im linearen gemischten Modell beispielsweise beim Test von Hypothesen der Gestalt $H_0 : \sigma_u^2 = 0$ gegen $H_1 : \sigma_u^2 > 0$ nicht erfüllt. Im vorliegenden Fall erfolgt daher eine Beschränkung, indem die Anwendung des LRT auf die Festlegung der Erwartungswertstruktur begrenzt wird.

3.2 Informationskriterien

Für einen Modellvergleich ohne die Voraussetzung einer hierarchischen Verknüpfung der zu vergleichenden Modelle stehen analytische Kriterien zur Verfügung. Häufig benutzt werden die Informationskriterien von AKAIKE (1969, 1974) sowie das Kriterium von SCHWARZ (1978). Die Berechnung dieser Kriterien zum Vergleich der Erwartungswertstruktur bei Nutzung der ML-Methode erfolgt nach:

$$AIC_{ML} = -2 \log L(\hat{\theta}) + 2(p_X + q)$$

und

$$BIC_{ML} = -2 \log L(\hat{\theta}) + (p_X + q) \log(n).$$

Dabei entsprechen p_X dem Rang der Versuchsplanmatrix X der festen Effekte, q der Anzahl zu schätzender Varianzkomponenten bzw. Parameter der räumlichen Funktion und n der Anzahl Beobachtungen bei lediglich einer Beobachtung pro Objekt bzw. der Anzahl Objekte bei mehreren Beobachtungen pro Objekt.

Der Vergleich der Kovarianzstruktur bei identischer Erwartungswertstruktur erfolgt durch Nutzung der REML-Methode gemäß:

$$AIC_{REML} = -2 \log L(\hat{\theta}) + 2q$$

und

$$BIC_{REML} = -2 \log L(\hat{\theta}) + q \log(n^*).$$

- (*n: - falls „random“ : Anzahl Stufen des ersten zufälligen Effekts,
- falls „repeated“: Anzahl Stufen des ersten Subjekts)

Die Berechnungsvorschriften der Informationskriterien sind so angegeben, dass das Modell mit dem vergleichsweise kleineren Wert für das Informationskriterium zu bevorzugen ist.

Auf eine Angabe des für kleine Stichprobenumfänge von HURVICH and TSAI (1989) vorgeschlagenen korrigierten Akaike-Kriteriums AICC wurde verzichtet, da bei den im vorliegenden Fall vorliegendem großen Stichprobenumfang zwischen AIC und AICC keine Unterschiede bestehen.

3.3 Residuenanalyse zur Überprüfung der Modellierungsergebnisse

Ein wichtiges Hilfsmittel zur Überprüfung der Modellierungsergebnisse stellt die Nutzung der Residuenanalyse dar (NOBRE and SINGER, 2007). Dabei können die aus einer OLS-Analyse resultierenden Residuen sowohl zur Überprüfung der Erwartungswert- als auch der Kovarianzstruktur verwendet werden. Bei räumlich erfassten Beobachtungen kann die Residuenanalyse besonders vorteilhaft eingesetzt werden, da dann in Abhängigkeit der räumlichen Position der Beobachtung ein Schätzwert für Erwartungswert und Varianz sowie Residuen zu deren Überprüfung zur Verfügung stehen.

Die Überprüfung der Erwartungswertstruktur erfolgt dabei mit dem Ziel, eine mögliche durch Nichtberücksichtigung fester Effekte vorliegende systematische Verzerrung der Residuen aufzudecken. Falls alle bedeutsamen festen Effekte beachtet wurden, sollte das Mittel der Residuen je Messtermin etwa Null sein und keine Änderung in Abhängigkeit räumlicher Koordinaten aufweisen. Das kann wiederum bei Nutzung einer lokal angepassten Regression überprüft werden (CLEVELAND et al., 1988; CLEVELAND and GROSSE, 1991).

Die OLS-Residuen dienen weiterhin der Überprüfung der Kovarianzstruktur. So kann die aus den OLS-Residuen geschätzte räumliche Korrelation der geschätzten Modell-Korrelation gegenüber gestellt werden.

4 Das Demonstrationsbeispiel

4.1 Versuchsplan und Datengrundlage

Die Vorgehensweise zur Modellwahl soll anhand eines Beispiels demonstriert werden, das wie folgt charakterisiert werden kann:

In einem OFE sind verschiedene Düngungsstrategien in ihrer Wirkung auf den Ertrag bei Winterweizen unter den Standortbedingungen des südlichen Sachsen-Anhalt zu vergleichen. Es werden drei Düngungsstrategien untersucht. Die räumliche Anordnung des Versuchs ist in Abbildung 1 skizziert.

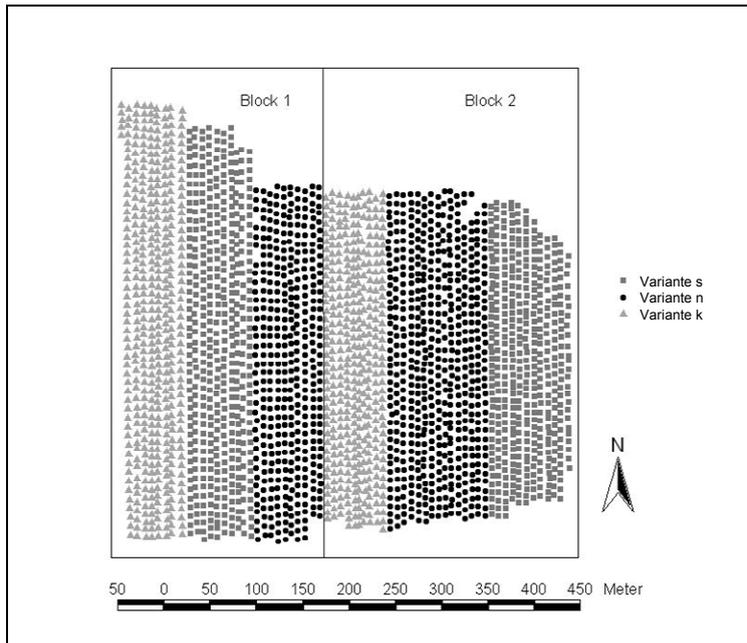


Abbildung 1: Lageplan für das Anwendungsbeispiel

Bei einer Ausdehnung des Versuchs von 444 x 333 m liegen je Stufe des Prüffaktors zwei Parzellen vor, die zwei Blocks zuzuordnen sind. Die statistischen Maßzahlen des Merkmal Ertrag sind in Tabelle 1 dargestellt. Neben dem Ertrag liegt für jeden Messwert im Merkmal Ertrag eine Messwert für die elektrische Leitfähigkeit des Bodens (EC25) vor.

Tabelle 1: Statistische Maßzahlen für das Anwendungsbeispiel ($N=2614$)

Merkmal	\bar{y}	s_y	\min_y	\max_y
Ertrag (dt ha ⁻¹)	99.5	8.03	64.6	124.4
Leitfähigkeit (EC25)	63.8	10.22	43.5	95.6

4.2 Auswertung bei Nutzung von Parzellenmittelwerten

Wie aus Abbildung 1 deutlich wird, entspricht der Versuch einer vollständigen Blockanlage mit zwei Parzellen je Prüfglied und mehr als einer Beobachtung je Parzelle. Die Zuordnung der Prüfglieder zu den Parzellen erfolgte zufällig. Damit liegt eine randomisierte Versuchsanlage vor. Basierend auf diese Anlage ist zunächst eine Auswertung bei Nutzung der Parzellenmittel möglich. Dabei erwies sich die Einbeziehung der elektrischen Leitfähigkeit auf der Ebene der Parzellenmittelwerte nach den in Abschnitt 3 dargestellten Methoden nicht als vorteilhaft. Daher wurde dieses Merkmal nicht als Kovariable in das Auswertungsmodell einbezogen.

Bei der Auswertung werden die Parzellenmittelwerte \bar{z}_{ij} als Realisation einer Zufallsvariable \bar{Z}_{ij} angesehen, für die das folgende Modell gilt:

$$\bar{Z}_{ij} = \mu + \alpha_i + bl_j + e_{ij}$$

wenn bedeuten:

μ = allgemeines Mittel

α_i = Effekt des i-ten Variante (fest)

bl_j = Effekt des j-ten Blocks (fest)

e_{ij} = zufälliger Resteffekt $e_{ij} \sim N(0, \sigma_e^2)$, $Cov(e_{ij}, e_{i',j'})$ $i, i' = 1, \dots, a$; $j, j' = 1, \dots, b$

Die Verwendung der Parzellenmittel für den Ertrag führt zu den in Tabelle 2 dargestellten Least Squares Means (LSMeans) und deren Differenzen. Für die Berechnung wurde SAS Proc Mixed in der nachfolgenden Form verwendet.

```
PROC MIXED DATA=parz_mean;
  CLASS variante block;
  MODEL mean_ertrag = variante block;
  LSMEANS variante /PDIFF CL;
RUN;
```

Tabelle 2: LSMeans, deren Differenzen, Standardfehler (SE), Freiheitsgrade (FG) und Konfidenzintervalle ($P=0.95$) bei Nutzung der Parzellenmittel

Variante	LSMean	SE	FG	Konfidenzintervall
k	94.3	1.02	2	[89.9;98.6]
n	101.3	1.02	2	[96.9;105.7]
s	102.2	1.02	2	[97.8;106.5]
Differenz				
k-n	-7.0	1.44	2	[-13.2;-0.8]
k-s	-7.9	1.44	2	[-14.1;-1.7]
n-s	-0.9	1.44	2	[-7.1;5.3]

Die Ergebnisse zeigen eine bedeutsame Differenzierung zwischen den Varianten, insbesondere auch bei Beachtung der Standardfehler. Das schlägt sich in den Konfidenzintervallen der Differenzen nieder, die für die Vergleiche k-n und k-s die Null nicht einschließen. Das entspricht einer Signifikanz für diese Vergleiche bei Anwendung des t-Tests. Allerdings stehen für die Konfidenzintervalle der LSMeans und deren Differenzen nur zwei Freiheitsgrade zur Verfügung. Das ist trotz der 2614 Beobachtungen zwangsläufig so, da in die Auswertung nur sechs Parzellenmittel eingegangen sind. Andererseits fordert das aber auch dazu heraus, ein Auswertungsmodell bei Zugrundelegung der Einzelwerte zu erarbeiten.

4.3 Auswertung bei Nutzung von Einzelwerten

4.3.1 Überlegungen zu Störfaktoren im Anwendungsbeispiel

Während bei Nutzung der Parzellenmittelwerte unter Zugrundelegung einer Versuchsanlage die Prüf- und Designfaktoren und damit das Auswertungsmodell im Wesentlichen festgelegt sind, ist das Auswertungsmodell bei der Nutzung von Einzelwerten in OFE keinesfalls gegeben.

Neben dem schon für die Parzellenmittelwerte verwendeten Prüffaktor "Düngungsvariante" und dem Designfaktor "Block" ist die Bedeutung weiterer Faktoren für die Wahl eines den Daten adäquaten Auswertungsmodells zu überprüfen. Die Festlegung der zu untersuchenden "Kandidatenfaktoren" ergibt sich aus sachlogischen Überlegungen. Im vorliegenden Fall sind das die folgenden festen Einflussfaktoren:

- Mähdrescher:
 - qualitativer Einflussfaktor;
 - die Ertragswerte wurden von drei Mähdreschern erfasst, die Mähdrescher arbeiteten auf allen Parzellen;
 - es ist zu prüfen, ob es zwischen Mähdreschern systematische Unterschiede in der Ertrags erfassung gibt;
 - weiterführende Informationen über die Druschspuren liegen nicht vor.
- Leitfähigkeit:
 - quantitativer Einflussfaktor:
 - es ist zu prüfen, ob auf der Ebene der Einzelwerte die Beachtung möglicher Bodenunterschiede bei Nutzung der Leitfähigkeit erforderlich ist.
- räumlicher Trend:
 - quantitativer Einflussfaktor;
 - basierend auf den zu jedem Ertragswert vorliegenden räumlichen Koordinaten ist zu prüfen, ob ein bedeutsamer räumlicher Trend vorliegt.

Eine weitere Aufgabe besteht in der Festlegung der Kovarianzstruktur. Die vorliegenden Daten sind durch eine Erfassung im Raum gekennzeichnet. Dabei ist zu erwarten, dass räumlich bedingte Abhängigkeiten zwischen den Beobachtungen vorliegen. Daher ist die Abhängigkeitsstruktur zu analysieren und zu prüfen, durch welche Kovarianzstruktur sie abgebildet werden kann.

Die weitere Vorgehensweise gliedert sich wie in Abschnitt 3 beschrieben.

4.3.2 Auswahl der Erwartungswertstruktur

Ausgangspunkt ist ein Modell mit dem Prüffaktor Variante und Designfaktor Block. Darauf aufbauend werden weitere Einflussgrößen entsprechend der sachlogischen Vorüberlegungen betrachtet. Für die quantitativen Einflussfaktoren ergeben sich dabei verschiedene Möglichkeiten der Modellbildung, beispielsweise durch eine lineare und zusätzlich quadratische Berücksichtigung der Kovariable. Eine Zusammenstellung ist in Tabelle 3 gegeben.

Tabelle 3: Anzahl Effekte im Modell, mit -2 multiplizierte Likelihoodfunktion, Prüfgröße des LRT und Differenzen von AIC und BIC zu Modell 7

Nr	feste Effekte	p (p _x)	-2 logL	$\Delta(-2 \log L)$ (Modell 7)	FG	$\Delta(AIC)$ (Modell 7)	$\Delta(BIC)$ (Modell 7)
1	variante block	5 (4)	17800.4	980.9	7	966.9	925.8
2	variante block md	7 (6)	17770.2	950.7	5	940.7	911.3
3	variante block md ec25	8 (7)	17661.5	842.0	4	834.0	810.6
4	variante block md ec25 ec25*ec25	9 (8)	17643.2	823.7	3	817.7	800.1
5	variante block md ec25 ec25*ec25 y	10 (9)	17206.3	386.8	2	382.8	371.0
6	variante block md ec25 ec25*ec25 y x	11 (10)	17187.6	368.1	1	366.1	360.2
7	variante block md ec25 ec25*ec25 y y*y x	12 (11)	16819.5	0		0	0

md = Mährescher, p_x = Rang der Designmatrix der festen Effekte, q = Anzahl der zufälligen Effekte, p = p_x+q, FG = Freiheitsgrade für LRT

Die Ergebnisse zeigen, dass Modell 7 allen anderen Kandidatenmodellen überlegen ist. Dabei führen LRT, AIC und BIC zu den selben Aussagen.

Die zwingende Notwendigkeit der Verbindung von LRT, AIC und BIC mit einer Analyse der OLS-Residuen kann sehr gut durch das vorliegende Beispiel demonstriert werden. So führt bereits Modell 6 zu einer bedeutsamen Verbesserung der angeführten Kriterien. Eine auf Modell 6 basierende Analyse der OLS-Residuen zeigt in y-Richtung das in Abbildung 2 dargestellte Bild. Es zeigt sich, dass noch ein bedeutsamer nichtlineare Trend der Residuen in dieser Richtung vorliegt.

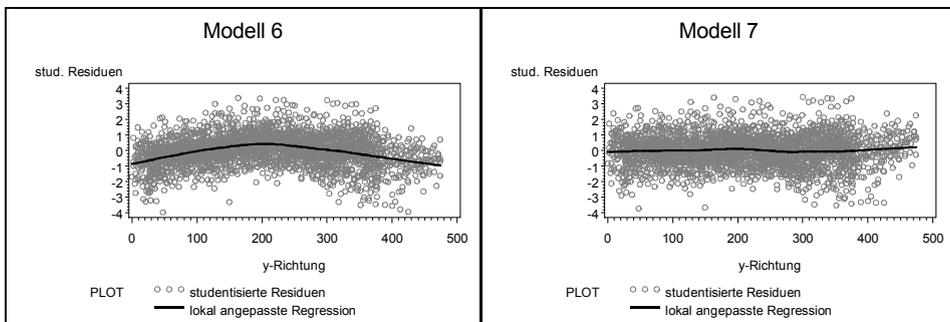


Abbildung 2: Studentisierte OLS-Residuen und deren lokal angepasste Regression in y-Richtung für Modell 6 und 7

Die Modellerweiterung um y^*y zu Modell 7 zeigt für die Residuen keinen Trend mehr (Abbildung 2). Das gilt ebenso auch für die hier nicht dargestellte X-Richtung. Entsprechend wird mit Modell 7 die Wahl der Erwartungswertstruktur zunächst abgeschlossen.

4.3.3 Auswahl der Kovarianzstruktur

Die Wahl des Auswertungsmodells wäre mit den vorherigen Arbeitsschritten nur für den Fall abgeschlossen, dass die Residuen als unabhängig angesehen werden können. Das ist im vorliegenden Fall aber nicht zu erwarten, da eine räumlich bedingte Abhängigkeit bedeutsam sein kann.

Ausgangspunkt der weiteren Analyse sind die OLS-Residuen für Modell 7 bei Nutzung der REML-Methode. Das Histogramm der Residuen (Abbildung 3) zeigt eine symmetrische Verteilung mit guter Annäherung an die Normalverteilung.

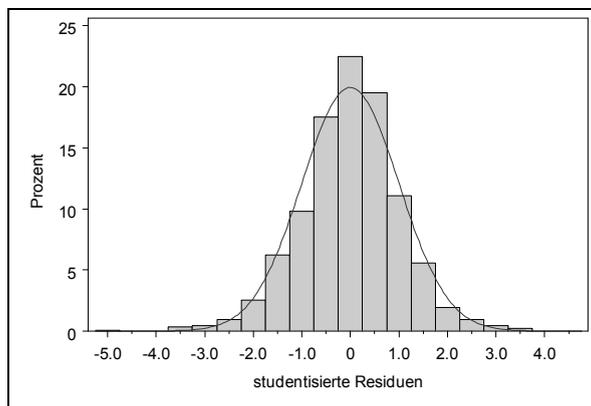


Abbildung 3: Histogramm der studentisierten OLS-Residuen für Modell 7

Da für die vorliegenden Residuen eine räumlich bedingte Abhängigkeit vermutet wird, wird die empirische räumliche Korrelation der Residuen berechnet. Die räumlich bedingten Varianzen und Kovarianzen können bei Nutzung von Proc Variogramm gewonnen werden. Der in Abbildung 4 dargestellte Verlauf zeigt eine ausgeprägte Abhängigkeit der räumlichen Korrelation sowohl vom Block als auch von der Richtung. Beachtenswert ist vor allem eine deutliche Anisotropie der Korrelation. Diese Vorinformation muss bei der Aufstellung von Kandidatenmodellen zur Optimierung der Kovarianzstruktur ausgenutzt werden.

Die in Abbildung 4 dargestellten Ergebnisse unterstreichen die erwartete Notwendigkeit einer weiterführenden Modellierung der Kovarianzstruktur. Eine Vernachlässigung dieses Sachverhaltes würden von 2614 unabhängigen Beobachtungen ausgehen. Bei weiterführenden Inferenzaussagen wären wegen zu geringer Schätzfehler und zu hoher Freiheitsgrade Unterschreitungen des nominalen Konfidenzniveaus und Überschreitungen des nominalen Fehlers 1. Art zwangsläufig die Folge.

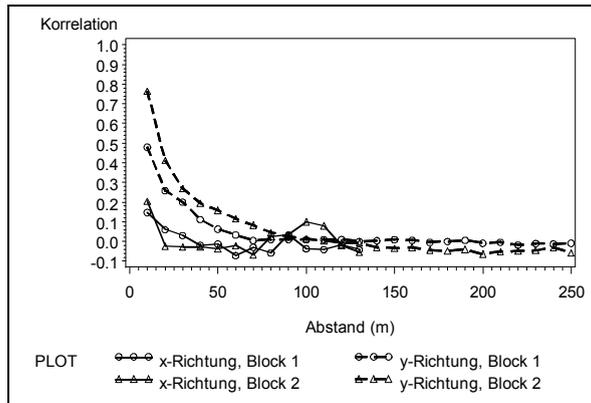


Abbildung 4: Empirische räumliche Korrelation der OLS-Residuen in Abhängigkeit von Abstand, Richtung und Block

Ausgewählte Ergebnisse der Modellwahl für die Kovarianzstruktur sind in Tabelle 4 zusammengestellt. Es wurden neben den in Tabelle 4 angegebenen Ergebnissen für die anisotrope Powerfunktion (TYPE=SP(POWA)) noch andere räumliche Funktionen untersucht, die jedoch wegen Konvergenzproblemen zu keinem Ergebnis führen oder schlechtere Werte der Informationskriterien aufweisen. Gleichmaßen werden die Ergebnisse für den isotropen Fall nicht mitgeteilt.

Die Ergebnisse zur Wahl der Kovarianzstruktur zeigen, dass die Beachtung räumlicher Abhängigkeiten zu einer deutlichen Verbesserung der Informationskriterien führt. Dieses Ergebnis war allerdings auch bei Beachtung der in Abbildung 4 dargestellten Korrelationen zu erwarten. Die Beachtung der räumlichen Abhängigkeiten in der Ausdehnung Versuch (SUB=intercept), Block (SUB=block) und Parzelle (SUB=parzelle) zeigt für den Bezugsrahmen Block die günstigsten Werte. Abgeleitet aus den in Abbildung 4 dargestellten empirischen Korrelationen wurden zusätzlich noch heterogene Korrelationen zwischen den Blocks berücksichtigt (Modell 13). Bei Nutzung dieses Modellansatzes und der zusätzliche Beachtung eines Parzelleneffektes ergab sich jedoch keine Lösung (Modell 12).

Tabelle 4: Anzahl der Varianzkomponenten bzw. Parameter im räumlichen Modell, mit -2 multiplizierte restringierte Likelihoodfunktion sowie Differenzen von AIC und BIC (feste Effekte gemäß Modell 7)

Nr	zufällige Effekte	q	-2 RlogL	AIC	$\Delta(AIC)$ (Modell 13)	BIC	$\Delta(BIC)$ (Modell 13)
8	Rest	1	16888.0	16890.0	1922.7	16895.9	1937.7
9	RANDOM parzelle Rest	2	16817.7	16820.7	1853.4	16818.1	1859.9
10	RANDOM parzelle REPEATED / SUB=intercept TYPE=SP(POWA) (x y) LOCAL	5	14969.8	14979.8	12.5	14975.3	17.1
11	RANDOM parzelle REPEATED / SUB=block TYPE=SP(POWA) (x y) LOCAL	5	14974.0	14984.0	16.7	14977.4	19.2
12	RANDOM parzelle REPEATED / SUB=block GROUP=block TYPE=SP(POWA) (x y) LOCAL				keine Lösung		
13	REPEATED / SUB=block GROUP=block TYPE=SP(POWA) (x y) LOCAL	7	14953.3	14967.3	0	14958.2	0
14	RANDOM parzelle REPEATED / SUB=parzelle TYPE=SP(POWA) (x y) LOCAL				keine Lösung		
15	REPEATED / SUB=parzelle TYPE=SP(POWA) (x y) LOCAL	4	14981.4	14989.4	22.1	14988.6	30.4

q = Anzahl der zufälligen Effekte

Die Schätzwerte der Varianzen bzw. Parameter der räumlichen Funktion erlauben eine Berechnung der geschätzten Korrelation, dargestellt in Abbildung 5.

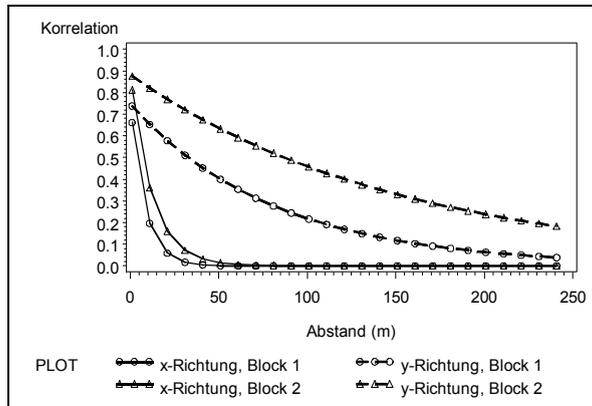


Abbildung 5: Geschätzte räumliche Korrelation in Abhängigkeit von Abstand, Richtung und Block (Modell 13)

Der Vergleich der Abbildungen 4 und 5 zeigt, dass die geschätzte Korrelation den Trend der empirischen Werte gut widerspiegelt. Dabei weisen die empirischen Werte einen deutlicheren Abfall der Korrelation in Abhängigkeit des Abstandes auf, als das mit dem geschätzten Funktionsverlauf abgebildet werden kann. Diese Unterschiede müssen für den vorliegenden Datensatz in Kauf genommen werden, da mit keinem verfügbaren räumlichen Modell eine bessere Abbildung gelingt.

4.3.4 Endgültige Festlegung der Erwartungswertstruktur

Abschließend ist die im 1. Schritt ermittelte Erwartungswertstruktur nochmals zu überprüfen. Für die vorliegende Auswertung zeigt sich, dass bei Nutzung von Modell 17 für den räumlichen Trend in x-Richtung kein Einfluss mehr nachweisbar ist ($P = 0.54$).

Bei der Auswertung wird die i -te Beobachtung von Variante i , erfasst in Block j mit Mährescher k , den räumlichen Koordinaten (x, y) und der elektrischen Leitfähigkeit $Ec25$, $z_{ijkl}(x, y; Ec25)$, als Realisation einer Zufallsvariable $Z_{ijkl}(x, y; Ec25)$ angesehen, für die das folgende Modell gilt:

$$Z_{ijkl}(x, y; Ec25) = \mu + \alpha_i + bl_j + md_k + \beta_1 Ec25 + \beta_2 Ec25^2 + \beta_3 y + \beta_4 y^2 + e_{ijkl}(x, y)$$

wenn bedeuten:

μ = allgemeines Mittel

α_i = Effekt des i -ten Prüfgliebes (fest)

bl_j = Effekt des j -ten Blocks (fest)

md_k = Effekt des k -ten Mähreschers (fest)

β_1, \dots, β_4 = feste Regressionskoeffizienten, zugehörig zu den Kovariablen $Ec25$, $Ec25^2$, y und y^2

$e_{ijkl}(x, y)$ = zufälliger Resteffekt.

Für den Resteffekt gelten die folgenden Verteilungsannahmen:

$$\text{Var}(e_{ijkl}(x, y)) = \sigma_{0,j}^2 + \sigma_e^2$$

$$\text{Cov}(e_{ijkl}(x, y), e_{ijkl'}(x', y')) = \sigma_{0,j}^2 \cdot \rho_{x,j}^{|x-x'|} \cdot \rho_{y,j}^{|y-y'|}$$

Dabei sind $\sigma_{0,j}^2$, $\rho_{x,j}$ und $\rho_{y,j}$ die Parameter der anisotropen Powerfunktion. Der Index j symbolisiert die Abhängigkeit der räumlichen Parameter von Block j .

Diese Modell entspricht in SAS Proc Mixed der Form:

```
PROC MIXED DATA=daten METHOD=REML;
  CLASS variante md block;
  MODEL ertrag = variante md block y y*y ec25 ec25*ec25 /
    DDFM=KR (FIRSTORDER);
  REPEATED /
    SUBJECT=BLOCK GROUP=block LOCAL TYPE=SP(POWA) (x y);
  LSMEANS variante / PDIFF CL;
RUN;
```

Zur Rückversicherung wurde eine nochmalige Residuenanalyse durchgeführt. Dabei zeigt sich, dass kein Trend der Residuen nachweisbar ist.

Bei der Formulierung des Auswertungsmodells ist die Festlegung der Freiheitsgradapproximation bedeutsam. Dabei wird mit der Einstellung DDFM = KR (FIRSTORDER) die Satterthwaite-Approximation der Freiheitsgrade bei zusätzlicher Modifikation der Standardfehler der geschätzten Effekte nach KENWARD and ROGER (2009) angewiesen. Durch die Option (FIRSTORDER) wird eine im Vorschlag von KENWARD and ROGER (1997) nicht beachtete Korrekturkomponenten berücksichtigt. Deren Nichtbeachtung führte bei komplexen Kovarianzstrukturen, wie beispielsweise bei räumlichen Kovarianzen, noch zu bedeutsamen Überschätzungen der Standardfehler (HU et al., 2006; SPILKE et al., 2010).

4.3.5 Auswertungsergebnisse bei Nutzung der Einzelwerte

Die bei Nutzung des Auswertungsmodells erhaltenen LSMeans der Varianten und deren Differenzen sind in Tabelle 5 zusammengestellt.

Ein Vergleich mit den aus den Parzellenmittelwerten resultierenden Schätzungen zeigt bedeutende Unterschiede insbesondere für Variante s. Die Unterschiede sind vor allem in der Korrektur von Bodentrend und elektrischer Leitfähigkeit begründet. Eine solche Korrektur wäre bei Nutzung der Parzellenmittel wegen der geringen Anzahl der Messwerte und damit verfügbaren Freiheitsgrade nicht möglich. Im vorliegenden Fall würde schon die Anzahl der Freiheitsgrade eine Beachtung weiterer Einflussgrößen wie beispielsweise Kovariablen verbieten. Somit bieten die Einzelwerte eine erhöhte Flexibilität bei der Modellformulierung für feste und zufällige Störgrößen. Die im vorliegenden Fall zwingend erforderliche Beachtung der räumlichen Kovarianzen berücksichtigt im Auswertungsmodell, dass keine 2614 unabhängigen Beobachtungen vorliegen. Das führt zwangsläufig zu einer Erhöhung der Schätzfehler der LSMeans und deren Differenzen. Entsprechend dürfen bei den interessierenden Inferenzaussagen die Restfreiheitsgrade nicht verwendet werden. Das wäre eine zu optimistische Annahme. Zur begründeten Bestimmung der Freiheitsgrade der Prüfgrößen bietet sich der methodische Ansatz von KENWARD and ROGER (2009) an. Diese Freiheitsgradberechnung muss aber

wie oben beschrieben im Model-Statement von Proc Mixed angewiesen werden, ansonsten wird diese Approximation nicht berechnet.

Tabelle 5: LSMMeans, deren Differenzen, Standardfehler (SE), Freiheitsgrade (FG) und Konfidenzintervalle ($P=0.95$) bei Nutzung der Einzelwerte

Variante	LSMean	SE	FG	Konfidenzintervall
k	95.3	1.30	40.6	[92.6;97.9]
n	100.6	1.21	38.9	[97.6;102.5]
s	100.1	1.22	40.8	[95.6;102.6]
Differenz				
k-n	-4.78	1.55	34.6	[-7.9;-1.6]
k-s	-4.83	1.55	33.2	[-7.9;-1.8]
n-s	-0.06	1.44	40.1	[-3.0;2.8]

Ein Vergleich der Intervallbreiten der Schätzungen der LSMMeans und deren Differenzen bei Nutzung der Parzellenmittel gegenüber den Einzelwerten zeigt, dass im vorliegenden Anwendungsfall die Nutzung der Einzelwerte zu einer bedeutsamen Reduzierung der Intervallbreiten um etwa 50 % führt. Das gleiche Verhältnis gilt entsprechend auch für die Grenzdifferenzen. Es ist jedoch zu betonen, dass diese Verhältnisse vom Anwendungsfall abhängen und nicht verallgemeinert werden können.

5 Abschließende Bemerkungen

Modellwahl im biometrischen Sinn bedeutet die Festlegung eines Auswertungsmodells bezüglich der festen und zufälligen Effekte sowie deren Kovarianzstruktur. Dabei darf der Erfolg der Modellwahl nicht an der Signifikanz der Stufen der Prüffaktoren gemessen werden, sonst wäre ein Modell mit ausschließlich festen Effekten immer zu bevorzugen. Der Erfolg der Modellwahl ist an der Identifizierung eines den Daten adäquaten Auswertungsmodells zu messen – das kann mit einer Erhöhung der Standardfehler, Abnahme der Freiheitsgrade verbunden sein und kann auch zum Verlust von Signifikanz führen. Analytische Kriterien stellen dabei ein wertvolles Hilfsmittel dar. Eine Rückversicherung durch Analyse der Residuen und den Vergleich der Kovarianzfunktion, geschätzt aus den OLS-Residuen und aus den Schätzwerten der Modellparameter, sind dabei unverzichtbar. Im vorliegenden Anwendungsfall kann dieser Vergleich anschaulich durch die jeweiligen räumlichen Korrelationen ausgedrückt werden. Die Notwendigkeit der Rückversicherung folgt auch aus dem Sachverhalt, dass das wahre Modell stets unbekannt ist. Die Kandidatenmodelle werden aus sachlogischer Sicht festgelegt. Damit kann aber immer die Gefahr verbunden sein, wichtige Störgrößen zu übersehen oder nicht richtig abzubilden. Somit verbietet sich bei der Modellwahl auch jeder Automatismus. Gerade wegen der Vielzahl möglicher fester und zufälliger Einflussfaktoren über die bekannten Prüf- und Designfaktoren hinaus kommt der biometrischen Modellwahl bei OFE eine besondere Bedeutung zu.

Danksagung

Für die Überlassung der Daten des Anwendungsbeispiels sei P. Wagner und T. Chudy herzlich gedankt.

Literatur

- AKAIKE, H. (1969): Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* 21: 243-247.
- AKAIKE, H. (1974): A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.
- CLEVELAND, W. S.; DEVLIN, S. J.; GROSSE, E.: (1988) Regression by Local Fitting. *Journal of Econometrics* 37: 87-114.
- CLEVELAND, W. S.; GROSSE, E. (1991): Computational Methods for Local Regression. *Statistics and Computing* 1: 47-62.
- FAHRMEIR, L.; HAMERLE, A. (1984): *Multivariate statistische Verfahren*. Berlin, New York, Walter de Gruyter.
- FAL, A. H. T.; CORNELIUS, P. L. (1996): Approximate F-Tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments. *Journal of Statistical Computation and Simulation* 54: 363-378.
- GIESBRECHT, F. G.; BURNS, J. C. (1985): Two-stage analysis based on a mixed model: large-sample asymptotic theory and small-sample simulation results. *Biometrics* 41: 477-486.
- GREENE, W. H. (2003): *Econometric analysis*. New York University, Prentice Hall, Pearson Education International, Fifth edition.
- HU, X.; SPILKE, J.; RICHTER, C. (2006): The influence of spatial covariance on the type I error and the power of different evaluation models. *Biometrical Letters* 43: 19-37.
- HURVICH, C. M.; TSAI, C. L. (1989): Regression and time series model selection in small samples. *Biometrika* 76: 297-397.
- KENWARD, M. G.; ROGER, J. H. (1997): Small Sample Inference for fixed effects from restricted maximum likelihood. *Biometrics* 53: 983-997.
- KENWARD, M. G.; ROGER, J. H. (2009): An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics and Data Analysis* 53: 2583-2595.
- NGO, L.; BRAND, R. (1997): *Model Selection in Linear Mixed Effects Models Using SAS Proc Mixed*. SAS Users Group International (22) San Diego, California, March 16-19.
- NOBRE, J. S.; SINGER, J. M. (2007): Residual Analysis for Linear Mixed Models. *Biometrical Journal* 49: 863-875.
- SCHWARZ, G. (1978): Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.
- SEARLE, S. R. (1971): *Linear Models*. New York, John Wiley & Sons, Inc.
- SPILKE, J.; PIEPHO, H.-P.; HU, X. (2005): A simulation study on tests of hypotheses and confidence intervals for fixed effects in mixed models for blocked experiments with missing data. *Journal of Agricultural, Biological, and Environmental Statistics* 10: 374-389.
- SPILKE, J.; RICHTER, C.; PIEPHO, H.P. (2010): Model selection and its consequences for different split-plot designs with spatial covariance and trend. *Plant Breeding* 129: 590-598.
- WOLFINGER, R. D. (1993): Covariance structure selection in general mixed models. *Communications in Statistics – Simulation and Computation* 22: 1079-1106.

On-Farm-Experiment mit georeferenzierten Daten und teilflächenspezifischer Applikation

CHRISTEL RICHTER und HEINRICH THÖLE

Zusammenfassung

Im Rahmen eines Forschungsprojektes am Institut für Agrartechnik in Potsdam-Bornim wurden in den Jahren 2005 bis 2007 auf mehreren Praxisschlägen Versuche zum Vergleich einer konventionellen einheitlichen und einer teilflächenspezifischen N-Düngung zu Getreide durchgeführt (THÖLE, 2010). Beide Prüfglieder wurden standortabhängig mit unterschiedlichen Fragestellungen gewählt. Eine Auswertung als Versuchsserie kam daher nicht in Frage. Im Folgenden wird einer dieser Versuche beschrieben und seine Auswertung demonstriert.

1 Versuchsfrage

Auf einem Alluvialstandort mit Wintergerste wurde die 1. und 2. N-Gabe mit 70 und 45 kg N ha⁻¹ einheitlich für beide Prüfglieder gestaltet. Bei der 3. N-Gabe wurde zwischen konventioneller einheitlicher und teilflächenspezifischer N-Düngung differenziert. Die teilflächenspezifisch ausgebrachte Düngermenge wurde in Abhängigkeit von der Bewertung des Bestandes vorgenommen. Diese erfolgte mittels des Crop-Meters im Traktorfrontanbau (EHLERT et al., 2003), dessen Messprinzip auf dem eines physikalischen Pendels basiert (Abbildung 1). Sein Auslenkungswinkel Φ aus der Senkrechten ist mit der vorhandenen Biomasse hoch korreliert (HAMMEN, 2001; EHLERT et al., 2003). Dieser Winkel wurde online nach einem vorgegebenen Algorithmus in eine Düngerapplikation umgesetzt.



Abbildung 1: Ansicht des Crop-Meters (Foto: Adamek, 2005)

Ausgangspunkt für die Wahl des Applikationsalgorithmus an dem gegebenen Standort war die Fachhypothese, dass besser etablierte Bestände, d. h. Bestände mit größeren Auslenkungswinkeln, erhöhte N-Gaben zum Fahnblattstadium in höhere Erträge umsetzen können. Die einheitliche N-Gabe wurde mit 41 kg ha^{-1} gewählt, die teilflächenspezifische Gabe entsprechend dem folgenden Ansatz (siehe auch Abbildung 2):

$$\text{teilflächenspezifische Gabe (kg N ha}^{-1}\text{)} = \begin{cases} 41 & \text{wenn } \Phi < 15^\circ \\ 1.2\Phi + 23 & \text{wenn } 15^\circ \leq \Phi \leq 30^\circ \\ 59 & \text{wenn } \Phi > 30^\circ \end{cases}$$

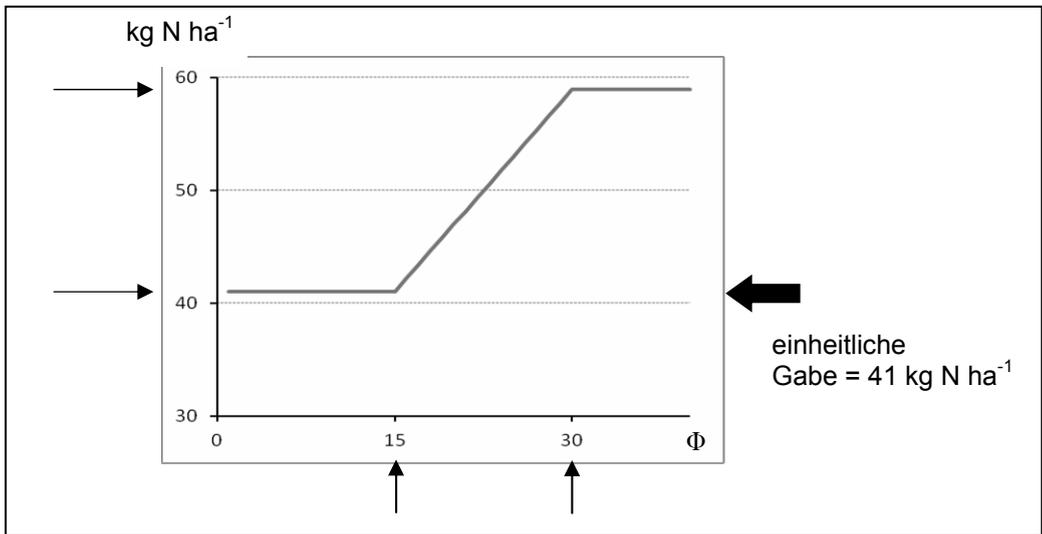


Abbildung 2: Definition der beiden Prüfglieder

Die mit diesem Einzelversuch zu beantwortende Versuchsfrage lautete demzufolge:
Führt der gewählte teilflächenspezifische Applikationsalgorithmus bei der gegebenen Sorte auf dem konkreten Feld zu höheren Kornerträgen als die einheitliche N-Düngung?

oder als statistische Hypothese formuliert:

$$H_0 : \mu_{\text{einh}} = \mu_{\text{teilfl}} \quad H_A : \mu_{\text{einh}} < \mu_{\text{teilfl}}$$

Da es sich um einen Einzelversuch handelte, ist der Aussagebereich beschränkt auf:

- den konkreten Applikationsalgorithmus unter Nutzung des Crop-Meters,
- den konkreten Schlag,
- die konkreten Witterungsbedingungen und
- alle weiteren Konstantfaktoren der Versuchsdurchführung.

2 Versuchsanlage

Der Versuch wurde als randomisierte Blockanlage angelegt, wobei jedes Prüfglied dreimal wiederholt wurde (insgesamt sechs Parzellen). Im klassischen Feldversuch werden die Blocks so gewählt, dass die Bedingungen innerhalb der Blocks möglichst homogen sind. Die Parzellen eines Blocks sollten dann die innerhalb des Blocks noch existierende Heterogenität in möglichst gleicher Weise widerspiegeln. Diese Herangehensweise garantiert, dass für alle Prüfglieder vergleichbare Ausgangsbedingungen vorhanden sind (*ceteris-paribus*-Prinzip). Hier wurde mit einem Sensoransatz (vgl. Leitfaden) gearbeitet, d. h. erst während der Applikation der Prüfglieder wird über die Winkelmessungen die Bestandesheterogenität erfasst. Somit konnte der Heterogenität bei der Planung der Versuchsanlage nur begrenzt nach visuellem Eindruck Rechnung getragen werden - die Richtungen von Blocks und Parzellen orientierten sich an den existierenden Fahrgassen. Auf Grund der fächerförmigen Streubilder der eingesetzten Zentrifugal-Düngerstreuers wurden für jede Parzelle zwei benachbarte Fahrgassen gewählt, da sich erst durch Überlappung die angestrebte gleichmäßige Düngerverteilung ergab. Die Parzellenbreite entsprach mit 60 m der zweifachen Arbeitsbreite des Düngerstreuers. Die Feldgeometrie bedingte, dass die Parzellen in Block 2 und 3 teilweise kleiner waren als in Block 1 (Abbildung 3).

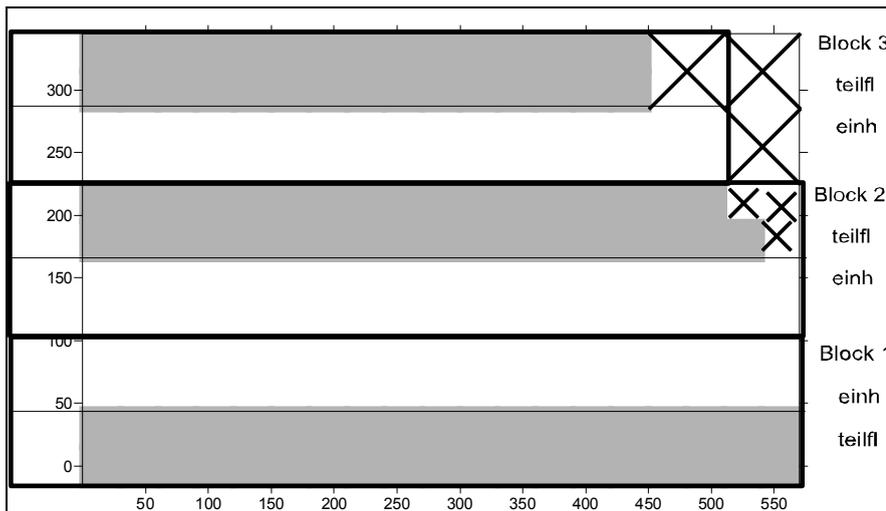


Abbildung 3: Anlage des Versuchs mit randomisierten Blocks

3 Erfasste Merkmale und deren Position

Alle Merkmale wurden je Parzelle mehrfach erfasst, wobei diese Mehrfacherfassungen keine echten Wiederholungen, sondern wiederholte Messungen innerhalb der jeweiligen Parzelle darstellen. Die Positionen der wiederholten Messungen wurden mittels GPS aufgezeichnet (Abbildung 4).

- Die *Auslenkungswinkel* des Crop-Meters wurden in den Fahrgassen im Abstand von ca. 2 sec (Sekunden) registriert. Obwohl der gemessene Winkel beim Prüfglied ‚einheitlich‘ für die ausgebrachte N-Menge keine Konsequenzen hatte, wurde er zwecks Kontrolle des Bestandeszustands auch dort erfasst. Für das Prüfglied ‚teilflächenspezifisch‘ wurde

die realisierte Düngerapplikation auf der Grundlage des Algorithmus aus dem gemessenen Winkel berechnet.

- Zwischen den Fahrgassen jedes Prüfglieds wurden je Parzelle zwei volle Schneidwerksbreiten im Kerndrusch geerntet und *Kornfrischmasse und -feuchte* ca. alle 5 sec ermittelt. Die verbleibenden Flächen und das Vorgewende wurden als Ränder behandelt und gingen nicht in die Analyse ein.
- Die *scheinbare elektrische Leitfähigkeit* wurde in den Fahrgassen bestimmt und in Form von Mittelwerten aus ca. 1000 Werten sec^{-1} gespeichert.

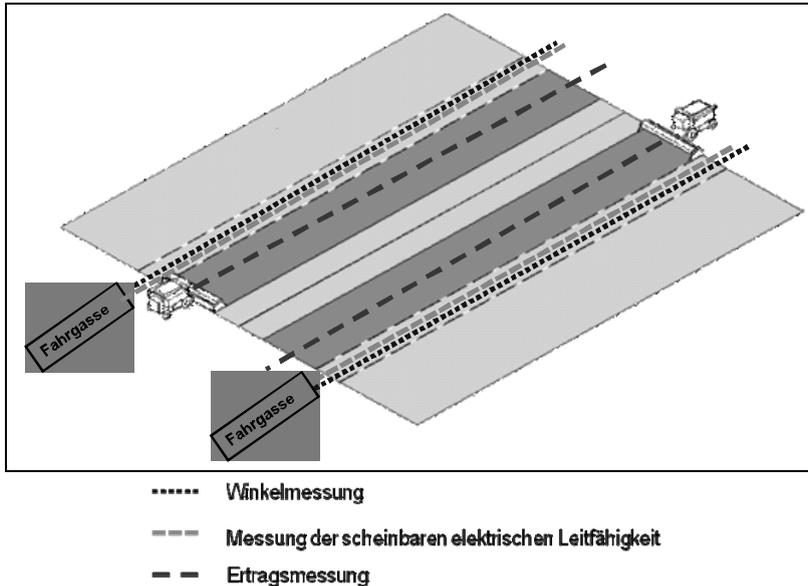


Abbildung 4: Darstellung der Merkmalerfassungen innerhalb einer Parzelle (modifiziert nach DAMMER, 2005)

4 Datenaufbereitung und Plausibilitätskontrolle

Vor der Analyse der Daten sollten diese auf Plausibilität überprüft und nicht plausible Daten gegebenenfalls von der weiteren Auswertung ausgeschlossen werden. Im vorliegenden Fall wurden u. a. folgende Kontrollen und Umrechnungen durchgeführt:

Winkelmessungen (Düngergaben)

- Berechnung der Fahrgeschwindigkeiten v (m s^{-1}),
- Entfernung aller Datensätze mit $v \leq 0$, sowie aller, wenn v nicht im Intervall $\bar{v} \pm 2s_v$ lag,
- Entfernung aller Datensätze aus dem Vorgewende,
- Kontrollwägungen je Fahrspur und Vergleich mit Düngerstreuer-Jobrechner.

Ertragsdaten

- Entfernung aller Datensätze mit Ertrags- und/oder Feuchtwerten = 0,
- Berechnung der Fahrgeschwindigkeiten v (m s^{-1}) und Entfernung aller Datensätze mit $v=0$,

- Nachkalibrierung der Erträge auf Basis Kontrollwägung je Kerndruschstreifen,
- Umrechnung auf 14 % Feuchte.

scheinbare elektrische Leitfähigkeit

- Umrechnung auf EC25 (25 °C Bodentemperatur).

Nach der Plausibilitätsanalyse ergaben sich folgende Stichprobenumfänge:

- Winkel (Düngung) N = 820,
- Ertrag N = 1028,
- EC25 N = 6360.

5 Aggregation und Zusammenführung von Merkmalsdaten

Dieser Schritt ist notwendig, wenn Zusammenhänge zwischen Applikations-, Ertrags- und potentiellen Kovariablen betrachtet werden sollen und diese an nicht gleichen Positionen und/oder in unterschiedlicher Intensität erfasst wurden. Im vorliegenden Beispiel liegen sowohl unterschiedliche Anzahlen der Merkmalswerte als auch unterschiedliche Positionen vor.

Für die Aggregation bieten sich zwei Herangehensweisen an:

1. Rasterbezug: Es werden gleichgroße Rasterzellen definiert und je Rasterzelle ein Mittelwert aus den darin liegenden Einzelwerten gebildet. Dieser Weg wurde für die Ertragsdaten und die Winkel gewählt. Dazu wurden Rasterzellen der Größe 30 m x 30 m gebildet, die sich an der Arbeitsbreite des Düngerstreuers orientierten. Im Ergebnis lag für beide Merkmale ein Mittelwert je Rasterzelle vor, dem die Koordinaten des Zellenmittelpunktes zugeordnet wurden.
2. Kriging: Das Kriging ist eine Methode zur Vorhersage von Merkmalsdaten an nicht beprobten Positionen (Punkt-Kriging) oder von Mittelwerten für nicht beprobte Flächeneinheiten (Block-Kriging). Grundlage für diese Methoden ist, dass ein Variogramm-Modell an das empirische Semivariogramm des entsprechenden Merkmals angepasst wird. Das Block-Kriging wurde für die EC25-Werte genutzt, wobei die EC25-Mittelwerte für die zuvor für die Ertrags- und Winkeldaten festgelegten Rasterzellen vorhergesagt wurden. Das Variogramm-Modell ist hierbei nicht bestmöglich an das empirische Semivariogramm über alle möglichen Distanzen auf der Versuchsfläche anzupassen, vielmehr ist eine besonders gute Anpassung für die kürzeren Distanzen anzustreben. Dieses resultiert daraus, dass der Vorhersagewert üblicherweise aus den Werten der 8 nächstliegenden Beobachtungspunkte berechnet wird. Die Modellanpassung wurde mit GSwin+ vorgenommen. Das sphärische Modell mit den Parametern: nugget = 0.01, range = 174.6 m und sill = 76.09 zeigte sich als besonders geeignet. Die Vorhersagefehler der so berechneten Rastermittelwerte werden durch die positionsabhängigen Kriging-Varianzen bestimmt. In den nachfolgenden Analysen wurde auf eine Wichtung der vorhergesagten EC25-Werte mit den Kriging-Varianzen verzichtet.

Nach der Aggregation lagen für alle Merkmale N = 212 Stichprobenwerte vor, wobei für die einheitliche N-Düngung $n_{\text{ein}} = 109$ und für die teilflächenspezifische $n_{\text{teil}} = 103$ war. Abbildung 5 zeigt die klassifizierte Merkmalsdaten ortsbezogen nach der Aggregation.

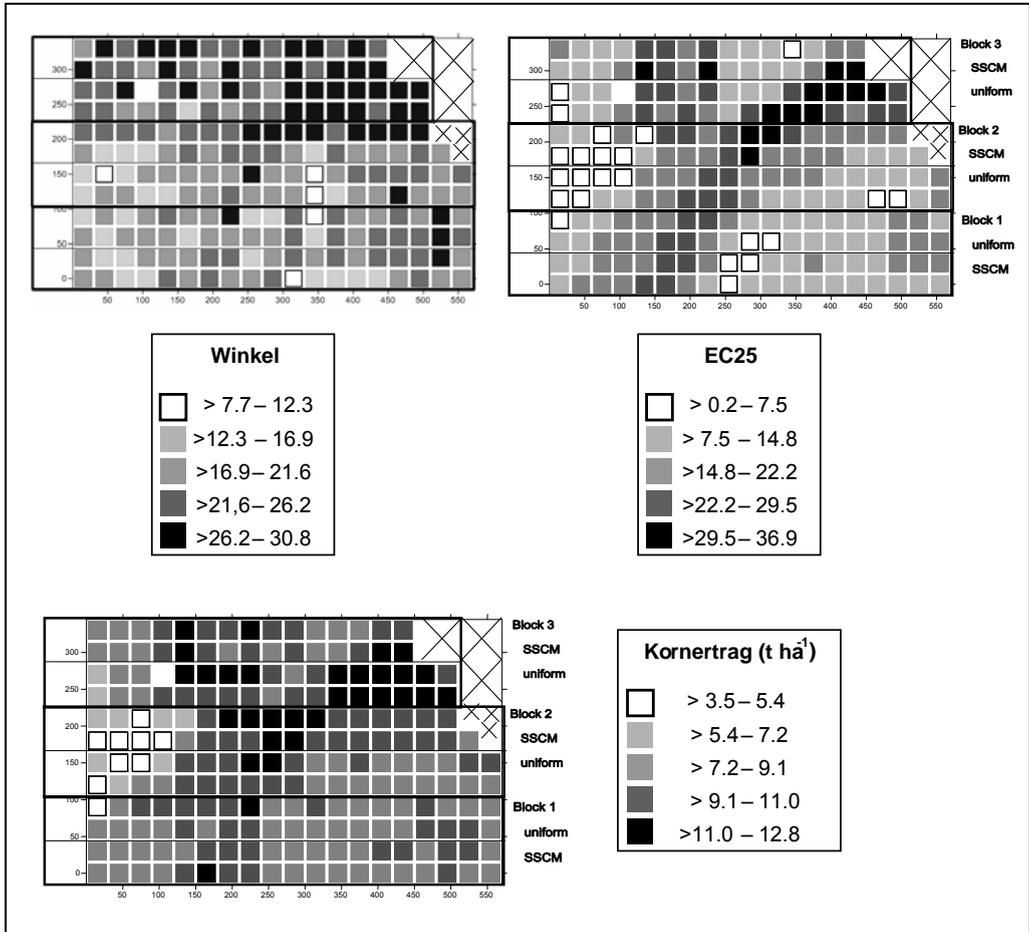


Abbildung 5: Lage der klassifizierten Merkmalsdaten nach Aggregation

In Abbildung 5 ist auffallend, dass die Ertragsdaten und die EC25-Werte ein ähnliches Muster auf der Fläche zeigen. Unabhängig vom Prüfglied tendieren höhere EC25-Werte zu höheren Kornertragsdaten.

6 Datenauswertung

Nach der Aggregation liegen die Daten in der Datei Gerste vor und enthalten die Variablen pg (Prüfglied mit den Kurzbezeichnungen ‚teilfl‘ und ‚einh‘), block, parzelle, ertrag, EC25 und winkel, sowie die Mittelpunktkoordinaten x und y von jeder Rasterzelle.

Zur Bewertung der Behandlungen werden mehrere Modellansätze demonstriert. Die Anpassungsgüte der verschiedenen Modellansätze wird durch das korrigierte Akaike-Informationskriterium (AICC) charakterisiert. Sämtliche Berechnungen erfolgten mit SAS Version 9.2 .

6.1 Mittelwertbildung über alle Einzelwerte je Parzelle und Auswertung als randomisierte Blockanlage

Es wurden je Parzelle die Mittelwerte der Ertragsdaten gebildet. Mit den erhaltenen sechs (2 Prüfglieder x 3 Blocks) Mittelwerten wurde eine klassische Varianzanalyse als randomisierte Blockanlage durchgeführt. Diese Auswertung ist gerechtfertigt, da der Versuch entsprechend angelegt wurde. Nur wenn der Versuch auf diese Weise auswertbar ist, erfüllt er die Grundprinzipien, die an die Planung eines OFE gestellt werden müssen.

Mittelwertberechnung für die Ertragsdaten je Prüfglied und Block:

```
PROC UNIVARIATE DATA = Gerste;
  VAR ertrag;
  BY pg block;
  OUTPUT OUT = mw MEAN = mwertrag;
RUN;
```

Varianzanalyse als randomisierte Blockanlage und Prüfgliedvergleich:

```
PROC MIXED DATA = mw METHOD = reml;
  CLASS pg block;
  MODEL mwertrag = pg block;
  LSMEANS pg / PDIF = controlu('einh');
RUN;
```

Man beachte, dass mit der Lsmeans-Anweisung und der Option pdiff standardmäßig ein zweiseitiger t-Test durchgeführt wird. Durch controlu('einh') wird die obige einseitige Alternativhypothese getestet. Die Ergebnisse sind in Tabelle 1 zusammengefasst.

Tabelle 1: Auswertungsergebnisse für den Ertrag nach 6.1

AICC _{REML}	Rest- varianz	Mittelwert		Differenz- teilfl-einh	s.e.d.	p-Wert	
		einh	teilfl			t-Test ¹⁾ Prüfglieder	F-Test Block
11.1	0.211	9.239	9.141	-0.0981	0.375	0.591	0.241

¹⁾ p-Wert für einseitigen t-Test

Als vorläufiges Fazit ergibt sich: Auf Grund des Tests kann die Nullhypothese nicht abgelehnt werden, d. h. ein signifikanter Mehrertrag kann für die teilflächenspezifische Variante nicht nachgewiesen werden.

6.2 Nutzung der Einzelwerte je Parzelle ohne Berücksichtigung der Georeferenzierung

Der folgende Auswertungsansatz berücksichtigt die wiederholten Messungen, nutzt aber nicht deren Georeferenzierung aus. Damit ist diese Vorgehensweise auch möglich, wenn keine Georeferenzierung vorliegt.

Die teilflächenspezifische Applikation hatte bei dem gewählten Applikationsalgorithmus nicht die Homogenisierung des Bestandes zum Ziel. Es ist vielmehr zu erwarten, dass beim Prüfglied ‚teilfl‘ eine stärkere Differenzierung des Ertrages als bei einheitlicher Düngung auftritt.

Aus diesem Grund werden die Ansätze mit und ohne Varianzhomogenität gegenübergestellt. Auch bei 6.1 wäre prinzipiell eine Prüfung der Varianzhomogenität denkbar, verbietet sich dort aber auf Grund der niedrigen Wiederholungsanzahl.

Ansatz bei Annahme von Varianzhomogenität:

```
PROC MIXED DATA = Gerste METHOD = reml;
  CLASS pg block parzelle;
  MODEL ertrag = pg block /DDFM = kr(firstorder);
  RANDOM parzelle;
  LSMEANS pg/ PDIFF = controlu('einh');
RUN;
```

Ansatz bei Annahme von Varianzheterogenität:

```
PROC MIXED DATA = Gerste METHOD = reml;
  CLASS pg block parzelle;
  MODEL ertrag = pg block / DDFM = kr(firstorder);
  RANDOM parzelle;
  REPEATED / GROUP = pg;
  LSMEANS pg/ PDIFF = controlu('einh');
RUN;
```

Tabelle 2: Anpassungskriterien und Varianzschätzungen für den Ertrag nach 6.2; fett: Modell mit besserer Anpassung

Varianz-homogenität	AICC _{REML}	Varianz Parzelle	Restvarianz	
			einh	teilfl
ja	824.6	0.1265	2.7662	
nein	824.4	0.1254	2.3668	3.1897

Auf Grund des kleineren AICC-Wertes ist der Ansatz mit Varianzheterogenität besser angepasst (Tabelle 2). Die Streuung zwischen den Parzellen charakterisiert gleichzeitig die durchschnittliche Kovarianz der wiederholten Messungen innerhalb der Parzellen (0.1265 bzw. 0.1254). Die eine bzw. zwei Restvarianzen beschreiben die durchschnittliche Streuung der wiederholten Messungen innerhalb der Parzellen. Im varianzheterogenen Ansatz ist, wie zuvor vermutet, die Restvarianz für das Prüfglied ‚teilfl‘ etwas größer.

Der auf den Parzellenmittelwerten basierende AICC-Wert (vgl. 6.1) ist mit den hier und im Folgenden aus den Einzelwerten berechneten AICC-Werten nicht vergleichbar, also daraus keine bessere oder schlechtere Anpassung ableitbar.

Tabelle 3: Prüfgliedvergleich für den Ertrag nach 6.2; fett: Modell mit besserer Anpassung

Varianz-homogenität	Mittelwert		Differenz teilfl - einh	s.e.d.	p-Wert	
	einh	teilfl			t-Test ¹⁾ Prüfglieder	F-Test Block
ja	9.2335	9.1476	-0.0859	0.3703	0.5807	0.2418
nein	9.2343	9.1486	-0.0857	0.3699	0.5805	0.2360

¹⁾ p-Wert für einseitigen t-Test

Interpretiert werden sollte das Testergebnis des besser angepassten Modells. Auch hier ergibt sich als vorläufiges Fazit (Tabelle 3): Auf Grund des Tests kann die Nullhypothese nicht abgelehnt werden, d. h. ein signifikanter Mehrertrag kann für die teilflächenspezifische Variante nicht nachgewiesen werden.

6.3 Zwischenschritt: Möglichkeit und Notwendigkeit der Berücksichtigung von Kovariablen

Die Diskussion der eventuellen Berücksichtigung von EC25, Auslenkungswinkel und weiterer Variablen als mögliche Kovariablen erfolgt unter verschiedenen Gesichtspunkten.

6.3.1 Kontrolle der Einhaltung des ceteris-paribus-Prinzips

Da vor Versuchsbeginn keine Kontrolle der einheitlichen Ausgangsbedingungen für die beiden Prüfglieder erfolgen konnte, kann hier nachträglich geprüft werden, ob diese hinsichtlich der EC25-Werte und des Bestandeszustands (Winkel) gegeben war. Dazu wird analog zu den SAS-Anweisungen aus 6.2 für diese beiden Merkmale vorgegangen. Der t-Test wird jedoch zweiseitig durchgeführt, da die Gleichheit der Bedingungen geprüft werden soll. Diese Überprüfung setzt nicht die exakte Kenntnis der Koordinaten der Beobachtungswerte voraus, jedoch sollten Winkel-, EC25- und Ertragswerte von den gleichen Positionen stammen.

Tabelle 4: Anpassungskriterien und Varianzschätzungen für die EC25- und Winkeldaten; fett: Modell mit besserer Anpassung

Merkmal	Varianz-homogenität	AICC _{REML}	Varianz Parzelle	Restvarianz	
				einfl	teilfl
Winkel	ja	1175.4	2.6211	14.7891	
	nein	1177.4	2.6216	15.0342	14.5292
EC25	ja	1449.9	7.6066	55.4699	
	nein	1451.8	7.6122	57.3897	53.4343

Tabelle 5: Prüfgliedvergleich für die EC25- und Winkeldaten; fett: Modell mit besserer Anpassung

Merkmal	Varianz-homogenität	Mittelwert		Differenz teilfl-einfl	s.e.d.	p-Wert	
		einfl	teilfl			t-Test ¹⁾ Prüfglieder	F-Test Block
Winkel	ja	21.5099	22.4935	0.9836	1.4244	0.5605	0.1545
	nein	21.5099	22.4935	0.9836	1.4244	0.5605	0.1546
EC25	ja	16.1082	15.9704	-0.1378	2.4753	0.9607	0.3954
	nein	16.1076	15.9695	-0.1380	2.4756	0.9607	0.3978

¹⁾ p-Wert für zweiseitigen t-Test

Aus den Tabellen 4 und 5 ist folgendes Fazit abzuleiten: Im Mittel sind die Winkel- und EC25-Werte für beide Prüfglieder zu Beginn des Versuchs identisch gewesen, auch die Variabilität ist bei beiden Prüfgliedern innerhalb der Parzellen als gleich zu betrachten. Damit bestanden zu Versuchsbeginn homogene Ausgangsbedingungen und die Ergebnisse aus 6.1 und 6.2 können interpretiert werden.

Wäre dieses nicht gegeben, so wäre bereits aus dieser Sicht zu prüfen, ob die entsprechenden Variablen als Kovariable in Betracht zu ziehen sind. Näheres wird in 6.3.2 erläutert.

6.3.2 Verbesserung der Modellanpassung und Erfüllung der Voraussetzungen für geostatistische Ansätze

Als Kovariablen kommen hier potentiell die Winkel- und EC25-Daten sowie Raumkoordinaten in Frage. Ihre mögliche Einbeziehung ist neben den in 6.3.1 erläuterten möglichen Gründen unter zwei Gesichtspunkten zu betrachten:

- (1) Wenn diese Merkmale einen Einfluss auf den Ertrag haben, dann könnte bei ihrer Berücksichtigung ein Teil der Variabilität des Ertrages erklärt und deren Einfluss eliminiert werden. Dabei würden die Ertragsmittelwerte standardmäßig auf das durchschnittliche Niveau der Kovariablen umgerechnet werden.
- (2) Der Ansatz eines geostatistischen Modells erfordert, dass in dem Bereich, in dem eine Korrelation unterstellt wird, kein globaler Trend vorliegen darf. Im vorliegenden Versuch heißt das, dass die Residuen des Ertrages nach Eliminierung der Prüfglied- und Blockeffekte keinen Trend zeigen sollten. In Abbildung 6 sind diese Residuen dargestellt. Ein Trend über die Versuchsfläche ist erkennbar und wie bereits in Abbildung 5 zu sehen war, scheint dieser insbesondere durch EC25 bedingt zu sein.

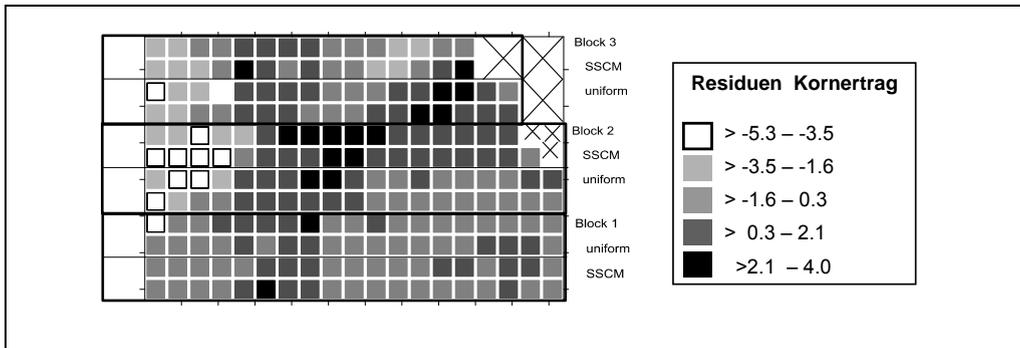


Abbildung 6: Lage der klassifizierten Residuen des Ertrages

Bei der Berücksichtigung von Kovariablen ist jedoch generell Vorsicht geboten. Es sollten keine Variablen in das Modell aufgenommen werden, die ursächlich mit der Definition der Prüfglieder zusammenhängen bzw. eine Eigenschaft der Prüfglieder sind. Damit entfällt der Winkel als Kovariable, da er das teilflächenspezifische Prüfglied definiert. Seine Einbeziehung käme einer Neudefinition des Prüfglieds gleich. Ob das Merkmal EC25 aus dieser Sicht als Kovariable sinnvoll ist, sollte in Abhängigkeit von seiner Wirkung geprüft werden.

Für das Prüfglied ‚teilfl‘ wäre folgende Ursache-Wirkungskette denkbar:

$$EC25 \rightarrow Winkel \sim \text{teilflächenspezifische Düngergabe} \rightarrow \text{Ertrag}$$

Auch für das Prüfglied ‚einh‘ wären die Ursache-Wirkungsbeziehungen $EC25 \rightarrow Winkel$ und $EC25 \rightarrow Ertrag$ möglich. Hier ist aber die Winkelmessung nicht an die Prüfglieddefinition gekoppelt. Würde sich für das Prüfglied ‚teilfl‘ die Abhängigkeit $EC25 \rightarrow Winkel$ bestätigen, wäre mit der Berücksichtigung von EC25 als Kovariable indirekt eine Korrektur des Prüfglieds verbunden. Somit soll zunächst für beide Prüfglieder die Beziehung $EC25 \rightarrow Winkel$ untersucht werden. Abbildung 7 zeigt, dass insbesondere für das Prüfglied ‚teilfl‘ die

Abhängigkeit schwach ausgeprägt ist, so dass nichts gegen die Einbeziehung von EC25 als Kovariable spricht. Aus Abbildung 8 folgt, dass ein quadratischer Ansatz für beide Prüfglieder sinnvoll erscheint.

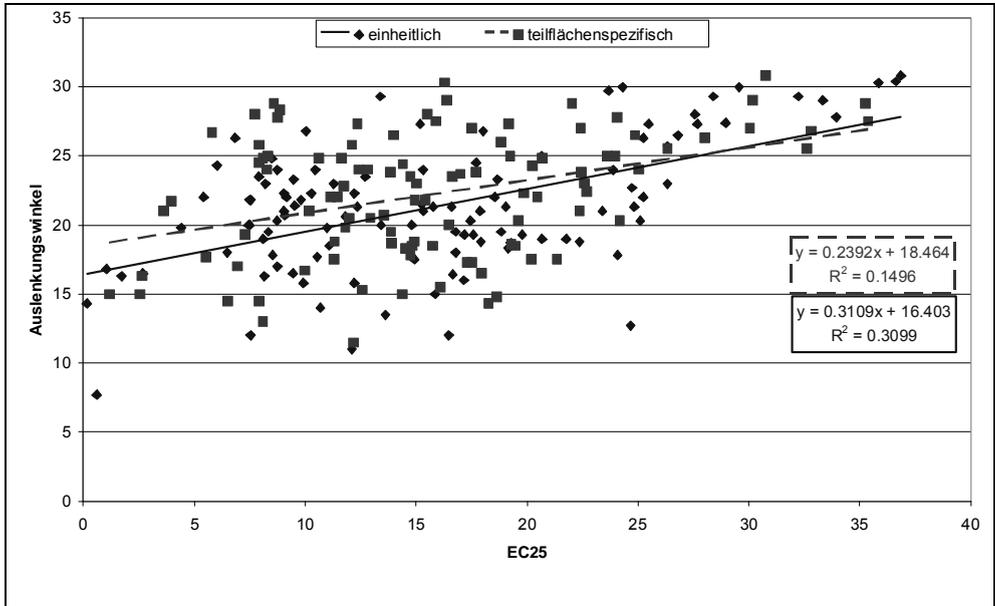


Abbildung 8: Abhängigkeit der Auslenkungswinkel von EC25 je Prüfglied

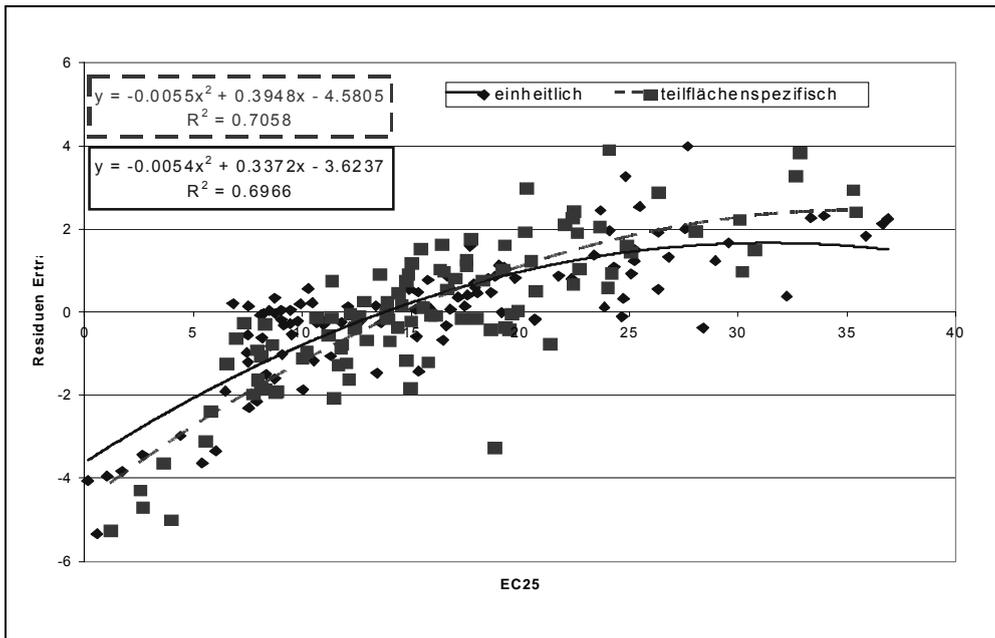


Abbildung 9: Abhängigkeit der Residuen des Ertrags von EC25 je Prüfglied

Werden EC25 und EC25² als Kovariablen betrachtet, so müssen deren Werte und die Ertragswerte von gleichen Positionen vorliegen, die konkreten Koordinaten sind nicht notwendig. Werden aber auch räumliche Koordinaten als Kovariable genutzt, so sind für alle Merkmalswerte die Positionsangaben (hier: Rasterzellenmittelpunkte) erforderlich. In 6.4.1 werden Vorgehen und Konsequenzen beschrieben.

6.4 Nutzung der Einzelwerte unter Berücksichtigung der Georeferenzierung

6.4.1 Einbeziehung von Kovariablen, ohne geostatistischen Ansatz

Neben dem linearen und quadratischen Ansatz für EC25 wurden auch die räumlichen Variablen x , x^2 , y , y^2 und $x*y$ als mögliche Kovariablen geprüft. Sollen letztere in das Modell aufgenommen werden, so ist die Georeferenzierung aller Merkmale notwendige Voraussetzung. Wird keine einheitliche, sondern eine prüfgedingspezifische Abhängigkeit des Ertrages von einer Kovariablen in Erwägung gezogen, beispielsweise des Ertrages von den EC25-Werten, so wäre dieses hier durch die Variablen $pg*EC25$ und $pg*EC25^2$ möglich. Dass dieses wahrscheinlich nicht notwendig ist, ist aber bereits aus Abbildung 8 erkennbar.

Sollen nunmehr Modelle mit verschiedenen Kovariablenanzahlen (fixe Einflussgrößen) untereinander verglichen werden, so sollte die Anpassungsgüte nicht an Hand des AICC-Wertes aus der REML-Methode beurteilt werden. Dieser enthält nur einen Strafterm für die Anzahl der Varianz-Kovarianz-Parameter. Stattdessen muss für die Festlegung der Kovariablen zur ML-Methode gewechselt werden, da der dabei angegebene AICC-Wert zusätzlich die Anzahl der Kovariablen bestraft (vgl. den Beitrag zur Wahl des Auswertungsmodells in diesem Band).

Bei der Anpassung verschiedener Kovariablenansätze zeigte sich, dass die Varianzheterogenität für den Ertrag nicht mehr bestätigt werden konnte, so dass nachfolgend nur noch der Ansatz mit Varianzhomogenität betrachtet wird. Im Folgenden ist beispielhaft gezeigt, wie ein Ansatz mit Kovariablen angewiesen werden kann. Dabei symbolisieren $xx = x^2$, $yy = y^2$, $xy = x*y$ und $ec = ec^2$. Diese Variablen sind vorher in einem Datenschnitt zu berechnen.

```
PROC MIXED DATA=Gerste METHOD = ml ;  
  CLASS pg block ;  
  MODEL ertrag = pg block x xx y yy xy ec25 ec pg*ec25 pg*ec /  
    DDFM=kr(firstorder) ;  
  RANDOM parzelle ;  
RUN ;
```

Folgende ausgewählte Ergebnisse wurden erhalten:

- | | |
|------------------------------------------------------|----------------------------|
| – ohne Kovariablen (entspricht Ansatz 6.2) | AICC _{ML} = 826.8 |
| – mit EC25 und EC25 ² | AICC _{ML} = 559.8 |
| – mit EC25, EC25 ² , x und x ² | AICC _{ML} = 492.4 |

Alle Regressionskoeffizienten der Kovariablen waren mit $p < 0.0001$ signifikant von Null verschieden. Die beste Anpassung von allen möglichen Kovariablenkombinationen zeigte der letzte Ansatz; Abbildung 9 zeigt die nunmehr erhaltenen Residuen. Der ursprüngliche Trend in Abbildung 6, verursacht durch EC25, ist nicht mehr erkennbar. Die Kovariablen x und x^2 sind als Stellvertreter-Variablen für eine unbekannte Variationsursache in x -Richtung zu verstehen.

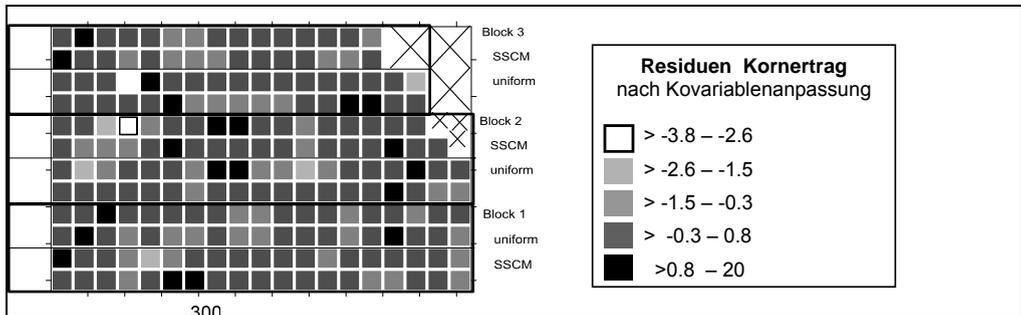


Abbildung 9: Lage der klassifizierten Residuen des Ertrags nach Anpassung der Kovariablen EC_{25} , EC_{25}^2 , x und x^2

Mit den ausgewählten Kovariablen werden dann wieder nach der REML-Methode die Varianzkomponenten geschätzt und anschließend die Prüfglieder verglichen. Die Ergebnisse sind in Tabelle 6 und Tabelle 7 dargestellt. Die angegebenen Ertragsmittelwerte sind die in 6.2 erhaltenen Ertragsmittelwerte, korrigiert auf das durchschnittliche beobachtete Niveau aller Kovariablen. Die Prüfglieddifferenz ist hier unabhängig von dem durchschnittlich erreichten Niveau der Kovariablen, da keine prüfgliedspezifische Abhängigkeit von den Kovariablen vorlag. Die ursprünglich durch die Variation der Kovariablen bedingte Variabilität des Ertrages wurde herausgerechnet, so dass sich eine kleinere Standardabweichung der Differenz ergibt als in 6.2. 50

6.4.2 Einbeziehung von Kovariablen mit geostatistischem Ansatz

Wenn die Kovariablen festgelegt sind, dann sollten sie im Folgenden Bestandteil des Modells sein. Die Ansätze 6.2 und 6.4.1 setzten eine konstante Korrelation der Ertragsdaten innerhalb der Parzellen voraus. In 6.2 war sie sehr klein und in 6.4.1 wurde sie sogar Null. Eine konstante Korrelation ist sicher nicht realistisch; vielmehr ist zu erwarten, dass sie eine Funktion der Distanz zwischen den Rasterzellen ist: für kürzere Distanzen wird sie größer als für weiter entfernte Zellen sein. Deshalb soll geprüft werden, ob und welches räumliche Modell diese Abhängigkeit am besten beschreiben kann. Bei gleicher Kovariablenanzahl sind die verschiedenen räumlichen Modelle wieder mit der REML-Methode anzupassen und deren AICC-Werte zum Vergleich der Anpassungsgüte heranzuziehen. Neben den verschiedenen möglichen und in PROC MIXED zur Verfügung stehenden räumlichen Modellen kann entschieden werden, ob die Korrelationen zwischen den Rasterzellen innerhalb der Parzellen oder aber innerhalb der Blocks oder sogar über den gesamten Versuch berücksichtigt werden sollen. Diese verschiedenen ‚Einflussphären‘ der berücksichtigten Korrelation können durch die Optionen subject = parzelle oder subject = block oder subject = intercept angewiesen werden. Die folgenden Anweisungen zeigen von allen angepassten räumlichen Modellen und den drei verschiedenen Einflussphären der Korrelation jene, welche letztlich zur besten Anpassung geführt haben:

```
PROC MIXED DATA=Gerste METHOD=reml;
  CLASS pg block;
  MODEL ertrag = pg block x xx ec25 ec / DDFM=kr(firstorder);
  REPEATED / TYPE = sp(sphga) (x y) SUBJECT = block;
  LSMEANS pg / PDIFF = controlu('einh');
RUN;
```

Das bestangepasste räumliche Modell war hier das geometrisch anisotrope sphärische Modell (sphga) mit Berücksichtigung der Korrelationen innerhalb der Blocks und den Parametern $sill = 0.826$, $range = 154.33$ m in Richtung 57.3° und 84.2 m in Richtung 147.4° (Winkel werden in SAS in Uhrzeigerichtung gemessen und die Richtung der y-Achse entspricht 0°).

Tabelle 6 fasst die Ergebnisse zur Modellanpassung der Ansätze 6.2, 6.4.1 und 6.4.2 zusammen, wobei die Ergebnisse der verschiedenen räumlichen Modelle nicht dargestellt sind. Der Vollständigkeit halber sind auch die Ergebnisse nach Ansatz 6.4.2 ohne Berücksichtigung der Kovariablen (kursiv) aufgenommen. Der Vergleich der $AICC_{ML}$ -Werte verdeutlicht den Verlust an Anpassungsgüte, wenn der geostatistische Ansatz ohne die Kovariablen gewählt worden wäre.

Tabelle 6: Vergleich der Modellanpassung bei Nutzung der Einzelwerte;
fett: Modell mit bester Anpassung

Ansatz	Kovariablen / Varianzhomogenität	subject	$AICC_{REML}$ ($AICC_{ML}$)
6.2	Ohne / nein	-	824.4 (826.8)
6.4.1	x, x ² , EC25, EC25 ² / ja	-	546.4 (492.4)
6.4.2 ¹⁾	Ohne / ja	Parzelle	566.5 (578.7)
		Block	533.1 (543.8)
		intercept	559.9 (569.2)
	x, x ² , EC25, EC25 ² / ja	Parzelle	472.8 (425.8)
		Block	468.4 (426.4)
		intercept	489.8 (456.1)

¹⁾ mit dem geometrisch anisotropen sphärischen Modell

In Tabelle 7 sind sämtliche Prüfgliedvergleiche zusammengestellt. In jedem Falle ergibt sich, dass die Nullhypothese nicht verworfen werden kann, d. h. die Überlegenheit der teilflächenspezifischen Düngung konnte nicht nachgewiesen werden. In anderen Versuchen könnten die Testaussagen widersprüchlich sein. Dann muss natürlich das Ergebnis mit der besten Modellanpassung interpretiert werden. Hier werden die Ergebnisse nur zu Demonstrationszwecken gegenüber gestellt.

Tabelle 7: Prüfgliedvergleiche für den Ertrag nach den erläuterten Ansätzen;
fett: Modell mit bester Anpassung

Ansatz	Kovariablen / Varianzhomogenität	Mittelwert		Differenz teilfl-einh	s.e.d.	p-Wert	
		einh	teilfl			t-Test ¹⁾ Prüfglieder	F-Test Block
6.1	ohne / ja	9.2391	9.1410	-0.0981	0.3753	0.5909	0.2410
6.2	ohne / nein	9.2343	9.1486	-0.0857	0.3699	0.5806	0.2360
6.4.1	x, x ² , EC25, EC25 ² / ja	9.2381	9.0968	-0.1413	0.1045	0.9110	0.0078
6.4.2 ²⁾	ohne / ja	8.9550	8.8338	-0.1212	0.1788	0.7497	0.6259
	x, x ² , EC25, EC25 ² / ja	9.2580	9.1184	-0.1396	0.1604	0.8064	0.5600

¹⁾ p-Wert für einseitigen t-Test

²⁾ mit dem geometrisch anisotropen sphärischen Modell und subject = block

6.5 Ansatz zur detaillierteren Bewertung der Behandlungen

Die bisherigen Ansätze gingen von einer mittleren Bewertung der beiden Prüfglieder aus. Im Folgenden wird ein weiterer Ansatz beschrieben, der den gewählten Applikationsalgorithmus in seiner Wirkung auf den Ertrag detaillierter betrachtet. Die bisher nicht gefundenen Unterschiede gefunden könnten darauf zurückzuführen sein, dass sich bei den gewählten teilflächenspezifischen Düngergaben im Mittel positive und negative Effekte gegenüber der einheitlichen Variante aufheben.

Aus diesem Grunde wurden die Auslenkungswinkel (und damit die Düngergaben) klassifiziert. Das Ziel besteht darin, die Prüfglieder je Winkelklasse zu vergleichen. Folgende Klassifizierung wurde vorgenommen (siehe auch Abbildung 10):

$$\text{Winkelklasse} = \begin{cases} 1 & \text{wenn } \Phi \leq 15^\circ \\ 2 & \text{wenn } 15^\circ < \Phi \leq 20^\circ \\ 3 & \text{wenn } 20^\circ < \Phi \leq 25^\circ \\ 4 & \text{wenn } \Phi > 25^\circ \end{cases}$$

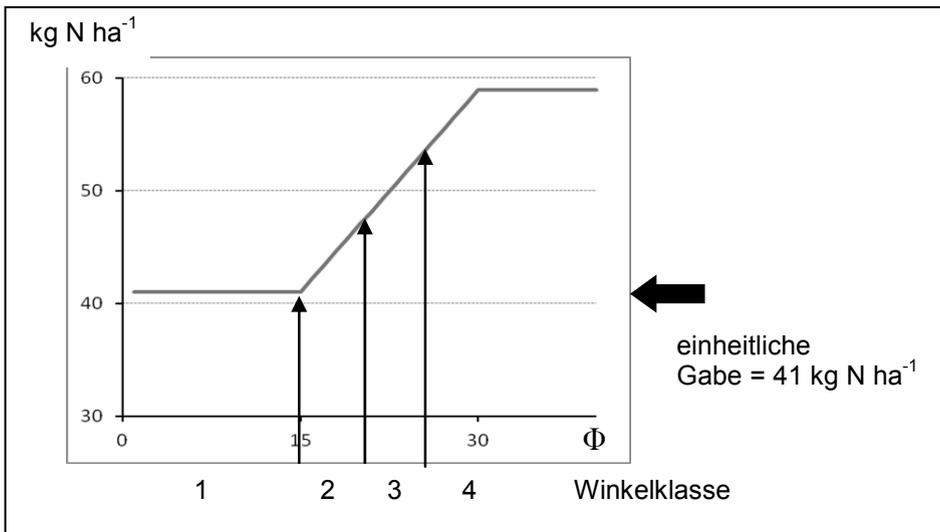


Abbildung 10: Definition der Winkelklassen für beide Prüfglieder

Die Winkelklassen wurden so gewählt, dass noch hinreichend große Stichprobenumfänge in den Klassen verblieben (Tabelle 8), wobei auf Grund dessen eine Klasse mit $\Phi > 30$ nicht möglich war. In der 1. Klasse waren die einheitliche und teilflächenspezifische Gabe identisch, so dass sich im zweiseitigen Test kein Ertragsunterschied ergeben dürfte. Dadurch ist auch hier nochmals eine Kontrolle der Versuchsmethodik gegeben. Wenn die zusätzlichen Gaben einen Effekt haben, ist zu erwarten, dass in den folgenden Winkelklassen die Ertragsdifferenzen zunehmen. Abbildung 11 zeigt, wie die Winkelklassen räumlich verteilt waren.

Tabelle 8: Stichprobenumfänge für die Prüfglieder in den einzelnen Winkelklassen

Prüfglied	Winkelklasse				Σ
	1	2	3	4	
einfl	8	33	45	23	109
teilfl	6	27	37	33	103
Σ	14	60	82	56	212

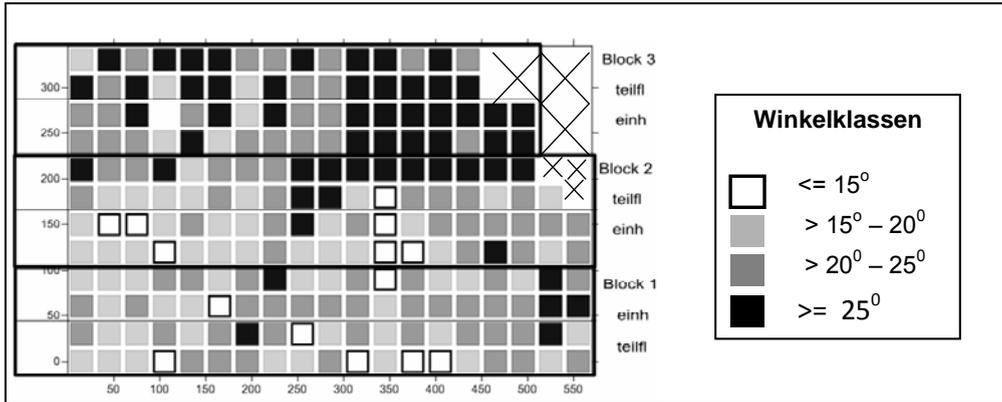


Abbildung 11: Lage der klassifizierten Winkel

Die Winkelklassen (Variable: winkkl) stellen jetzt einen zusätzlichen Klassifikationsfaktor dar. Damit wird eine zweifaktorielle Auswertung unter Berücksichtigung von Wechselwirkungen zwischen Prüfgliedern und Winkelklassen möglich, die unter Umständen eine Erklärung für die nichtsignifikante (mittlere) Prüfglieddifferenz in den vorangegangenen Analysen geben könnte. Auch hier kann mit oder ohne Kovariablen gearbeitet werden:

```
PROC MIXED DATA=Gerste METHOD=reml;
  CLASS pg winkkl block;
  MODEL ertrag = pg winkkl pg*winkkl block x xx ec25 ec/
  DDFM=kr(firstorder);
  REPEATED / TYPE=sp(sphga) (x y) SUBJECT = block;
RUN;
```

Das geometrisch anisotrope sphärische Modell mit subject = block stellte sich wiederum als das bestangepasste räumliche Modell heraus. In Tabelle 9 sind die Ergebnisse unter Nutzung des geostatistischen Modells mit und ohne Kovariablen gegenübergestellt. Die LS-Means-Anweisungen sollten spezifisch für die interessierenden Vergleiche gewählt werden.

Basierend auf dem AICC_{ML} ist auch hier der erhebliche Anpassungsgewinn durch die Hinzunahme der Kovariablen erkennbar. Wechselwirkungen zwischen den beiden Faktoren und Unterschiede zwischen den Prüfgliedern im Mittel der Winkelklassen können im F-Test nicht nachgewiesen werden. Während sich die Winkelklassen im Mittel der Prüfglieder im Ansatz ohne Kovariablen unterscheiden, geht dieser im zweiten Ansatz verloren (Tabelle 9). Die F-Tests aus Tabelle 9 prüfen zweiseitige Hypothesen. Beim Vergleich der Winkelklassen interessieren jedoch primär einseitige Hypothesen: Zeigen besser etablierte Bestände zum Fahnenblattstadium auch höhere Erträge? Die Ertragsmittelwerte für die Winkelklassen sind

in Tabelle 10 angegeben. Klar erkennbar ist die Tendenz, dass mit höherer Winkelklasse der Ertrag ansteigt. Offensichtlich sind größere Auslenkungswinkel zum Zeitpunkt der Düngung auch mit höheren Biomassen zum Zeitpunkt der Ernte verknüpft.

Tabelle 9: Anpassungskriterien und F-Tests für den Ertrag bei bestangepasstem geostatistischen Modell mit und ohne Kovariablen und diskretisierten Winkeln

Kovariablen / Varianzhomogenität	AICC _{REML} (AICC _{ML})	Variationsursache	p-Wert F-Test
ohne / ja	518.9 (534.2)	PG	0.5466
		Winkelklasse	0.0003
		WW PG x Winkelklasse	0.3416
		Block	0.7589
EC25, EC25 ² , x, x ² / ja	470.6 (431.2)	PG	0.5683
		Winkelklasse	0.0808
		WW PG x Winkelklasse	0.7238
		Block	0.6708
		x	<.0001
		x ²	0.0028
		EC25	<.0001
		EC25 ²	<.0001

Tabelle 10: Vergleich zwischen den Winkelklassen im Mittel der Prüfglieder nach Ansatz 6.5 für die Erträge

Kovariablen	Mittelwerte je Winkelklasse und einseitiger t-Test ¹⁾			
Ohne	1	2	3	4
	8.33	8.64	8.89	9.31
	a	a	b	c
EC25, EC25 ² , x, x ²	1	2	3	4
	8.82	9.10	9.21	9.34
	a	b	bc	c

¹⁾ $\mu_i < \mu_j$ für alle $i < j$ und $i, j = 1, 2, 3, 4$ Winkelklasse; Mittelwerte mit gleichem Buchstaben unterscheiden sich nicht signifikant bei $\alpha = 0.05$.

Analog zu der in 6.1 bis 6.4 geprüften einseitigen Hypothese ist der einseitige Vergleich der Prüfglieder je Winkelklasse von Interesse. Ausgenommen davon ist die 1. Klasse, wo die zweiseitige Alternative getestet werden soll (Tabelle 11).

Die Betrachtung je Winkelklasse ergibt hier keine zusätzlichen Informationen – in allen Winkelklassen kann keine Überlegenheit der teilflächenspezifischen Düngung nachgewiesen werden. In der 1. Winkelklasse kann von Gleichheit der mittleren Erträge ausgegangen werden, ein zusätzlicher Beweis, dass keine groben methodischen Fehler vorgelegen haben.

Tabelle 11: Vergleich der Prüfglieder je Winkelklasse bei bestangepasstem geostatistischem Modell mit und ohne Kovariablen

Ansatz	Winkel- klasse	Mittelwert		Differenz teilfl-einh	s.e.d.	p-Wert t-Test Prüfglieder
		einh	teilfl			
Ohne Kovari- ablen	1	8.3439	8.3206	-0.0232	0.4136	0.9553 ¹⁾
	2	8.7428	8.5358	-0.2070	0.2362	0.8086 ²⁾
	3	8.8515	8.9358	0.0843	0.2015	0.3385 ²⁾
	4	9.4681	9.1484	-0.3198	0.2521	0.8962 ²⁾
mit EC25, EC25 ² , x, x ²	1	8.7730	8.8766	-0.1036	0.3436	0.7633 ¹⁾
	2	9.2121	8.9937	-0.2184	0.2064	0.8539 ²⁾
	3	9.2559	9.1607	-0.0952	0.1840	0.6968 ²⁾

¹⁾ zweiseitiger t-Test ²⁾ einseitiger t-Test

6.6 Zusammenfassung der Analysenschritte

Abbildung 12 zeigt eine Übersicht über die in 6.1 bis 6.5 demonstrierten Analysenschritte.

7 Schlussfolgerung

Als fachliche Schlussfolgerung muss für diesen Versuch festgestellt werden, dass der Ertrag primär durch den Bestandeszustand vor der 3. Düngergabe bestimmt war, die erhöhten teilflächenspezifischen Düngergaben zeigten keinen positiven Effekt. Die Einschätzung des Bestandes mittels des Crop-Meters zum Fahnblattstadium spiegelte sich in den Ertragsrelationen wider.

Analysenschritt		Motivation
<p>6.1 Mittelwertbildung über alle Einzelwerte je Parzelle = klassische Blockanlage</p> <p>6.2 Nutzung der Einzelwerte je Parzelle ohne Berücksichtigung der Georeferenzierung</p> <p>6.4 Nutzung der Einzelwerte je Parzelle unter Berücksichtigung der Georeferenzierung</p> <p>6.4.1 ohne geostatistischen Ansatz, aber mit Kovariablen</p> <p>6.4.2 mit geostatistischem Ansatz (mit und ohne Kovariablen)</p>	<p>mittlere</p> <p>Aussage über die beiden Prüfglieder</p>	<p>Analyse gerechtfertigt auf Grund der Randomisation (Basisauswertung)</p> <p style="text-align: center;">↓</p> <p>Nutzung der Information aus wiederholten Messungen bei konstanter Korrelation in Parzellen</p> <p style="text-align: center;">↓</p> <p>Reduzierung der Restvariabilität (u. U. Vermeidung von Verzerrungen der PG-Differenz)</p> <p style="text-align: center;">↓</p> <p>Nutzung der Information aus wiederholten Messungen bei flexiblem Ansatz für die Korrelation</p> <p style="text-align: center;">↓</p>
<p>6.5 Nutzung der Einzelwerte je Parzelle unter Berücksichtigung der Georeferenzierung bei Klassifizierung der N-Düngung</p> <p><i>demonstriert:</i> mit geostatistischem Ansatz (mit und ohne Kovariablen)</p>	<p>detailliertere</p>	<p style="text-align: center;">↓</p> <p>wie zuvor und zusätzlich: Information über die Wirkungsweise des gewählten Applikationsalgorithmus im Verhältnis zur einheitlichen Düngung</p>

Abbildung 12: Überblick über die gewählten Analysenschritte

Literatur

- DAMMER, K. H. (2005): Demonstration der Langzeitwirkung bedarfsorientierter Fungizidbehandlung mit dem CROP-Meter. Bornimer Agrartechnische Berichte, Heft 41.
- EHLERT, D.; HAMMEN, V. C.; ADAMEK, R. (2003): On-line sensor pendulum-meter for determination of plant mass. Precision Agriculture 4, 139-148.
- HAMMEN, V. C. (2001): On-line sensing of cereal crop biomass. Dissertation, Humboldt-Universität zu Berlin, Germany.
- THÖLE, H. (2010): Ansätze zur statistischen Auswertung von On-Farm-Experimenten mit georeferenzierten Daten. Dissertation, Humboldt-Universität zu Berlin, Germany.

Analyse von On-Farm-Experimenten am Beispiel der Wirkung von Fungizidbehandlungen im Stadium der Rapsblüte (BBCH 65)

DOMINIK DICKE und ANDREAS BÜCHSE

Zusammenfassung

Ziel dieser Versuchsserie war es, die Wirkung von zwei Fungiziden zur Blütenbehandlung in Winterraps im Vergleich zur jeweils unbehandelten Kontrolle zu untersuchen. Hierzu wurden in den Jahren 2009-2011 in Praxisschlägen jeweils randomisierte Blockanlagen mit drei bzw. vier Wiederholungen angelegt. Die Größe der 9 bzw. 12 Parzellen betrug zwischen 21 m und 30 m in der Breite und zwischen 70 m und 270 m in der Länge. Aus jeder Parzelle wurden mit einem praxisüblichen Mähdruscher Streifen im Kerndrusch mit GPS-Ertragserfassung geerntet. Die kalibrierten Werte aus der Ertragserfassung wurden mit einem geostatistischen Modell verrechnet. Durch Fungizideinsatz zur Blüte konnte im Jahr 2009 gegenüber der Kontrolle ein signifikanter Mehrertrag erzielt werden. In 2010 und 2011 waren die Ertragsunterschiede zwischen den Varianten dagegen gering. Die Präzision der Versuche war ausreichend, um Ertragseffekte im Bereich von 3 bis 5 % statistisch abzusichern. Diese Versuchsmethodik kann mit der handelsüblichen Maschinenausstattung vom Landwirt selbst auf seiner Ackerfläche (On-Farm) praktiziert werden. Hierdurch können die Ergebnisse aus amtlichen Prüfungen in Kleinparzellen ergänzt werden. Durch die so ermöglichte Kombination aus hoher Präzision und Praxisnähe können standortspezifische Beratungsempfehlungen im Ackerbau effizient abgeleitet werden.

1 Einleitung

Amtliche Beratungsempfehlungen im Pflanzenschutz basieren auf Versuchsergebnissen. Üblicherweise wird die Wirkung verschiedener Pflanzenschutzvarianten in Kleinparzellen auf Versuchsfeldern getestet, die an verschiedenen Orten innerhalb des Beratungsgebietes etabliert sind. Diese Versuche werden mit Spezialmaschinen angelegt, geführt und geerntet; hierbei werden praxisnahe Bedingungen angestrebt, zwangsläufig unterscheiden sich diese jedoch von den praxisüblichen Verhältnissen. Zudem ist aus technischen und finanziellen Gründen oftmals die Anzahl an Wiederholungen begrenzt, so dass pflanzenbaulich interessante Effekte zuweilen nicht statistisch abgesichert werden können. In den letzten Jahren sind die technischen Möglichkeiten im Rahmen der Online-Ertragserfassung, Sensorik und GPS-Technik rasant gewachsen. Eine GPS-Ertragserfassung während des Mähdrusches oder Messung der scheinbaren elektrischen Bodenleitfähigkeit via EM 38 (CORVIN und LESCH, 2005) dient zur Erstellung von Karten für Precision-Farming-Anwendungen (GRAEFF und CLAUPEIN, 2007). Sensordaten können zudem gezielt genutzt werden, um Interaktionen zwischen Bodeneigenschaften und pflanzenbaulichen Maßnahmen im Versuch zu untersuchen (DICKE, 2008). RITTER et. al. (2008) entwickelten eine auf Sensordaten basierende Methode, um die Ertragswirkung unterschiedlicher Herbizidvarianten auf Praxisflächen unter realen Produktionsbedingungen zu studieren. DICKE et. al. (2012) prüften die Herbizidempfindlichkeit unterschiedlicher Maissorten durch Integration von Sensorik im Feldversuch. Versuche unter realen Produktionsbedingungen sind schon seit je her eine bedeutsame Komponente des landwirtschaftlichen Versuchswesens. Mit „Produktionsbedingungen“ ist gemeint, dass die Anlage der Varianten auf Flächen mit einer praxisüblichen Größe und der gebräuchlichen landwirtschaft-

lichen Maschinenausstattung durchgeführt wird. Sollen aus diesen Versuchen jedoch amtliche Empfehlungen abgeleitet werden, dann ist die Einhaltung der Grundsätze des landwirtschaftlichen Versuchswesens (Blockbildung, Wiederholung, zufällige Verteilung der Versuchsglieder (Randomisation)) zwingend erforderlich. Diese Versuche fallen unter den Begriff „On-Farm-Experimente“ (OFE). Sie haben einen gleich hohen Anspruch an statistische Präzision wie auch Kleinparzellenversuche auf den amtlichen Versuchsfeldern. Die einfache Erprobung einer Variante, die ein einzelner Landwirt auf seinen Ackerflächen ohne Einhaltung der Regeln des Versuchswesens durchführt, um einen orientierenden Überblick zu bekommen, sollte strenggenommen nicht unter dem Begriff OFE eingeordnet werden - die Ergebnisse sind nicht verallgemeinerbar.

Es liegt nahe zu prüfen, inwieweit Sensordaten und GPS-Technik zur Klärung von pflanzenbaulichen Fragestellungen eingesetzt und ins Versuchswesen integriert werden können.

Seit 2009 führt der hessische Pflanzenschutzdienst in Zusammenarbeit mit Landwirten OFE-Versuche in Raps aber auch Getreide durch. Die Durchführung sowie die Ergebnisse der letzten drei Jahre zur Wirkung der Blütenbehandlung in Raps werden im Folgenden erläutert. Ziel der vorliegenden Arbeit war es, unter On-Farm-Bedingungen in Raps die Wirkung von Fungiziden im Vergleich zur unbehandelten Kontrolle zu untersuchen. Hierbei wurden durch georeferenzierte Sensorik gewonnene Ertragsdaten genutzt. Verschiedene statistische Modelle wurden verglichen.

2 Material und Methoden

In den Jahren 2009 bis 2011 wurden in Winterrapsbeständen am Standort Grund-Schwalheim bei Echzell sowie Ober-Erlenbach in Hessen auf 4 bis 6 ha großen Rapsschlägen Versuche angelegt, um die ertragssichernde Wirkung der Fungizide Proline (0.7 l ha^{-1}) (Variante 1) und Harvesan (0.8 l ha^{-1}) (Variante 2) gegen Weißstängeligkeit im Vergleich zur unbehandelten Kontrolle (Variante 3) zu überprüfen. Die sonstigen Betriebsmittel (Pflanzenschutz, Düngung etc.) wurden von den Landwirten betriebsüblich über die gesamte Fläche einheitlich appliziert. Die Versuche wurden jeweils als randomisierte Blockanlagen mit drei bzw. vier Wiederholungen und Parzellenbreiten von bis zu 30 Metern angelegt. Die Parzellenlängen entsprachen jeweils der Ausrichtung der Fahrgassen mit Aussparung der Vorgewende. Die Fungizidbehandlung wurde nach vorgegebenen Versuchsplan vom jeweiligen Landwirt mit seiner handelsüblichen Feldspritze zur Blüte (BBCH 65) durchgeführt ($300 - 400 \text{ l Wasser ha}^{-1}$, Fahrgeschwindigkeit: ca. 4 km h^{-1}). Die folgende Tabelle 1 enthält die Kennzahlen zu den Versuchen der Jahre 2009 bis 2011.

Tabelle 1: Kennzahlen zu den Versuchen (2009-2011)

Versuchsjahr	Wiederholungen	Parzellenbreite	Parzellenlänge	Druschbreite (Kerndrusch)
2009	3	30 m*	140-270 m	5.7 m
2010	3	21 m**	160 m	7.5 m
2011	4	21 m**	70-76 m	7.5 m

* Parzellenbreite entspricht 2 Spritzbreiten von 15m = 2 Fahrgassen. Je Parzelle wurden mit voller Druschbreite 2 Durchfahrten realisiert (bei allen Versuchen)

** Parzellenbreite entspricht Spritzbreite von 21 m = 1 Fahrgasse

Zur Ernte wurden die Parzellen jeweils im Kerndrusch mit einem Mähdrescher (Hersteller Claas) und GPS-Ertragserfassung, bei voller Schnittbreite gemäht. Im Abstand von jeweils fünf Sekunden wurde ein Ertragswert mit zugehöriger GPS-Koordinate abgespeichert, so dass

je Parzelle eine hohe Anzahl an Messpunkten (Messwiederholungen) realisiert werden konnten. Je Parzelle wurden zwei Durchfahrten (Erntespuren) angestrebt.

Das Erntegut jeder Parzelle wurde jeweils auf einer Radlastwaage gegengewogen. Der auf der Radlastwaage ermittelte Wert aus der jeweiligen Parzelle wurde durch den Durchschnittswert der Sensordaten der jeweiligen Parzelle dividiert. Der resultierende Wert wurde wiederum mit den Sensordaten der jeweiligen Parzellen multipliziert. Damit wurden die Sensordaten parzellenweise kalibriert, d. h. auf das durchschnittliche Ertragsniveau der realen Wiegung gebracht. Die jeweils ersten sechs bis zehn Ertragsmesspunkte bei den Einfahrten in die Parzellen wurden verworfen, da hier erst ein kontinuierlicher Ertragsmessstrom aufgebaut wird und diese Werte daher fehlerbehaftet sind. Ernteproben mit unterschiedlichem Feuchtegehalt wurden auf einen einheitlichen Trockensubstanzgehalt von 91 % korrigiert. Die folgende Abbildung 1 zeigt ein Luftbild der Versuchsanlage mit Ertragsmesspunkten der Online-Ertragserfassung.

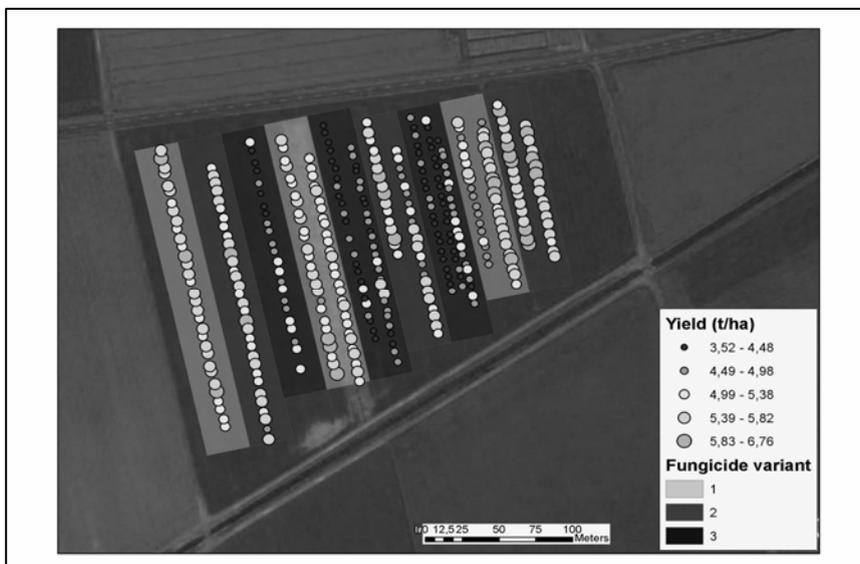


Abbildung 1: Luftbild der Versuchsanlage 2009 mit Messpunkten aus der Online-Ertragserfassung

2.1 Statistische Verrechnung

2.1.1 Allgemeine Vorbemerkungen

Die statistische Auswertung von OFE mit Online-Ertragserfassung unterscheidet sich nicht grundsätzlich von der Auswertung von Ertragsdaten jedes anderen landwirtschaftlichen Experimentes. Folgende Aspekte verdienen jedoch besondere Aufmerksamkeit:

- Der Mähdrescher ist sachgemäß mit gewogenen Proben zu kalibrieren; die Rohdaten sind aufzubereiten und insbesondere auf einheitliche Feuchte zu adjustieren; eine eventuell verfügbare Georeferenzierung muss kontrolliert werden und hinsichtlich Störgrößen abgeglichen werden (Fahrrichtung, Fahrgeschwindigkeit, Zeitverzug zwischen GPS-Signal, Ernteschnitt, Wiegung).
- Die aufbereiteten Rohdaten sind kritisch zu untersuchen und eventuelle Fehlmessungen sind vor der eigentlichen statistischen Analyse zu eliminieren. Das Problem von unplau-

siblen Werten und Ausreißern stellt sich durch die bei On-Farm-Experimenten gegebene hohe Datendichte in besonderem Maße. Vor einem „Ausreißertest“ sollten die Daten aufgrund ihrer meist sehr großen Zahl in jedem Falle grafisch aufbereitet werden (Box-Plot, Streudiagramm). Ein statistischer Ausreißertest kann dann beispielsweise mit dem Grubbs-Test oder einem 3- oder 4-sigma-Test erfolgen. Bei Letzterem wird jeweils die Differenz zwischen dem verdächtigen Wert und dem Versuchsglied- oder Parzellenmittelwert gebildet und diese Differenz dann durch die Standardabweichung dividiert. Falls die Werte einer Normalverteilung folgen, so ist mit einer Wahrscheinlichkeit von etwa 1:350 ein Wert zu erwarten, der mehr als drei Standardabweichungen vom Mittelwert entfernt ist. Die Wahrscheinlichkeit für einen mehr als vier Standardabweichungen vom Mittelwert entfernten Wert beträgt dagegen nur etwa 1:15000. Solche Werte sollten also praktisch nicht vorkommen und man kann wohl in jedem Falle von einem Ausreißer ausgehen. Die Entscheidung, ob ein Einzelwert zu verwerfen ist, sollte jedoch immer auch fachlich begründet sein und nicht ausschließlich statistisch. Es ist zu bedenken, dass bei sehr großen Datenmengen auch extreme Werte zwangsläufig eine gewisse Wahrscheinlichkeit haben. Wenn Extremwerte verworfen werden, so ist zu überlegen, extrem niedrige und extrem hohe Werte gleichermaßen auszuschließen (beidseitiges Zensieren der Stichprobe).

- c) Bei OFE sind die Versuchsflächen in aller Regel sehr groß. Daher sind räumliche Trends zu erwarten, die im Rahmen der statistischen Auswertung bei Verfügbarkeit räumlicher Information berücksichtigt werden können bzw. müssen. Die Art dieser Trends ist nicht von vornherein klar. Sind die Trends linear, quadratisch, ein- oder zweidimensional? Es kann passieren, dass der Einschluss eines Trends in das Auswertungsmodell die Präzision der erhaltenen Schätzungen von Prüfgliedeffekten sogar vermindert. Hier liegt eine der größten Herausforderungen bei der Auswertung eines großflächigen OFE mit Georeferenzierung.

2.1.2 Datenaufbereitung im Beispielversuch

Im Versuchsjahr 2009 wurden die aus dem Mähdrusch erhaltenen Ertragsdaten zunächst mit den Parzellenerträgen der Radlastwaage kalibriert (Tabelle 2). Die kalibrierten Ertragswerte wurden dann einem 4-sigma-Test unterzogen. Für jede Parzelle wurden der arithmetische Mittelwert (\bar{y}) und die Standardabweichung (s) berechnet. Zur besseren Vergleichbarkeit wurde innerhalb der Parzellen eine z-Transformation der Beobachtungen y nach der folgenden Vorschrift vorgenommen: $z_y = (y - \bar{y}) / s$. Die Abweichung eines Einzelwertes vom Parzellenmittel sollte nicht mehr als vier Standardabweichungen betragen. Alle 367 Werte bewegten sich im geforderten Intervall und wurden für die weitere Auswertung verwendet (Tabelle 3).

Die wiederholten Messungen in einer Parzelle für statistische Tests als unabhängige Wiederholungen zu betrachten ist nicht zulässig. Es wurden daher zwei mögliche Auswertungsstrategien verfolgt:

- a) Über alle Einzelmessungen einer Parzelle wurde ein Mittelwert gebildet. Hierdurch wurden 12 Parzellenmittelwerte erhalten, die dann varianzanalytisch geprüft wurden.
- b) Die „Pseudo“-Wiederholungen innerhalb einer Parzelle wurden in einem geostatistischen Modell entsprechend berücksichtigt und modelliert. Hierbei wurden auch räumliche Trends überprüft.

Die folgenden Tabellen und Abbildungen beinhalten die Rohdaten, die Ermittlung von Extremwerten, den Boxplot der Ertragswerte sowie ein Ablaufschema der Datenanalyse.

Tabelle 2: Rohdaten Versuch 2009

Uhrzeit	Block	Fungi	Ertrag roh	Ertrag kalibriert	GPS Nord	GPS Ost
184627	2	1	5.3	6.03776	N50.4140019	E8.9187338
184632	2	1	5.1	5.80992	N50.4140668	E8.9187005
184637	2	1	4.7	5.35424	N50.4141146	E8.9186834
184642	2	1	5.8	6.60736	N50.4141829	E8.9186672
184647	2	1	5.9	6.72128	N50.4142341	E8.9186493
184652	2	1	4.7	5.35424	N50.4142990	E8.9186331
184657	2	1	4.4	5.01248	N50.4143503	E8.9186160
184702	2	1	4.6	5.24032	N50.4144322	E8.9185835
184707	2	1	4.3	4.89856	N50.4144834	E8.9185835
184712	2	1	4.9	5.58208	N50.4145517	E8.9185502
184717	2	1	5.1	5.80992	N50.4145996	E8.9185332
184722	2	1	4.9	5.58208	N50.4146644	E8.9184999

Tabelle 3: Untersuchung von Extremwerten mittels z-Transformation (Versuch 2009)

Parzelle	Block	Variante	N	Mittelwert	Min.	Max.	Std.Abw.	z_min	z_max
1	1	Proline	32	5.61	5.15	6.22	0.26	-1.75	2.32
2	1	Harvesan	32	5.58	4.81	6.70	0.33	-2.33	3.38
3	1	Kontrolle	23	4.87	4.40	5.43	0.32	-1.47	1.74
4	2	Proline	55	5.54	5.01	6.87	0.34	-1.54	3.86
5	2	Kontrolle	51	4.65	3.60	5.82	0.41	-2.54	2.83
6	2	Harvesan	38	5.48	4.89	6.42	0.34	-1.73	2.72
7	3	Kontrolle	61	4.60	3.99	5.43	0.41	-1.51	2.05
8	3	Proline	40	5.50	4.70	6.47	0.48	-1.67	2.03
9	3	Harvesan	35	5.88	5.35	6.91	0.31	-1.70	3.29

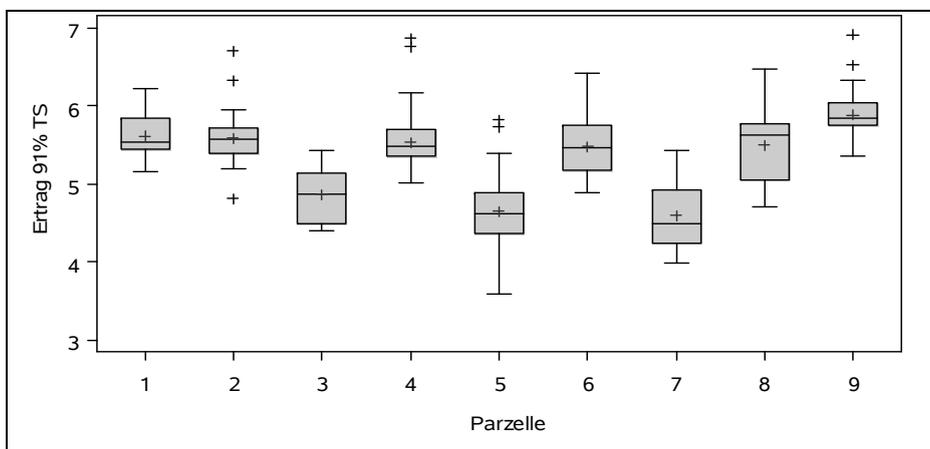


Abbildung 2: Boxplot der kalibrierten und feuchtekorrigierten Ertragswerte aus 2009; Parzellen 3, 5 und 7 sind die unbehandelte Kontrolle

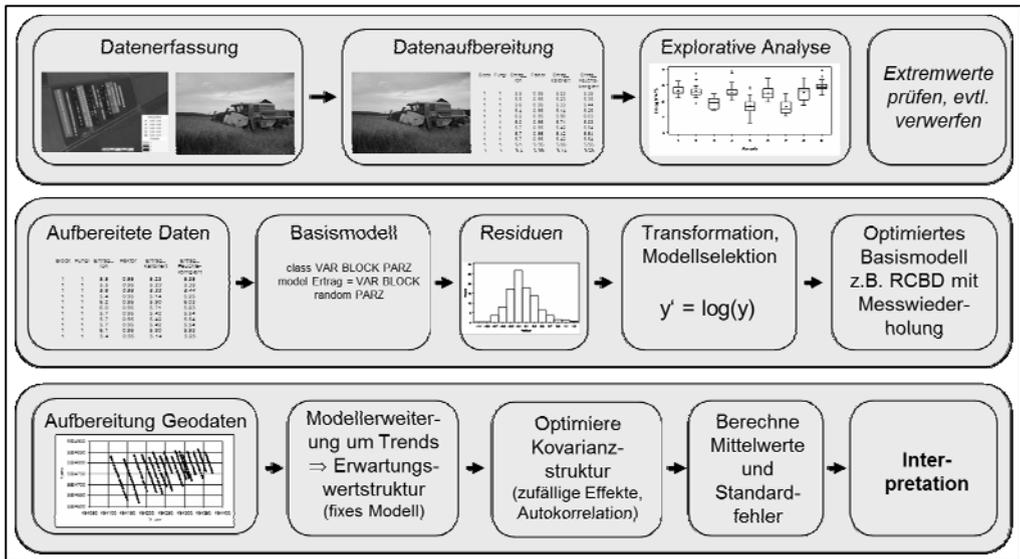


Abbildung 3: Ablaufschema der Datenanalyse

2.1.3 Auswertung über Parzellenmittelwerte

Je Parzelle wurden aus den Einzelwerten die Parzellenmittelwerte berechnet. Die erhaltenen 12 Mittelwerte wurden dann mittels Varianzanalyse und multiplem Mittelwertvergleich (t-Test) auf signifikante Varianteneffekte geprüft. Die Datenanalyse erfolgte mit der Prozedur MIXED im Statistikpaket SAS, Version 9.2.

```

PROC SORT DATA=raps;
  BY variante block;
RUN;
PROC MEANS DATA=raps NOPRINT ;
  BY variante block;
  VAR ertrag;
  OUTPUT OUT=means MEAN=ertrag N=n;
RUN;
PROC MIXED DATA=work.means LOGNOTE;
  CLASS variante block ;
  MODEL ertrag=variante block
    /DDFM=KENROG (firstorder) OUTP=res1;
  LSMEANS variante block/ PDIFF CL ;
  ODS OUTPUT LSMEANS=lsmeans DIFFS=diffs;
RUN;
QUIT;

```

2.1.4 Auswertung der Versuche mit Geostatistik

Für die geostatistische Auswertung wurden die Koordinatenachsen zunächst so transformiert, dass die Ernterichtung parallel zu einer der Hauptachsen verlief und ein Eckpunkt des Versuchsfeldes die Koordinaten 0;0 erhielt. Hierzu wurde eine Rotation ($x' = x \cdot \cos \alpha + y \cdot \sin \alpha$; $y' = -x \cdot \sin \alpha + y \cdot \cos \alpha$) und eine Translation ($x' = x - x_0$; $y' = y - y_0$) durchgeführt. Abbildung 4 zeigt die Ertragsmesspunkte nach der Transformation.

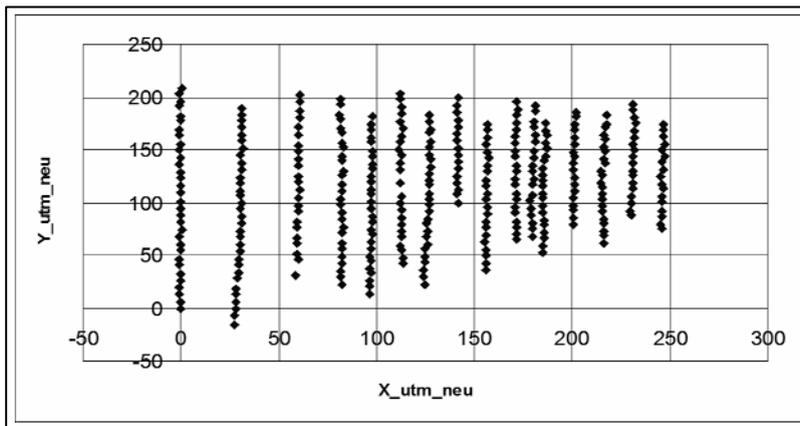


Abbildung 4: Plot der transformierten Messpunkte auf neuen Achsen

Für die weitere Auswertung wurden viele unterschiedliche Modelle geprüft. Der Modellvergleich wurde jeweils durch grafische Aufbereitung der Residuen (Differenzen zwischen vom Modell vorhergesagten und gemessenen Beobachtungen) und mit dem Akaike-Information-Kriterium (AIC) vorgenommen.

Die Korrelation zwischen Messpunkten innerhalb einer Parzelle wurde zunächst durch einen zufälligen Parzelleneffekt modelliert (Modell EW2). Es zeigte sich, dass zudem ein Effekt für die Druschspur sinnvoll ist, da hier besonders starke Abhängigkeiten auftraten (Modell EW3).

Bei grafischer Aufbereitung der erhalten Residuen zeigte sich zudem, dass die Druschrichtung sehr spurspezifische Effekte hervorrufen kann. Daher wurde ein spurspezifischer Trend in y-Richtung berücksichtigt (Modell EW4). Für die Residuen aus Modell EW4 wurde dann ein Variogramm erzeugt. Dieses stellt die Ähnlichkeit zwischen Messpunkten in Abhängigkeit von ihrer räumlichen Distanz dar (Abbildung 5). Hierbei zeigte sich eine räumliche Abhängigkeit bei zu einer Entfernung von etwa 50 Metern in Druschrichtung (y-Richtung). Die Form des Variogramms konnte am besten durch die räumliche Struktur „Gaussian“ mit Nugget-Effekt beschrieben werden (Tabelle 4). Ein Nugget-Effekt bedeutet, dass auch Beobachtungen, die am gleichen Punkt gemacht werden, eine Varianz aufweisen. Dieses kann als Messfehler der Online-Ertragsbestimmung interpretiert werden.

EW2: „Blockanlage mit Messwiederholung“

```
CLASS variante block parzelle  
MODEL ertrag=variante block  
RANDOM parzelle
```

EW3: „Blockanlage mit hierarchischer Struktur der Messwiederholung“

```
CLASS variante block parzelle druschspur  
MODEL ertrag=variante block  
RANDOM parzelle parzelle*druschspur
```

EW4: „Blockanlage mit Messwiederholung und Random Regression“ (Es werden spurspezifische Trends in Richtung der y-Koordinate berücksichtigt)

```
CLASS variante block parzelle druschspur richtung  
MODEL ertrag_feuchtekorrigiert=variante block  
RANDOM parzelle  
RANDOM int y / SUBJECT=parzelle*druschspur TYPE=UN
```

SPALT-SPALT-ANLAGE mit Spureffekten, Trend und Korrelationsstruktur Gaussian

```
CLASS variante block richtung druschspur parzelle  
MODEL ertrag_feuchtekorrigiert=block variante  
RANDOM int yutm_neu /  
SUBJECT=Druschspur*Parzelle TYPE=unrepeated/SUBJECT=  
Druschspur*Parzelle TYPE=sp(gau)(yutm_neu) local
```

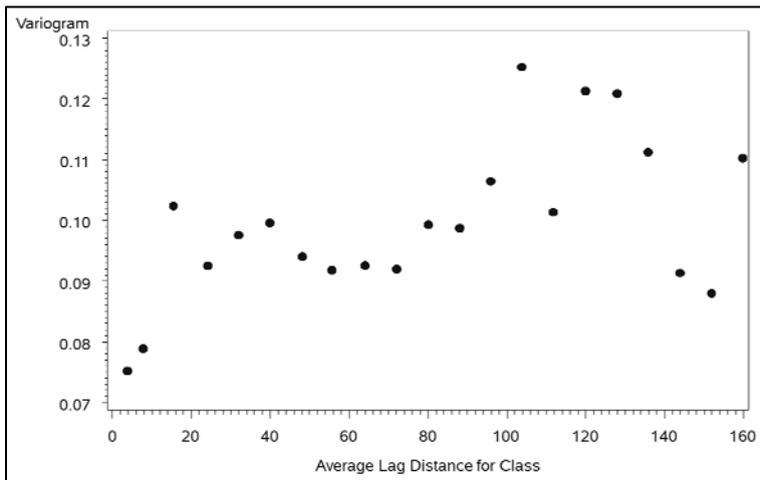


Abbildung 5: Variogramm der Residuen aus Modell EW4

Tabelle 4: Vergleich verschiedener räumlicher Korrelationsstrukturen

Modell	Typ der räumlichen Korrelationsstruktur	AIC
EW4	Keine räumliche Struktur	259.4
EW4 – power	Power + Nugget Effekt	Keine Konvergenz
EW4 – exp	Exponentiell + Nugget Effekt	248.4
EW4 – gau	Gaussian + Nugget Effekt	246.5
EW4 – sph	Sphärisch + Nugget Effekt	255.6
EW4 – lin	Linear + Nugget Effekt	Keine Konvergenz

3 Ergebnisse

3.1 Auswertung über Parzellenmittelwerte – Daten 2009

Die mittleren Erträge für Variante 1 und 2 lagen bei 5.55 bzw. 5.65 t ha⁻¹. Für die unbehandelte Kontrolle ergaben sich 4.70 t ha⁻¹. Der Unterschied zwischen den zwei behandelten Varianten war nicht statistisch signifikant (Tabelle 5). Beide Varianten unterschieden sich aber signifikant um 0.85 bzw. 0.95 t h⁻¹a von der Kontrolle. Die Grenzdifferenz lag bei 0.37 t ha⁻¹.

Tabelle 5: Mittelwertdifferenzen bei Auswertung über Parzellenmittelwerte

Variantenvergleich	Differenz	StdFeh	FG	p-Wert	95 %-Konfidenzintervall	Grenzdifferenz
Proline - Harvesan	-0.0999	0.1326	4	0.4930	-0.4680 0.2682	0.368
Proline - Kontrolle	0.8453	0.1326	4	0.0031	0.4772 1.2134	0.368
Harvesan - Kontrolle	0.9452	0.1326	4	0.0020	0.5771 1.3133	0.368

3.2 Auswertung mit geostatistischem Modell – Daten 2009

Die mittleren Erträge für Variante 1 und 2 lagen bei 5.61 t h⁻¹a. Für die unbehandelte Kontrolle ergaben sich 4.50 t ha⁻¹. Der Unterschied zwischen den zwei behandelten Varianten lag bei 0, zwischen den behandelten Varianten und der Kontrolle bei 1.11 t ha⁻¹ (Tabelle 6). Diese Differenz war eindeutig signifikant mit einer Grenzdifferenz von 0.28 bzw. 0.25 t ha⁻¹.

Tabelle 6: Mittelwertdifferenzen bei Auswertung mit geostatistischem Modell

Variantenvergleich	Differenz	StdFeh	FG	p-Wert	95 %-Konfidenzintervall	Grenzdifferenz
Proline - Harvesan	-0.00	0.122	3.84	0.97	-0.35 0.34	0.34
Proline - Kontrolle	1.11	0.105	4.58	0.0002	0.83 1.38	0.28
Harvesan - Kontrolle	1.11	0.109	8.63	0.0001	0.86 1.36	0.25

3.3 Ergebnisse 2010 und 2011

In 2010 und 2011 konnten weder zwischen den Fungizidmaßnahmen noch zwischen Fungizidmaßnahmen und unbehandelter Kontrolle Unterschiede gesichert werden. Im Mittel über alle drei Versuchsjahre wurde durch die Fungizidbehandlung ein Ertragsverlust von 0.35 t ha⁻¹ vermieden. Gemessen am mittleren Ertragsniveau waren das rund 6%.

Bezüglich der Grenzdifferenz konnte die komplexe geostatistische Auswertung keinen Vorteil gegenüber der einfachen Auswertung über die Parzellenmittelwerte erreichen. Tabelle 7 zeigt die Mittelwerte der Varianten in allen drei Versuchsjahren.

Tabelle 7: Mittelwerte bei Auswertung über Parzellenmittelwerte 2009-2011

Versuchsjahr	Kontrolle	Harvesan	Proline	Grenzdifferenz _{5%} (t ha ⁻¹)
2009	4.70 (a)	5.65 (b)	5.54 (b)	Blockanlage: 0.37 /Geostatistisch: 0.28
2010	5.87 (a)	5.97 (a)	6.07 (a)	Blockanlage: 0.21 /Geostatistisch: 0.26
2011	6.07 (a)	6.00 (a)	6.09 (a)	Blockanlage: 0.19
Mittel 2009-11	5.55	5.87	5.90	

*Mittelwerte mit gleichem Buchstaben unterscheiden sich nicht signifikant

4 Diskussion und Ausblick

In den hier dargestellten Versuchen zur Fungizidanwendung in Winterraps erwies sich ein OFE durchaus als geeignet, um die Leistungsfähigkeit verschiedener Produktionsverfahren zu untersuchen. Sehr wichtig war die Beachtung der Grundprinzipien des landwirtschaftlichen Versuchswesens: *Wiederholung, Blockbildung und Randomisation*. Ferner zeigte sich die Kalibrierung der Online-Ertrags Erfassung mit den Ergebnissen der Wiegung am Feldrand als wichtige qualitätssichernde Maßnahme. Die Einhaltung dieser Prinzipien garantiert in jedem Falle eine Auswertbarkeit über klassische varianzanalytische Modelle. Eine erweiterte geostatistische Datenanalyse kann dann zusätzlich durchgeführt werden, ist aber nicht zwingend und bringt auch nicht in jedem Falle einen wesentlichen Vorteil für die Aussagegenauigkeit. Soll der Einfluss von Bodenparametern mit berücksichtigt werden oder ist eine teilflächenspezifische Bewirtschaftung Teil der Versuchsfrage, dann ist ein geostatistischer Ansatz die Methode der Wahl. Dieses war in den hier vorgestellten Versuchen aber nicht das Ziel.

In Zukunft werden immer mehr Mähdrescher mit Ertrags erfassungssystemen auf den Betrieben bzw. in den Lohnunternehmen zu finden sein. Sensordaten über Bestandsinformationen und auch Bodenverhältnisse werden zunehmend zur Verfügung stehen, so dass die Wirkungen von Betriebsmitteln auch unter wechselnden Standortbedingungen immer besser beurteilt werden können. Durch die so ermöglichte Kombination aus hoher Präzision und Praxisnähe können standortspezifische Beratungsempfehlungen im Ackerbau effizient abgeleitet werden. Informationstechniken sollten in der Zukunft durch Zusammenarbeit mit Praktikern im Versuchswesen vermehrt genutzt werden.

In der Diskussion um die relative Vorzüglichkeit von OFE kommt oft die Frage auf, welche Bedeutung denn noch den Kleinparzellenversuchen zukommt. Hierbei ist zu betonen, dass Kleinparzellenversuche ein unabdingbarer Bestandteil des pflanzenbaulichen Versuchswesens sind und immer bleiben werden. Insbesondere wenn sehr viele Prüfglieder (Sortenprüfung, Düngungs- und Pflanzenschutzversuche) untersucht werden sollen, müssen die Parzellen klein und die Rahmenbedingungen standardisiert und damit vergleichbar sein, da sonst die Fläche des Gesamtversuches zu groß wird, Wachstumsbedingungen zu unterschiedlich werden und Pflegemaßnahmen nicht mehr homogen durchgeführt werden können. Des Weiteren werden kleinparzellerte Exaktversuche dringend benötigt, um unter möglichst definierten Bedingungen Grundlagen zur Entwicklung von Entscheidungsregeln zu erarbeiten. Die im Kleinparzellenversuch unter solchen definierten Bedingungen für eine Ackerbauregion gewonnenen Informationen können dann als Grundlage für die Prüfung unter speziellen Standortbedingungen im OFE dienen. Somit können sich Kleinparzellenversuche und OFE im Sinne der Optimierung von Beratungsaussagen ergänzen und methodisch unterstützen.

In Zukunft plant der hessische Pflanzenschutzdienst in Zusammenarbeit mit landwirtschaftlichen Betrieben weitere Fragestellungen mit OFE zu bearbeiten, so im Jahr 2012 die Ertrags-

wirkung von Insektizidmaßnahmen in Winterraps. Aber auch in anderen pflanzenbaulich interessanten Themenbereichen, z. B. Wirkung von organischen Düngern, Wirkung unterschiedlicher Bodenbearbeitungsverfahren oder Fragen zur Bewässerungsintensität, die in Kleinparzellen nur bedingt geprüft werden können, erscheint uns diese Methode sehr gut geeignet.

Literatur

- CORVIN, D. L.; LESCH, S. M. (2005): Apparent soil electrical conductivity measurements in agriculture. *Computer and Electronics in Agriculture* 46: 11-46.
- DICKE, D. (2008): Detecting effects of within-field heterogeneity on yield structure and crude protein content of winter wheat using sensor data and GIS. *Journal of Plant Diseases and Protection. Special Issue XXI*: 155-160.
- DICKE, D.; JACOBI, J.; BÜCHSE, A. (2012): Quantifying herbicide injuries in maize by use of remote sensing. *Julius-Kühn-Archiv, Proceedings 25. German Conference on Weed Biology and Weed Control, Vol. 1*: 199-205.
- GRAEFF, S.; CLAUPEIN, W. (2007): Applicability of precision farming datasets for the area of bioenergy. *Mitt. Ges. Pflanzenbauwiss.* 19: 274-275.
- RITTER, C.; DICKE, D.; WEIS, M.; OEBEL, H.; PIEPHO, H.-P.; BÜCHSE, A.; GERHARDS, R. (2008): An on-farm research approach to quantify yield variability and derive decision rules for site-specific weed management, *Prec. Agric.* 9: 133-146.

On Farm Research – ein Gemeinschaftsprojekt der Landwirtschaftskammer Schleswig-Holstein und der Gutsverwaltung Helmstorf

ULFRIED OBENAUF, IMKE BORCHARDT, CHRISTOPH LUBKOWITZ und
CARSTEN KOCK

Zusammenfassung

Der Beitrag beschreibt die Fragestellungen und bisherigen Ergebnisse seit 2007 durchgeführter OFE unter Praxisbedingungen auf einem Betrieb in Schleswig-Holstein. Auf einer Fläche von 300 ha werden Versuche zur Bodenbearbeitung und zur teilflächenspezifischen Stickstoffdüngung durchgeführt, wobei Aussaat und Grunddüngung gleichfalls teilflächenspezifisch gestaltet werden. Besonders wegen der Komplexität der Fragestellungen und strikter Orientierung an Produktionsbedingungen hat das Projekt Pilotcharakter über Schleswig-Holstein hinaus.

1 Zielstellung und Inhalte Projekt On Farm Research

Mit dem Projekt On Farm Research sollen Entscheidungshilfen für eine zukunftssträchtige Entwicklung des Marktfruchtbaus in Schleswig-Holstein erarbeitet werden. Es geht um die Erprobung intelligenter Technik zur Reduzierung des Betriebsmitteleinsatzes, zur Reduzierung der Kosten der Arbeitserledigung und zur Verbesserung der Produktionsverfahren (Ertragsoptimierung, Umweltwirkung, Qualitätssicherung) im typischen Marktfruchtbetrieb. Versuchsinhalte sind die Erprobung und Umsetzung teilflächenspezifischer Bewirtschaftungsverfahren auf den Gebieten:

- Bodenbearbeitung (Pflugsaat, Mulchsaat, Direktsaat),
- Stickstoffdüngung (betriebsüblich, N-Sensor, Einmaldüngung),
- Aussaat,
- Grunddüngung (P, K, Mg, Kalk) und
- Pflanzenschutz.

Die Bewertung des Erfolges der teilflächenspezifischen Produktionsverfahren erfolgt anhand von Ertrags- und Qualitätsmerkmalen des Erntegutes. Daran schließt sich die betriebswirtschaftliche Bewertung der Produktionsverfahren und deren Umsetzung im Betrieb als Beispiel leistungsfähigen Marktfruchtbaus in Schleswig-Holstein an, auch als Betriebsbeispiel mit bundesweitem Vorbildcharakter. Der Projektstart erfolgte mit der Herbstaussaat 2007; das Projekt befindet sich derzeit im 5. Versuchsjahr.

Finanziell und personell ist das Projekt von der Landwirtschaftskammer Schleswig-Holstein mit einer Jahreszuwendung aus dem Kammerhaushalt ausgestattet. Eine weitere finanzielle Förderung erfährt und erfährt das Projekt durch die Stiftung Schleswig-Holsteinische Landschaft (Beteiligung bei Einstiegsinvestitionen zur technischen Ausstattung) und eine jährliche Zuwendung auf Antrag durch die Landwirtschaftliche Rentenbank Frankfurt am Main.

1.1 Der Projektbetrieb und die Versuchsflächen

Der Projektbetrieb liegt im typischen Marktfruchtanbaugebiet Schleswig-Holsteins (Östliches Hügelland) und ist mit ausgeprägter Bodenheterogenität ausgestattet (Quelle: Reichsboden-

schätzung, EM-38 Karten). Detailangaben zum Betrieb wie Betriebsgröße, Boden, Klima, Anbauverhältnisse, personelle und technische Ausstattung können dem Betriebsspiegel bei www.lksh.de/landwirtschaft/pflanze/ unter „On Farm Research“ entnommen werden. Das Projekt wird auf insgesamt 300 Hektar der Betriebsfläche durchgeführt (Abbildung 1), hat eine Laufzeit von zehn Jahren und erfasst alle im Betrieb angebaute typischen Fruchtarten und Vorfrucht-Nachfruchtkombinationen.

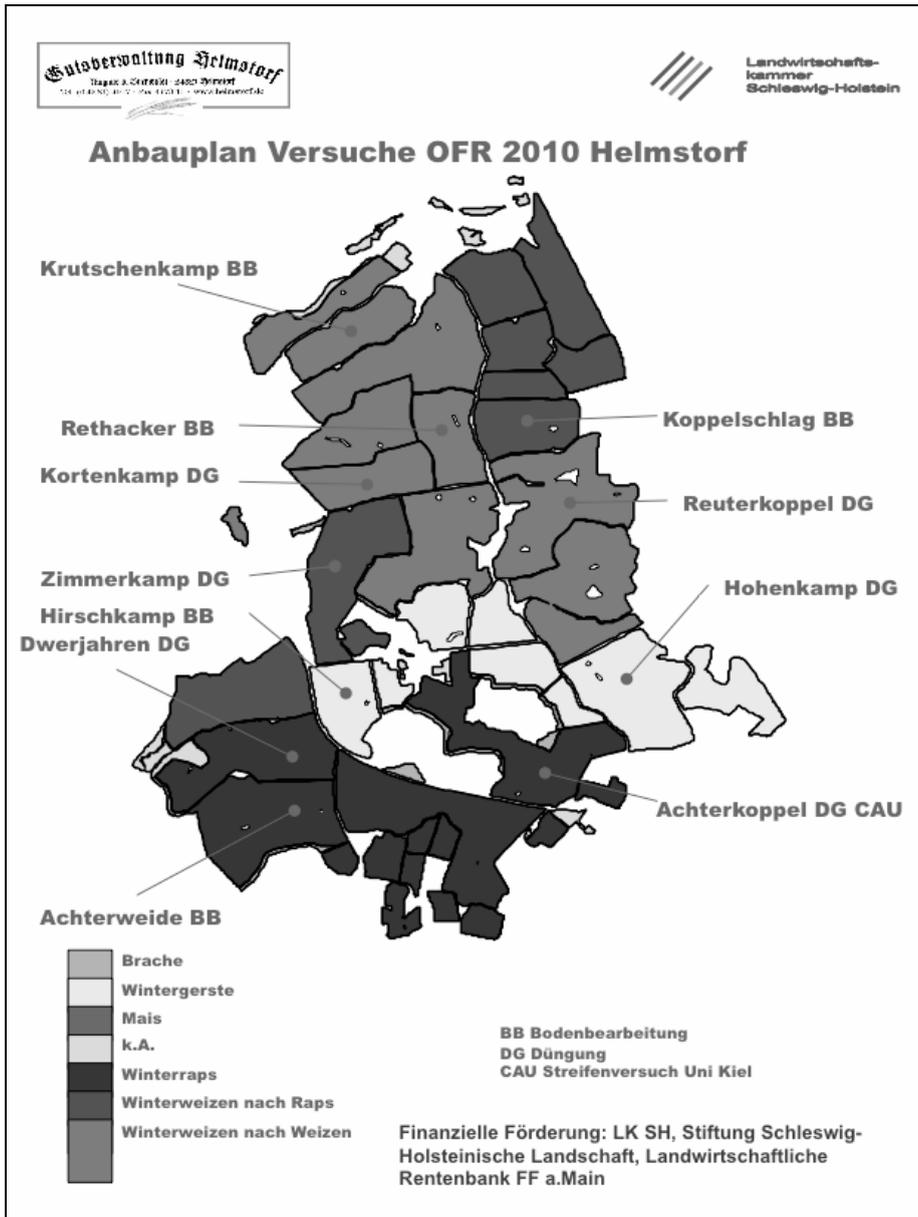


Abbildung 1: Übersicht zum Projekt On Farm Research Gut Helmstorf am Beispiel des Anbauplanes 2010 (RAW-Winterraps, WW – Winterweizen, WG – Wintergerste)

1.2 Bodenheterogenität und Versuchsdesign

Basis für die Bewertung der Bodenheterogenität der Versuchsschläge (und aller Schläge des Gesamtbetriebes) war die Kartierung über EM 38 und die Umsetzung der ermittelten differenzierten elektrischen Leitfähigkeit in „pragmatische“ drei Bodenklassen – leicht – mittel – schwer (z. B. für die teilflächenspezifische Aussaat). Anpassungen wurden mittels Betriebsleiterkenntnissen vorgenommen, wenn es um die Unterscheidung von Lehm- und Tonkuppen sowie humosen Senken und deren Zuordnung zu Bodenscannerklassen mit gleicher elektrischer Leitfähigkeit, aber unterschiedlichen Bodenzuständen ging (Abbildung 2). In Einzelfällen erfolgte ein ergänzender Abgleich mit Angaben aus der Reichsbodenschätzung.

Die Anlage der Versuche orientierte sich an den zwei separat betrachteten Problemstellungen: Vergleich von drei Bodenbearbeitungsvarianten (Pflugsaat, Mulchsaat, Direktsaat) – kurz: *Bodenbearbeitungsversuch* – und Vergleich von drei N-Düngungsvarianten (Betriebsüblich, N-Sensor, Einmaldüngung) – kurz: *N-Düngungsversuch* – (vergleiche auch Abbildung 1). Für diese Versuche wurden die Versuchsschläge aus Gründen der technischen Durchführbarkeit unter Praxisbedingungen in drei Teilstücke ohne Wiederholungen auf dem Einzelschlag unterteilt (Abbildung 3 und 12). Für die konkrete Festlegung der Größe, Form und Lage der Teilstücke waren drei Überlegungen ausschlaggebend: (1) Nutzung des existierenden Fahrgassensystems, (2) Anlage der Teilstücke so, dass in jedem die zuvor ermittelten Bodenscannerklassen vertreten sind (vergleiche z. B. Abbildung 2 und 3) und (3) Wahl der Teilstückgröße so, dass relativ neue Varianten mit schwer kalkulierbarem finanziellen Erfolg das Risiko für den Landwirt überschaubar halten. Letztere Überlegung führte z. B. dazu, dass die Teilstückfläche für die Direktsaat kleiner als die der anderen beiden Varianten ist (Abbildung 3). Die Anlage der beiden Versuche erfolgte im ersten Versuchsjahr auf mehreren Schlägen und wurde in den Folgejahren auf den gleichen Schlägen fortgeführt. Dabei war geplant, die Fruchtarten entsprechend zweier Fruchtfolgen (1. Raps – Winterweizen – Winterweizen und 2. Raps – Winterweizen – Gerste) anzubauen. Dieses ist zum großen Teil umgesetzt worden; in einzelnen Fällen musste allerdings auf Grund betrieblicher Besonderheiten und/oder witterungsbedingter Probleme zum Zeitpunkt der Aussaat davon abgewichen werden. Da kumulative Wirkungen der Bodenbearbeitung bzw. Düngung geprüft werden sollen, wurden die Varianten den Teilstücken nur im ersten Versuchsjahr – vergleichbar einem Dauerversuch – randomisiert zugeordnet. Die Anzahl der gewählten Schläge orientierte sich an der Überlegung, dass in jedem Versuchsjahr jedes Fruchtfolgeglied auf einem Schlag vertreten sein soll. Damit existiert in beiden Versuchen für jedes Fruchtfolgeglied und jede Variante nur eine echte Wiederholung in jedem Jahr.

Innerhalb der Parzellen werden Aussaat und Grunddüngung teilflächenspezifisch, jedoch nach einem einheitlichen Algorithmus für alle Bodenbearbeitungs- und N-Düngungsvarianten, umgesetzt. Deren Realisierung wird in Abschnitt 2 und 3 beschrieben. Diese teilflächenspezifischen Applikationen werden nicht randomisationsbasiert vorgenommen, sie sind ausschließlich an der existierenden Heterogenität innerhalb der Parzellen orientiert. Damit entstehen innerhalb der Teilstücke weitere Stratifizierungen.

Die Umsetzung der Varianten des N-Düngungsversuchs und erste Ergebnisse sind im Abschnitt 4 und für den Bodenbearbeitungsversuch im Abschnitt 5 dargelegt.

Da in beiden Versuchen in jedem Jahr nur eine Wiederholung der Varianten in Kombination mit dem Fruchtfolgeglied vorhanden ist, können diese Varianten je Jahr nur mittels der Methoden der beschreibenden Statistik verglichen werden. Die vorgenommenen wiederholten Messungen innerhalb der Teilstücke durch Kartierung der Ertragsdaten (siehe weiter unten in

diesem Abschnitt) verändern diese Situation nicht (siehe auch den Beitrag von C. RICHTER in diesem Band). Erst nach einer oder mehreren vollständigen Rotation(en) kann auf der Basis kumulativer Berechnungen je Teilstück und Schlag mit Methoden der Inferenzstatistik gearbeitet werden. Diese Analysen stehen im Projekt noch aus; bisher wurde generell nur mit beschreibenden Verfahren gearbeitet. So wird z. B. in den Abbildungen 13 bis 16 die Wirtschaftlichkeit der N-Düngungs- und Bodenbearbeitungsvarianten für eine vollständige Rotation (kumulative Betrachtung) unter Nutzung beschreibender Methoden dargestellt. Ob und in welcher Form bei den Auswertungen die Stratifizierungen durch Bodenscannerklassen und variierte Grunddüngung berücksichtigt werden sollen oder müssen, bedarf noch weiterer Klärungen.

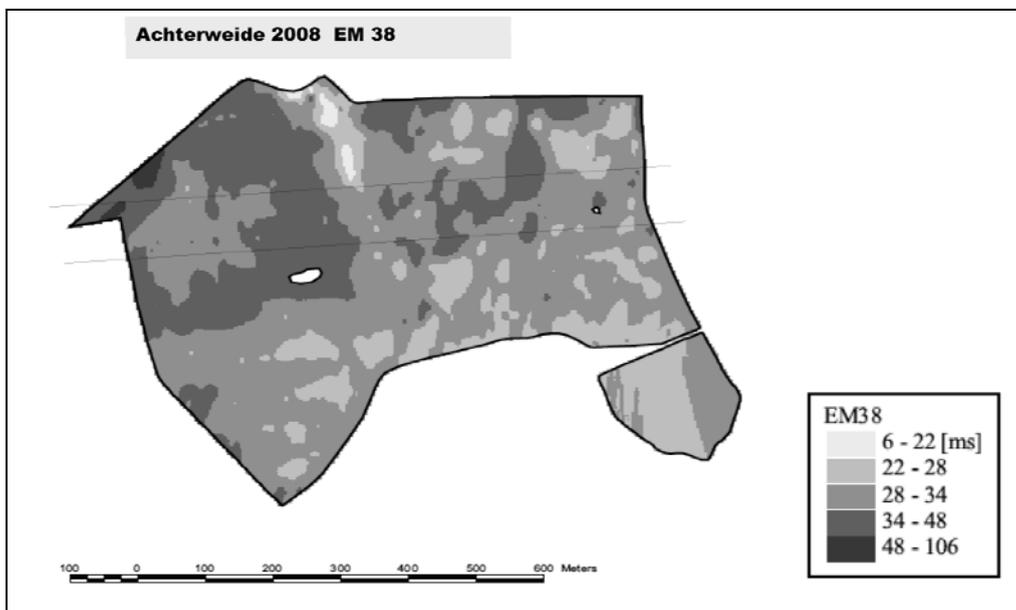


Abbildung 2: Übersicht über die EM 38 Kartierung des Projektschlages Achterweide zur Erprobung differenzierter Bodenbearbeitungsverfahren

Für die Bewertung der teilflächenspezifischen Bewirtschaftung stehen hier immer ganze Produktionssysteme zum Vergleich. Die Auswahl und die inhaltliche Ausgestaltung der Maßnahmen in den einzelnen Produktionssystemen („Bestvarianten“) erfolgt:

- auf der Basis mehrjähriger validierter Exaktversuchsergebnisse aus dem klassischen ein- und mehrfaktoriellen (Parzellen)Versuchen in der Region bzw. unter vergleichbaren Anbaubedingungen oder
- ggf. durch Modifizierung und Anpassung unter Zuhilfenahme von Betriebsleiter- und Expertenwissen (z. B. Festlegung teilflächenspezifischer Aussaatmengen in Anpassung an die Bodenklassen leicht – mittel – schwer).

An dieser Stelle sei mit Nachdruck darauf hingewiesen, dass diese Vergleichsverfahren vorausgehende und nachgelagerte Exaktversuche nicht ersetzen können! Man ist damit bestenfalls in der Lage (großtechnisch) drei Bestvarianten zu vergleichen, oder bei einfaktorieller Fragestellung – dieses Versuchsdesign erfreut sich für „Entscheidungen“ zur Sortenwahl auf

Praxisbetrieben zunehmender Beliebtheit – für drei bis fünf Sorten eine „betriebsspezifische“ Auswahl zu treffen. Dabei darf nicht vergessen werden, dass für eine sinnvolle und möglichst treffsichere Auswahl dieser drei bis fünf Prüfglieder (Verfahren oder Sorten) Vorleistungen aus Exaktversuchen nötig und zurzeit auch noch vorhanden sind. Sonst stochert man z. B. bei ca. 200 zugelassenen Weizensorten im Nebel, wenn sie nicht im Vorscreening der Landessor-tenversuche auf aussichtsreiche Kandidaten in der jeweiligen Anbauregion verdichtet wurden.

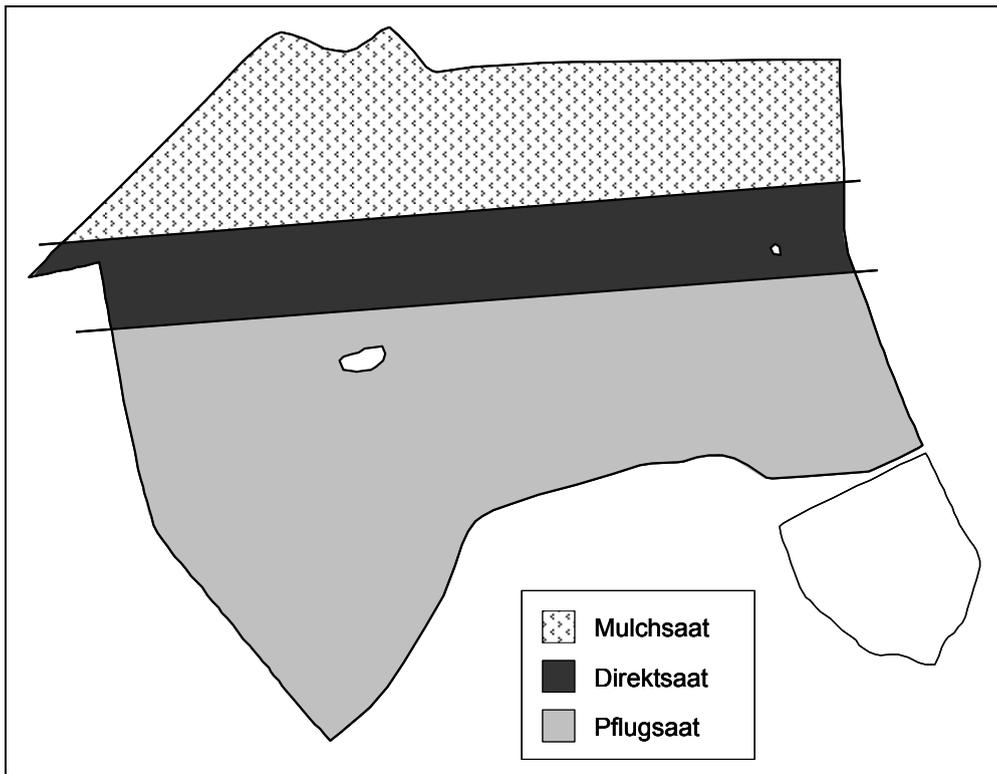


Abbildung 3: Lage der Teilstücke im Bodenbearbeitungsversuch am Beispiel Achterweide

Die Erhebungen und Versuchsdaten aller Versuchsschläge werden detailliert in einer handgeführten Ackerschlagkartei (Übertragung aus der betrieblichen Ackerschlagkartei) geführt. Nur so wird eine exakte Dokumentation aller Maßnahmen in jeder Versuchsvariante und im Jahresverlauf gewährleistet (Übersicht 1). Diese Daten sind Basis für alle weiteren Bewertungen der Ergebnisse (z. B. Nährstoffbilanzen, Wirtschaftlichkeit der Produktionssysteme etc.).

Schlag:	Reuterkoppel (42 ha)	Frucht::	SWW Inspiration	Versuch:	DG
Erntejahr:	2010	Vorfrucht:	WW		
Datum	Maßnahme			EC	
	Betriebsüblich	Einmaldüngung	Sensor		
06./07.08.09 07.08.09 07./08.08.09 08./09.08.09 27./28.08.09 21./22.09.09 21./22.09.09 23.09.09 24.09.09 28.09.09 02.10.09 01.11.09	Klärschlamm ohne Kalkanteil Kanten abschlegeln Spurlockerer 1. Stoppelbearbeitung Centaur 5 – 8 cm 2. Stoppelbearbeitung Centaur 7 – 10 cm Saatvorbereitung Centaur 12 – 15 cm Aussaat Rapid 255 K/m ² (100 %) = 152 kg, TKG 56.2 (100 %, 150 %, 200 % 12 m Walze 360 kg/ha Triple Phosphat per GPS 136 kg/ha PK 7/40 per GPS 0.6 l Herold, 33 l AHL = 12 N 5 kg Bittersalz, 1.25 kg Mangansulfat, 33 l AHL = 12 N				
24.03.10		574 kg/ha Alzon S 40 + 6 = 230 N		25	
25.03.10	375 kg/ ha Hydro Sulfan 24-6 90 N		375 kg/ ha Hydro Sulfan 24-6 90 N	25	
29.03.10			3.5 ha mit 500 g/ha Atlantis + 1 l FHS + 30 l AHL = 11 N	25	
12.04.10	302 kg/ha KAS = 82 N		220 kg/ha KAS = 59 N	27	
15.04.10 26.04.10 10.05.10 02.06.10	5 kg BS, 2 kg Mn, 0.2 gk Kupfer soll WG, 0.1 l Tallius, 30 l AHL = 11 N 1.6 l CCC, 0.3 l Medax top 0.5 l CCC + 0.3 l Mdeax top + 0.4 l Champion + 0.4 l Input + 0.1 l Tallius + 100 g Hoestar super 0.5 l Diamant, 0.25 l Input, 0.25 l Achat, 0.3 l Medax top, 3.3 kg BS			29 30 32 39	
03.06.10 17.06.10	243 kg KAS = 66 N	0.75 l Prosaro	197 kg KAS = 53 N	39 55	
Herbst-N (kg/ha)	24	24	24		
Frühjahrs-N (kg/ha)	249	241	218		
N-Gesamt (kg/ha)	273	265	242		
Ertrag (dt/ha)	89.4	95.5	94.4		
Protein (%)	12.9	12.9	13.9		
N-Entzug (kg N/ha)	174	186	198		
N-Bilanz (kg N/ha)	99 (75 ohne Herbst-N)	79 (55 ohne Herbst-N)	44 (20 ohne Herbst-N)		

Übersicht 1: Dokumentation (Ackerschlagkartei) der produktionstechnischen Daten der Versuchsschläge und –varianten an einem Beispiel

Neben den Produktionsdaten werden die Erntedaten aus den exakten Wägungen der Versuchsvarianten (auf Basis der Teilstücke, Daten aus dem Wiegebuch) und aus der Ertragskartierung (Daten vom Mährescher) erhoben und den Bodenscanner-/Aussaatklassen zugeordnet. Die kartierten Ertragsdaten sind als wiederholte Messungen innerhalb der Teilstücke zu verstehen, wobei ihre Zuordnung zu den Aussaatklassen und Grunddüngungsrasterzellen (siehe Abschnitt 3) gegebenenfalls zu berücksichtigen ist. Die Qualitätsdaten (Protein- bzw. Ölgehalte) werden als Durchschnittsproben aus den Teilstücken im Ernteverlauf an der Fuhrwerkswaage erhoben (Übersicht 2 und Abbildung 4).

Schlag:	Dwerjahren (35 ha)	Frucht:	SWW Inspiration	Versuch:	DG		
Erntejahr:	2009	Vorfrucht:	WW				
Daten aus dem Wiegebuch							
<i>Variante</i>	<i>Erntemenge Wiegebuch korrigiert (kg)</i>	<i>Flächenanteile GIS (ha)</i>	<i>Teilerträge GIS (dt/ha)</i>	<i>relativer Teilertrag (%)</i>			
Vorgewende	141704	13.9	101.9	90			
Betrieb	49938	4.4	113.5	100			
Einmal	106367	9.4	113.2	100			
Sensor	72171	6.8	106.1	94			
	<i>Gesamtfläche GIS (ha)</i>	<i>Gesamtertrag GIS (dt/ha)</i>	<i>Gesamtfläche Kock (ha)</i>	<i>Gesamtertrag Kock (dt/ha)</i>			
	34.5	107.30	34.59	107.02			
Daten vom Mähdrescher (gefiltert: 7197 von 8432 Werten verwendet, Feuchte: 1-25, Geschwindigkeit: 0-15, Schneidwerksbreite: nur voll)							
<i>Variante</i>	<i>Aussaat (%)</i>	<i>Fläche (ha)</i>	<i>Ertrag (dt/ha)</i>	<i>Anzahl</i>	<i>Min (dt/ha)</i>	<i>Max (dt/ha)</i>	<i>Standardabweichung (dt/ha)</i>
Vorgewende	100	4.0	91	767	11	160	29
	150	4.6	102	936	13	152	26
	200	5.4	103	1067	10	156	23
Betrieb	100	2.7	113	563	11	151	20
	150	1.5	118	317	11	144	15
	200	0.1	124	28	108	137	8
Einmal	100	7.4	116	1549	10	160	13
	150	1.7	114	363	32	147	14
	200	0.2	116	48	49	143	14
Sensor	100	3.7	113	815	10	159	15
	150	2.9	111	675	21	158	16
	200	0.3	88	69	12	137	31
<i>Variante</i>	<i>Ertrag*Fläche</i>	<i>Fläche (ha)</i>	<i>gew. Ertrag (dt/ha)</i>	<i>rel. Ertrag</i>			
Vorgewende	360	14.0	99	86			
	469			86			
	559			86			
Betrieb	305	4.3	115	100			
	177			100			
	18			100			
Einmal	859	9.4	116	100			
	198			100			
	26			100			
Sensor	415	6.8	111	97			
	321			97			
	25			97			
	3732	34.5	108				
Analysen							
<i>Variante</i>	<i>TKG FuKa</i>	<i>Feuchte FuKa</i>	<i>HL FuKa</i>	<i>Protein Agrolab</i>			
Betrieb	51.8	16.3	81.2	11.7			
Einmal	49.4	16.3	81.0	11.6			
Sensor	53.6	16.1	78.9	10.9			

Übersicht 2: Übersicht über die tabellarisch erhobenen Versuchsdaten – Mähdrescher an einem Beispielsschlag

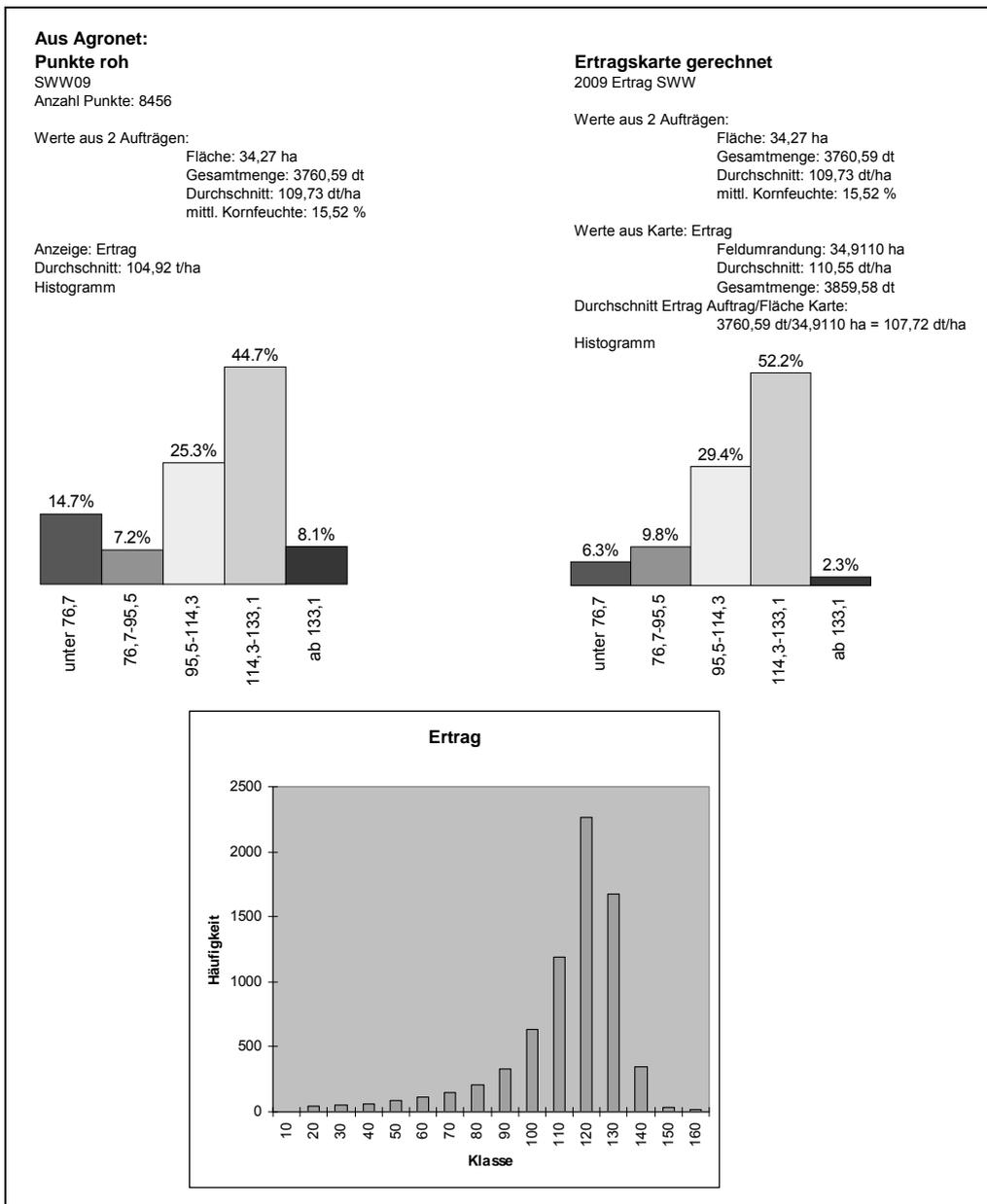


Abbildung 4: Beispiele für die grafische Auswertung erhobener Versuchsdaten – Mähdrescher an einem Beispielschlag

Um neben der Bewertung auf der Basis der Teilstücke kleinteilige Informationen zum Zusammenhang zwischen dem „Erfolg“ der Versuchsvariante und den Bodenscanner-/Aussaatklassen zu bekommen, werden bei Getreide zusätzlich Quadratmeterproben zum Verlauf der Ertragsbildung, zum Ertrag und zur Qualität erhoben (Abbildung 5, Übersicht 3). Die Ergebnisse dieser Quadratmeterproben sollen eine wesentliche Basis für die noch ausstehende statistische Bewertung und Interpretation der Versuchsergebnisse bilden.

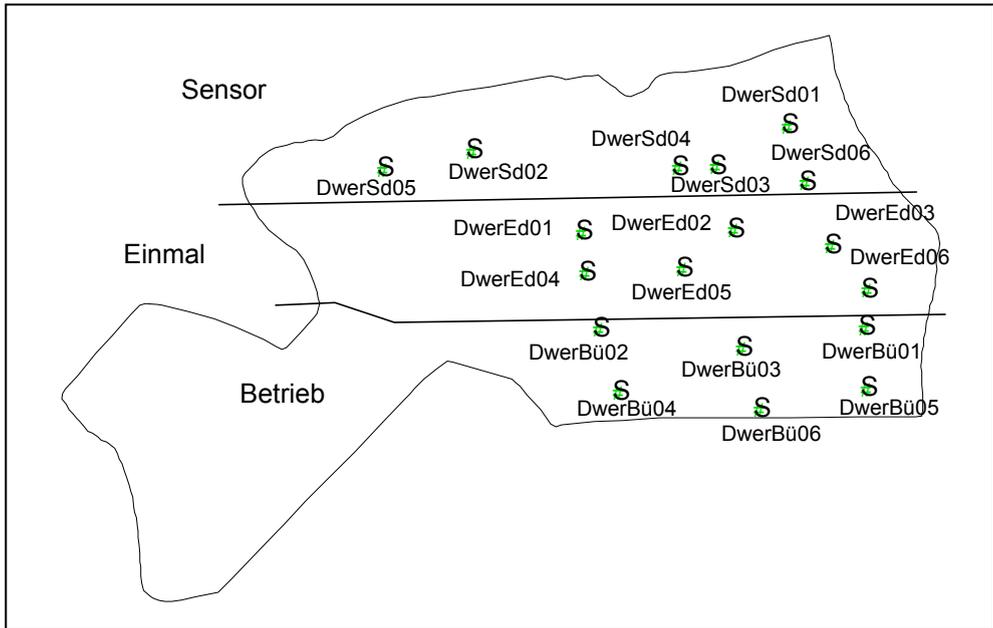


Abbildung 5: Lage der Quadratmeterproben innerhalb der Teilstücke im Beispielsschlag Dwerjahren

Schlag: Erntejahr:	Dwerjahren (35 ha) 2009		Frucht: Vorfrucht:	SWW Inspiration WW		Versuch: DG	
<i>Probe</i>	<i>Relief</i>	Aussaat 100 % = 250 Kö/m ² , 150 kg/ha (%)		Pflanzenanzahl im Frühjahr (pro m ²)		Ährentragende Halme zur Ernte (pro m ²)	
Betrieb							
DwerBü01	leichter Hang	100		206		627	
DwerBü02	Hang	100		172		600	
DwerBü03	Ebene	200		268		603	
DwerBü04	Kuppe	150		218		447	
DwerBü05	Ebene	150		239		563	
DwerBü06	leichter Hang	100		224		434	
Einmal							
DwerEd01	Hang	100		188		578	
DwerEd02	Kuppe	100		188		600	
DwerEd03	Hang	100		208		635	
DwerEd04	Hang	100		172		583	
DwerEd05	leichter Hang	150		222		610	
DwerED06	Ebene	150		216		661	
Sensor							
DwerSd01	Ebene	150		164		488	
DwerSd02	Hang	150		172		542	
DwerSd03	Kuppe	100		164		461	
DwerSd04	Kuppe	150		194		490	
DwerSd05	Hang	100		190		528	
DwerSd06	Ebene	100		144		491	
<i>Probe</i>	<i>Ertrag</i> FuKa (dt/ha)	<i>Feuchte</i> FuKa (%)	<i>TKG</i> FuKa (g)	<i>HL</i> FuKa (kg)	<i>TKG</i> feuchte- korrigiert (g)	<i>HL</i> feuchte- korrigiert (kg)	<i>Protein</i> Agrolab (% i.d. TS)
Betrieb							
DwerBü01	144.8	15.9	52.4	79.4	51.2	80.4	10.6
DwerBü02	137.2	15.6	50.4	79.8	49.5	80.6	11.6
DwerBü03	139.9	14.6	53.6	80.2	53.2	80.5	10.9
DwerBü04	120.2	14.8	52.2	80.7	51.7	81.1	11.5
DwerBü05	104.4	15.6	52.2	79.2	51.2	80.0	11.0
DwerBü06	134.6	15.4	54.4	81.4	53.5	82.1	11.3
Einmal							
DwerEd01	139.8	15.2	53.4	80.2	52.7	80.8	10.9
DwerEd02	129.9	15.3	49.6	80.1	48.9	80.8	11.0
DwerEd03	122.5	15.5	51.0	79.2	50.1	80.0	10.5
DwerEd04	128.6	15.4	55.0	77.5	54.1	78.2	11.2
DwerEd05	121.3	15.6	53.8	76.8	52.8	77.6	10.8
DwerED06	141.0	15.0	51.6	80.6	51.0	81.1	11.9
Sensor							
DwerSd01	133.0	15.5	56.4	80.5	55.4	81.3	11.5
DwerSd02	129.5	15.6	54.2	80.4	53.2	81.2	11.0
DwerSd03	127.4	15.6	55.0	80.7	54.0	81.5	11.1
DwerSd04	144.0	15.5	54.8	81.1	53.8	81.9	11.7
DwerSd05	126.5	15.3	53.0	81.2	52.2	81.9	11.9
DwerSd06	123.7	14.9	52.6	80.8	52.0	81.3	11.8

Übersicht 3: Erhobene Versuchsdaten – Quadratmeterproben im Beispielschlag Dwerjahren

2 Teilflächenspezifische Aussaat

Um zu Beginn des jeweiligen Anbaus eine an die Bodenunterschiede innerhalb der Schläge angepasste Saatmenge teilflächenspezifisch auszubringen, wurden den auf die jeweilige „Bodengüte“ angepassten drei Bodenscannerklassen (schwer, mittel, leicht) Aussaatmengen zugeordnet. Dabei bildete die optimale Standardsaatmenge zur jeweiligen Sorte (Getreide, Raps) und zum jeweiligen Saattermin die "Aussaatklasse 100 %" (leichte Bodenteilstücke). Aus den betrieblichen Erfahrungen wurden für die mittleren Bodenteilstücke 150 % der Standardsaatmenge und für die schweren Bodenteilstücke (Lehm- und Tonkuppen) 200 % der Standardsaatmenge gewählt (Abbildung 6).

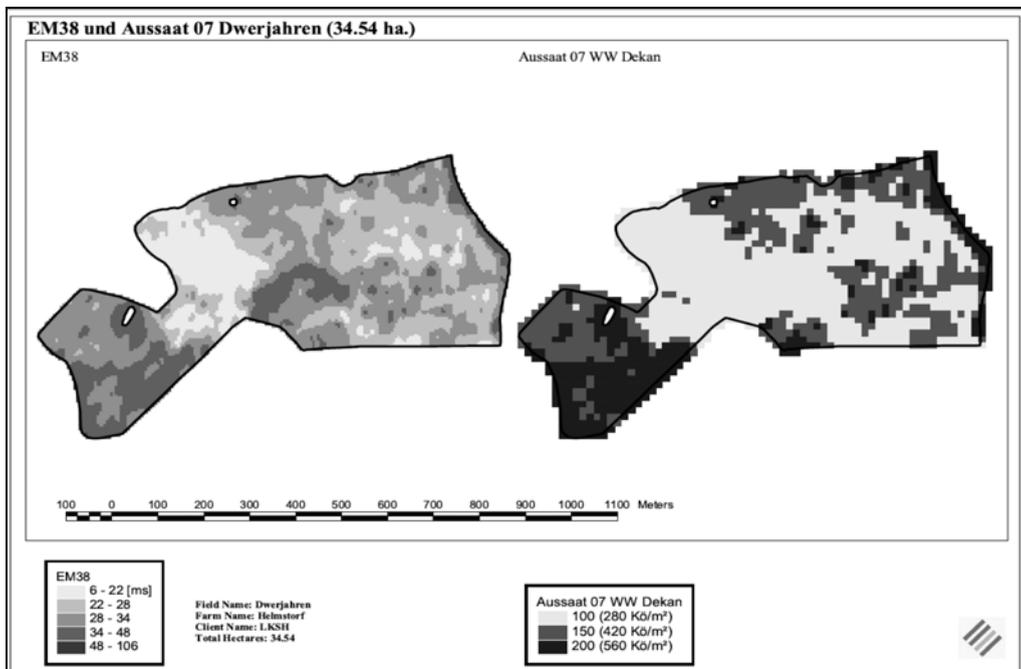


Abbildung 6: Bodenscannerklassen aus EM 38 und zugeordnete Saatmengen
(Beispiel für eine Aussaatkarte für die teilflächenspezifische Aussaat)

Für letztere galt z. B. für Getreide die Betriebserfahrung – ein Saatkorn gleich ein ährentragender Halm – als Bemessungsgröße für den Einstieg in die teilflächenspezifische Aussaat. Nach drei Versuchsjahren wurden auf Grund der Erhebungen (Aufgangszahlen, ährentragende Halme) auf den Rasterzellen der jeweiligen Bodenscannerklassen die mittlere Saatmenge auf 140 % und die Saatmenge für die schweren Bodenteilstücke auf 180 % angepasst. Die Aufgangszahlen aus der schwierigen Herbstsaat 2011 zeigten mit den an die Bodenscannerklassen angepassten differenzierten Saatmengen für alle drei Bodenscanner-/Aussaatklassen die gleiche Keimpflanzendichte.

Der mittel- und langfristige Erfolg der teilflächenspezifischen Aussaat wird sich u. a. am schlagspezifischen Ertragsverlauf (Ertragshöhe, Ertragsicherheit über die Jahre) messen lassen. Kurzfristigen „Erfolg“ signalisieren die Aufgangszahlen (Keimpflanzendichte) und die dazugehörigen Ertragshebungen auf den der jeweiligen Bodenscannerklasse zugeordneten

Rasterzellen. Hierüber lässt sich auch erst mittel- und langfristig eine Verbesserung über dieses automatisierte Verfahren der Saatmengenanpassung auf die Bodenfruchtbarkeit auf den bisher oft mit zu geringem Pflanzenbestand (und demzufolge ohne ausreichende Humusversorgung über das geerntete Stroh) ausgestatteten Lehm- und Tonkuppen erwarten.

Am Beispiel der Versuchsschläge aus 2009 und 2010 werden die differenzierten Saatmengen anhand der Flächenanteile der Bodenscannerklassen leicht – mittel – schwer auf den einzelnen Versuchsschlägen deutlich (Abbildung 7).

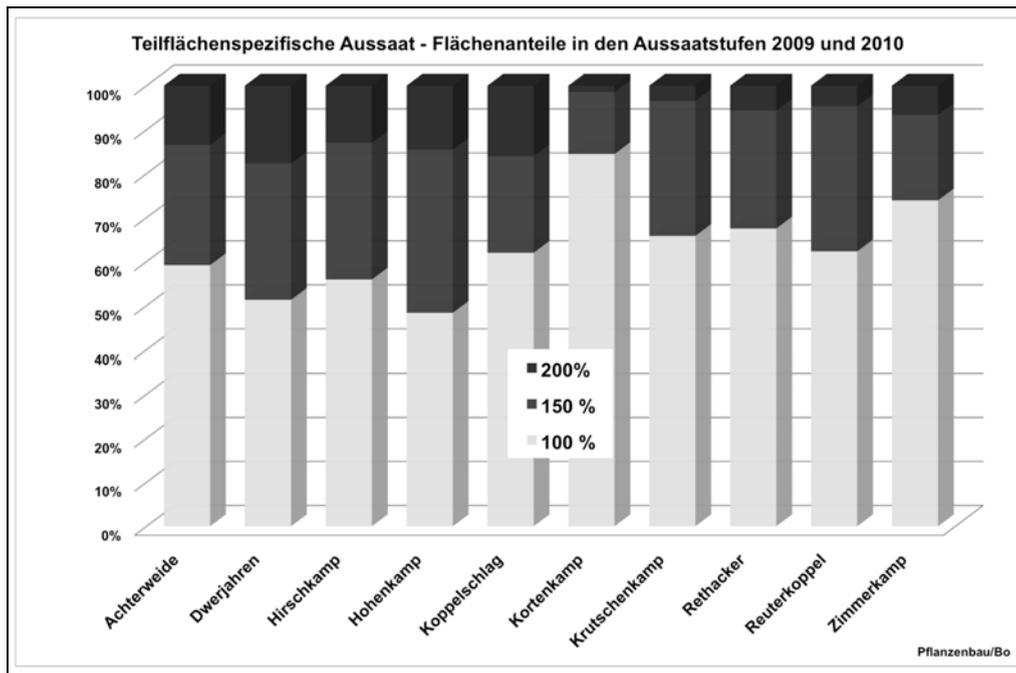


Abbildung 7: Flächenanteile für die an die Bodengüte angepassten Aussaatmengen – Versuchsschläge 2009 und 2010

Die durch die teilflächenspezifische Aussaat entstehenden Saatgutmehrkosten – hier am Beispiel des Jahres 2010 – sollten mittel- und langfristig durch die o. g. Faktoren der Ertragsentwicklung aufgefangen werden (Abbildung 8). Einen Vergleich für die teilflächenspezifische Aussaat vs. schlageinheitliche Saatmenge gibt es in diesen Versuchen nicht. Letzteres war auch ohne das über die jetzt vorhandenen Aussaatkarten automatisierte Aussaatverfahren nicht Stand der Technik im Betrieb. Bisher erfolgte die Saatmengenanpassung während der Aussaat manuell durch den Schlepperfahrer über die Plus/Minus-Taste am Drillcomputer im „Sichtmodus“ zur Unterscheidung schwerer und leichter Schlagabschnitte. Das führte vor allem bei langer Tagesarbeitszeit und nachts zu unbefriedigenden Ergebnissen. Die automatisierte teilflächenspezifische Aussaat ist nach Erkenntnissen aus Exaktversuchen und aus der oben geschilderten mehrjährigen Betriebserfahrung vor Ort inzwischen die im Gesamtbetrieb gesetzte Basis für die weiteren schlagspezifischen Produktionsmaßnahmen.

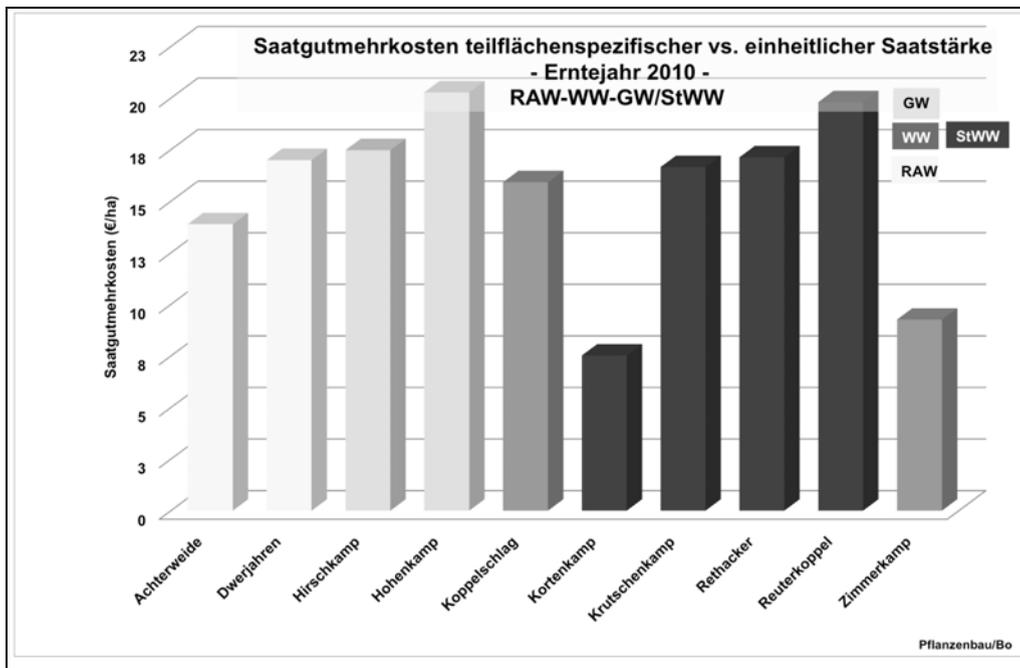


Abbildung 8: Saatgutmehrkosten teilflächenspezifische Aussaat vs. schlageinheitlicher Saatmengen am Beispiel der Versuchsschläge 2010 (RAW-Winterraps, WW – Winterweizen, WG – Wintergerste, StWW – Stoppelweizen)

3 Teilflächenspezifische Grunddüngung

Um zur Erprobung der teilflächenspezifischen Grunddüngung auf die aus der EM 38 Kartierung bekannte Bodenheterogenität auf den Versuchsschlägen ausreichend zu reagieren, wurde die Bestimmung der Versorgung mit den Grundnährstoffen P, K, Mg und der pH-Wert auf der Basis einer 1-ha-Raster Beprobung ermittelt (Abbildung 9). Die Beprobung wurde auf Empfehlung des Landwirtschaftlichen Labors Dr. Janssen GmbH und des Maschinenrings Wagrien e. V. auf der Basis der Kreisbeprobung (Durchmesser je Rasterzelle 25 Meter) durchgeführt.

Die Schlag- und Düngungshistorie des Betriebes, gekennzeichnet durch

- mehrjährigen Einsatz kalkstabilisierten Klärschlammes mit inzwischen zu hohen pH-Werten und zunehmender Gefahr der P-Festlegung und
- zurückliegend kostenextensive Grunddüngung im P-Bereich mit unzureichenden P-Gehalten auf einzelnen Schlagteilen,

realisierte auf den untersuchten Versuchsschlägen ein deutlich differenziertes Nährstoffraster, das für eine zukünftige Anpassung der Versorgung über eine teilflächenspezifische Grunddüngung bestens geeignet ist (Abbildung 10).

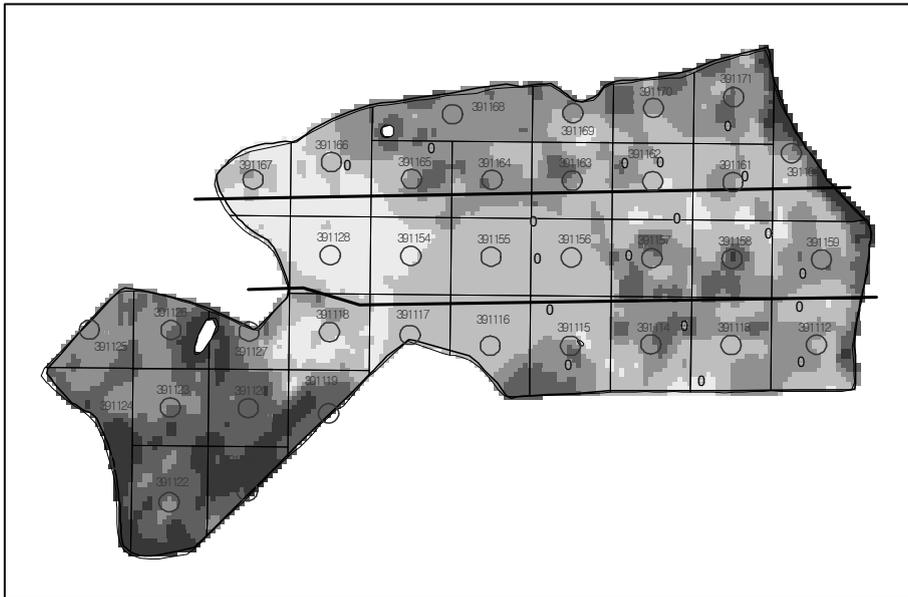


Abbildung 9: Bodenproben auf Basis 1ha-Raster – Versuchsschlag Dwerjahren

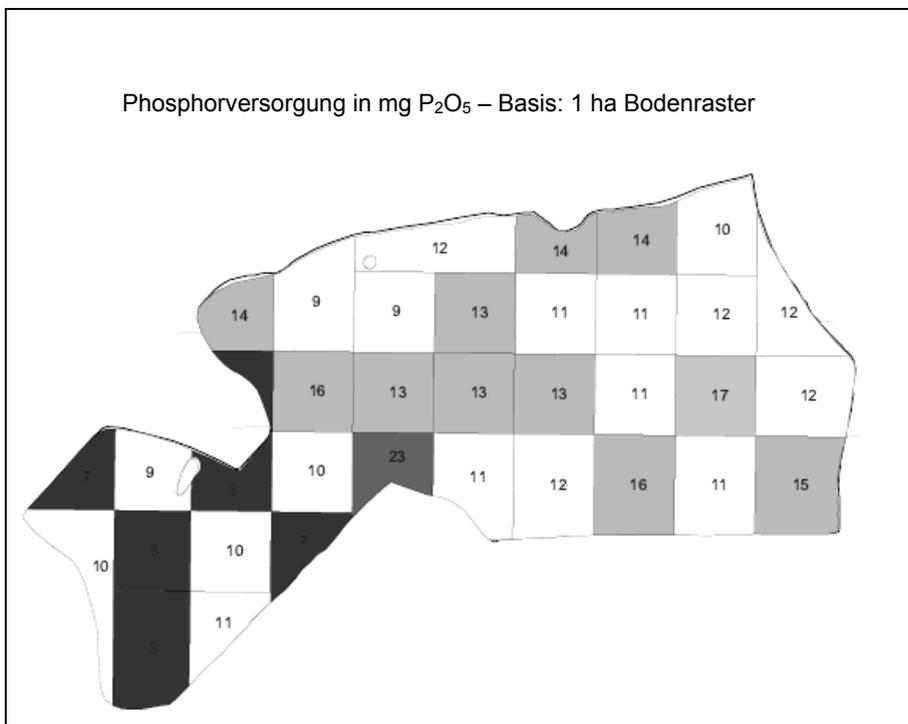


Abbildung 10: Teilflächenspezifischer Gehalt an P_2O_5 – Versuchsschlag Dwerjahren

Die Bestimmung der Bodenart erfolgte, wie i. d. R. generell empfohlen, mit der Fingerprobe, was sich im Nachhinein als problematisch herausstellte. So ergab die Fingerprobe für alle Versuchsschläge als Bodenart schlageinheitlich sL/uL, was nach der aus der EM 38 Kartierung abzuleitenden Bodenheterogenität nicht plausibel sein konnte. Deshalb wurde vorerst probeweise für einen Versuchsschlag (Koppelschlag) im Vergleich zur Fingerprobe eine Texturbestimmung über die Schlämmanalyse vorgenommen, mit einem von der Fingerprobe deutlich abweichenden Ergebnis (Abbildung 11).

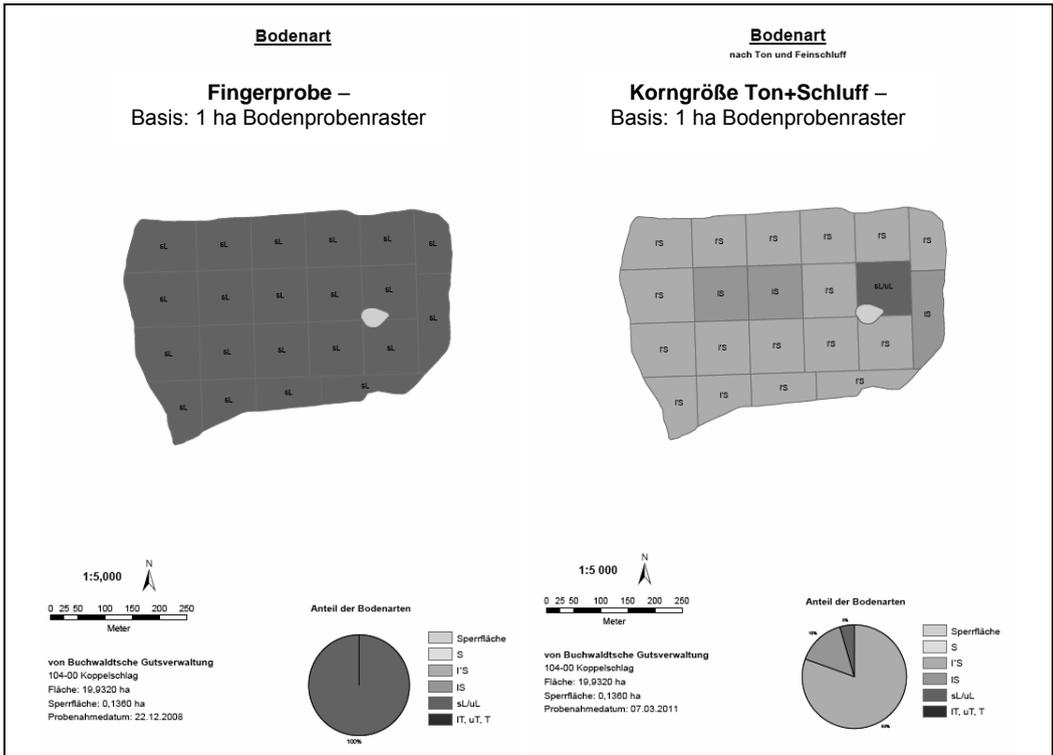


Abbildung 11: Bestimmung der Bodenart nach Fingerprobe (links) und nach Schlämmanalyse nach Ton und Feinschluff (rechts) – Versuchsschlag Koppelschlag

Das hat erhebliche Konsequenzen für die teilflächenspezifische Düngungsempfehlung für K_2O , MgO und CaO , die bekanntermaßen in Abhängigkeit von den Versorgungsstufen an die Bodenart angepasst werden muss. Folgt man der Düngungsempfehlung nach Fingerprobe, findet z. B. eine hoffnungslose Überkalkung der Teilflächen und damit des gesamten Schlages statt. Auch der an die tatsächliche Bodenart angepasste Düngerbedarf für K_2O und MgO wird teilflächenspezifisch nicht getroffen und führt zur Überdüngung verbunden mit insgesamt deutlich höheren Düngekosten (Anmerkung: Berechnungsdetails für die Schläge und die Hochrechnung auf den 1000-Hektar-Betrieb werden in kürze in weiterführenden Fachveröffentlichungen zum Projekt OFR dargestellt).

Abgesehen von den begrenzten technischen Möglichkeiten mit einem 36 m Schleuderstreuer teilflächenspezifisch an das Bodenprobenraster angepasst zu düngen (auch hier muss an dieser Stelle auf weiterführende Fachveröffentlichungen zum Projekt OFR verwiesen werden) muss hier folgendes Fazit gezogen werden:

- Eine teilflächenspezifische Düngungsempfehlung ohne Bestimmung der Bodenart der Teilfläche, insbesondere wenn es sich um heterogene Standorte handelt, führt zu falschen Düngungsempfehlungen und gegebenenfalls zu unnötiger Kostenbelastung des Düngerkontos im Betrieb.
- Unter solchen Bedingungen ist die bei der Laboruntersuchung zu den Nährstoffgehalten mitgelieferte „kostenlose“ Fingerprobe zur Bestimmung der Bodenart letztendlich wertlos bzw. fachlich falsch!

4 N-Düngungsversuch – seine Varianten und erste Ergebnisse

Um die N-Düngung teilflächenspezifisch zu gestalten, bedient man sich in der praktischen Umsetzung vor allem sensorgestützter Lösungen. Im Projekt kommt bis zum derzeit laufenden fünften Projektjahr der Yara-N-Sensor der Firma agricon zum Einsatz. Neben der Sensor-Variante wurde die Düngung „Betriebsüblich“ und in einer Einmaldüngungsvariante mit stabilisiertem Stickstoff – nicht teilflächenspezifisch zu Vegetationsbeginn ausgebracht – erprobt. Die beschriebene Dreiteilung der Versuchsschläge und anschließende Zuordnung der Varianten ist beispielhaft in Abbildung 12 dargestellt, wobei je Versuchsschlag die Varianten den Teilstücken randomisiert zugeordnet wurden, aber über die Jahre und Fruchtarten mit konstanter Schlagteilung gearbeitet wird, um über die Projektlaufzeit möglicherweise auflaufende kumulative Effekte der differenzierten N-Düngungsstrategien zu erfassen.

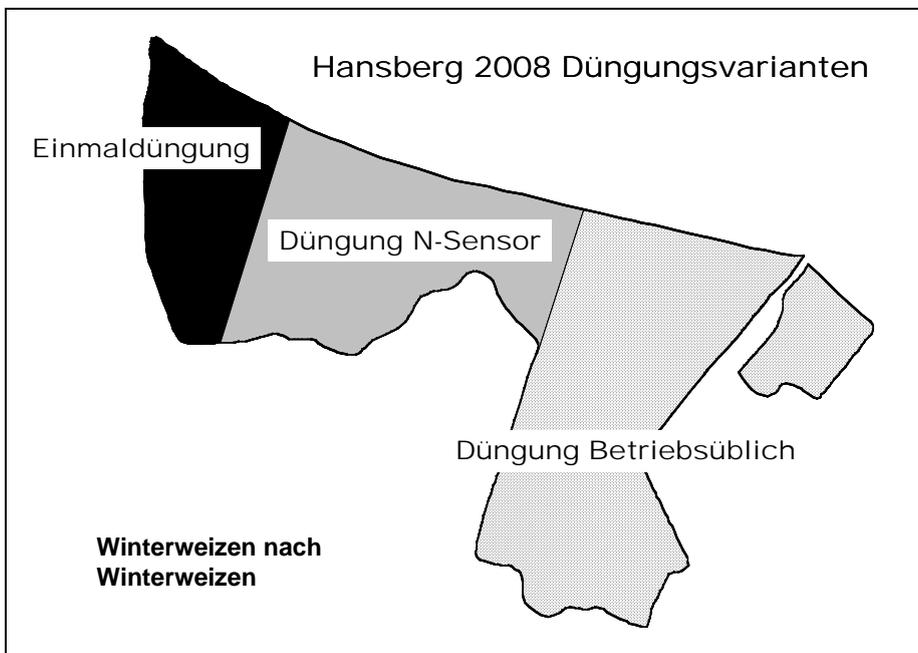


Abbildung 12: Lage der Teilstücke im N-Düngungsversuch am Beispiel Hansberg 2008

Es sei an dieser Stelle explizit darauf hingewiesen, dass die Variante „Betriebsüblich“ nicht konstant gedüngt wird, sondern der Streuerfahrer auf Grund seiner langjährigen Erfahrung und detaillierten Schlagkenntnis „steuernd“ über die Plus/Minus-Taste bei der Ausbringmenge in Abhängigkeit vom Bestandesbild und den aktuell befahrenen Schlagteilen (Lehmkuppe, hu-

mose Senke) entsprechend eingreift. Es war beim Strategievergleich nicht das Ziel, den Betrieb „in die Steinzeit zurück zu versetzen“ und betriebsüblich ein flächenkonstantes Streuen im Vergleich zum N-Sensor zu organisieren, wie das bei der Darstellung diverser Ergebnisse zum N-Sensor oft getan wird.

Auf methodische Details beim Einsatz des N-Sensors, Einzelergebnisse des im Projekt intensiv durchgeführten N-Monitorings, Probleme des Sensoreinsatzes und der daraus abgeleiteten N-Empfehlungen in Abhängigkeit vom Witterungsverlauf, der N-Aufnahme des Pflanzenbestandes, z. B. beim Raps, die Möglichkeit der sicheren teilflächenbezogenen Platzierung der empfohlenen N-Menge und der erreichten Erträge, Qualitäten und N-Bilanzen kann hier im Einzelnen nicht eingegangen werden. Auch hier sei an dieser Stelle auf weiterführende Fachveröffentlichungen zum Projekt OFR und die Ergebnisdarstellung im Internet unter der Adresse www.lksh.de/landwirtschaft/pflanze/ unter "On Farm Research" verwiesen. Gleiches gilt für Methodik, Inhalte und Ergebnisse der Variante „Betriebsüblich“ und „Einmaldüngung“.

Die Ergebnisse subsumieren sich letztendlich in einer Bewertung der Wirtschaftlichkeit der verglichenen Düngungsstrategien. Hierzu wurde nach aufwendiger Datenerhebung im Projektbetrieb und in Anpassung und Abgleich mit vergleichbaren Betrieben aus der Marktfuchtberatung Schleswig-Holsteins über:

- die Leistung (Ertrag, Qualität, Preis),
- die Direktkosten (Saatgut, Düngemittel, Pflanzenschutzmittel, Hagelversicherung, Zinsansatz Umlaufvermögen) und
- die Arbeitserledigungskosten (Maschinen- und Personalkosten für Bodenbearbeitung, Aussaat, Pflanzenschutz, Düngung, Ernte, Management, Zinsansatz Maschinenvermögen, Zinsansatz variable Kosten gesamt).

die Direkt- und Arbeitskostenerledigungsfreie Leistung der Produktionsverfahren ermittelt.

Für die erprobten Düngungsstrategien hat sich im Mittel des Betriebes und der angebauten Fruchtarten Raps-Weizen-Gerste die Variante „Betriebsüblich“ als das wirtschaftlichste Verfahren, gefolgt von der „Einmaldüngung“ erwiesen. Dagegen fiel die Sensordüngung im wirtschaftlichen Erfolg ab (Abbildung 13).

Für die einzelnen Fruchtarten liegen die entsprechenden Ergebnisse ebenfalls im Internet unter www.lksh.de/landwirtschaft/pflanze/ unter "On Farm Research" vor. Auch bei Raps schnitt die Variante „Betriebsüblich“ am wirtschaftlichsten ab, „Einmaldüngung“ und „N-Sensor“ lagen gleich auf. In Wintergerste lagen „Betriebsüblich“ und „Einmaldüngung“ gleich auf, die N-Sensor-Variante fiel dagegen auch hier wirtschaftlich ab. Besonders erfolgreich war die „Einmaldüngung“ in Weizen nach Weizen (Abbildung 14). Unter diesen Anbaubedingungen und den in den letzten Jahren sich veränderten Witterungsabläufen, lang anhaltende Vorsommertrockenheiten mit unzureichender N-Wirkung bei geteilten N-Gaben nehmen zu, war eine ausreichend hohe N-Versorgung zu Vegetationsbeginn unter Nutzung der aus dem Winter verbliebenen Wasserreserven ertragsentscheidend. Es ist davon auszugehen, dass sich unter diesen Verhältnissen zunehmend vereinfachte Düngungsstrategien durchsetzen werden. Wenn das Wasser in wichtigen Abschnitten des Vegetationsverlaufes fehlt, nutzt es weder etwas, mit mehrfacher Gabenteilung zu versuchen „der Pflanze ins Maul zu düngen“, noch sich mit aufwendiger N-Tester- und Sensortechnik an einem erhofften Entwicklungsverlauf der Pflanzen entlang zu hangeln.

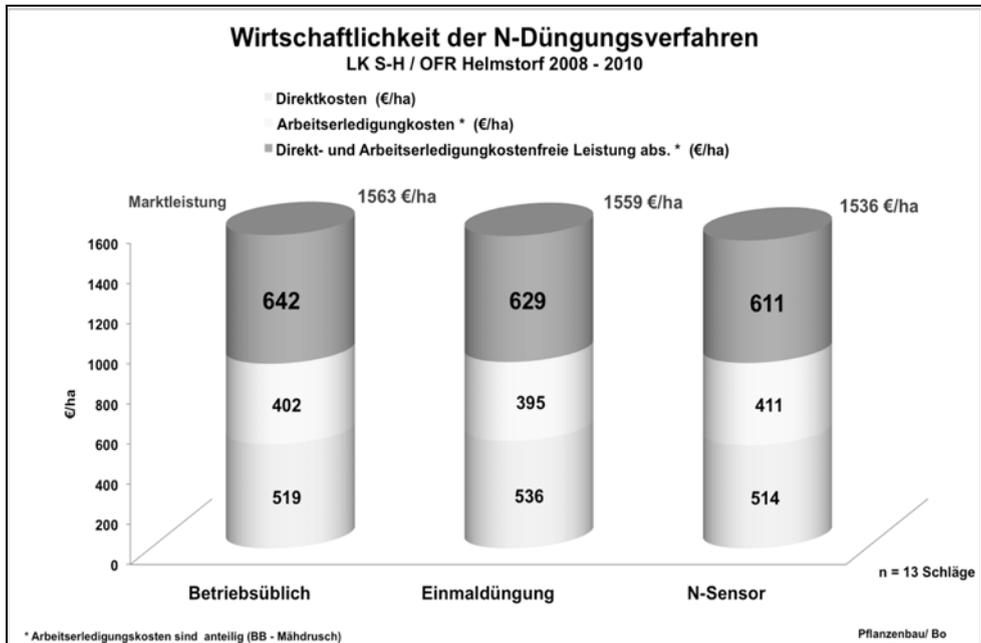


Abbildung 13: Wirtschaftlichkeit differenzierter N-Düngungsstrategien im Projekt On Farm Research – Betriebsmittel Raps - Weizen - Gerste

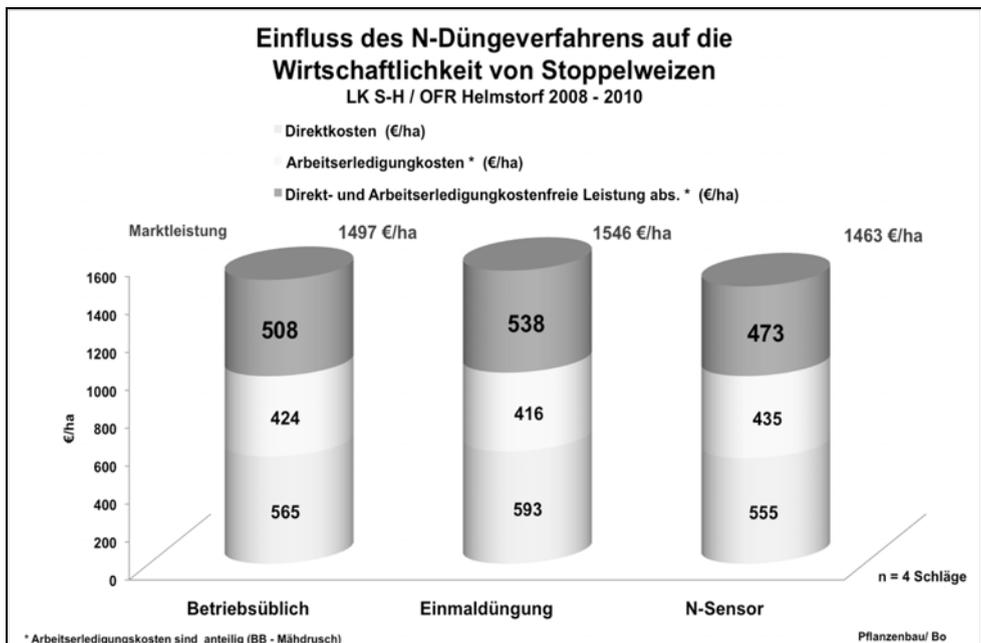


Abbildung 14: Wirtschaftlichkeit differenzierter N-Düngungsstrategien im Projekt On Farm Research - Stoppelweizen

5 Bodenbearbeitungsversuch – seine Varianten und erste Ergebnisse

Die Effizienz von Bodenbearbeitungsverfahren steht besonders auf schweren und zudem noch heterogenen Standorten im Focus zukünftiger Bearbeitungsstrategien. Im Projekt werden Pflugsaat, Mulchsaat und Direktsaat, letztere als extremste Variante extensiver Bodenbearbeitungsverfahren, mehrjährig verglichen. Dabei haben wir z. B. für die Direktsaat die betriebsübliche Fruchtfolge Raps-Weizen-Gerste bzw. Raps-Weizen-Weizen bewusst – aus Gründen der Wirtschaftlichkeit des Anbaus dieser Fruchtarten im Vergleich zu Leguminosen und anderer Sommerungen, mit der Möglichkeit Zwischenfruchtanbau einzufügen – unter den Anbaubedingungen Schleswig-Holsteins entgegen jeder Lehrmeinung nicht verlassen. Auch hier steht die Langjährigkeit der Erprobung (Projektlaufzeit zehn Jahre) zur Erschließung von positiven oder negativen Langzeiteffekten im Vordergrund. Die Bodenbearbeitungsvarianten wurden aus diesen Gründen ebenfalls stationär je Versuchsschlag angelegt (s. Abbildung 3). In den ersten beiden Jahren wurde die Direktsaatvariante mit einer Claydon-Drille etabliert, seit dem dritten Versuchsjahr läuft hier ein Seed Hawk 600 C der Fa. Väderstad. Seitdem gehört u. a. die Unterfußdüngung zur Aussaat (i. d. R. mit DAP) zum Verfahren.

Der Versuch, die Mulchsaat im Verfahrenvergleich teilflächenspezifisch, d. h. Steuerung der Bearbeitungstiefe des Grubbers in Abhängigkeit von der Bodenart, des Reliefs, der Ernte- und damit anfallenden Strohmenge, zu realisieren, scheiterte bisher an den fehlenden technischen Möglichkeiten. Ab Herbstbestellung 2012 soll es voraussichtlich dazu für den eingesetzten Kockerling Vector eine technische Lösung zur Ansteuerung der Bearbeitungstiefe über eine hinterlegte Karte mit den o. g. Faktoren geben.

Aus Platzgründen kann hier ebenfalls nicht auf die Einzelschlag- und Einzeljahresergebnisse und deren ausführliche Interpretation eingegangen werden. Details finden sich wieder unter www.lksh.de/landwirtschaft/pflanze/ unter "On Farm Research".

Auch hier subsumieren sich die Ergebnisse letztendlich in einer Bewertung der Wirtschaftlichkeit der verglichenen Bodenbearbeitungsstrategien. Im Betriebsdurchschnitt bei Betrachtung aller Bearbeitungsverfahren über alle drei Fruchtarten hat in den ersten drei Versuchsjahren der Pflugsaatbetrieb am wirtschaftlichsten abgeschnitten (Abbildung 15). Das resultiert aber nur aus dem Ertragsvorteil von Wintergerste nach Pflug im Vergleich zu Direkt- und Mulchsaat. Bei Raps lagen Pflugsaatbetrieb und Mulchsaat gleich auf, Direktsaat fiel im Mittel der ersten drei Jahre wirtschaftlich ab. Wintergerste muss unter diesen Anbaubedingungen gepflügt werden. Interessant war aber, dass hier die Direktsaat der Mulchsaat überlegen war. Weizen nach Raps, sowohl in Pflugfruchtfolge gemulcht, als auch in der durchgängigen Mulchsaat, lagen gleich auf, die Direktsaat fiel deutlicher ab. Im Stoppelweizen war in den Jahren 2008 bis 2010 die Pflugsaat der Mulchsaat überlegen, gefolgt von der Direktsaat (Abbildung 16).

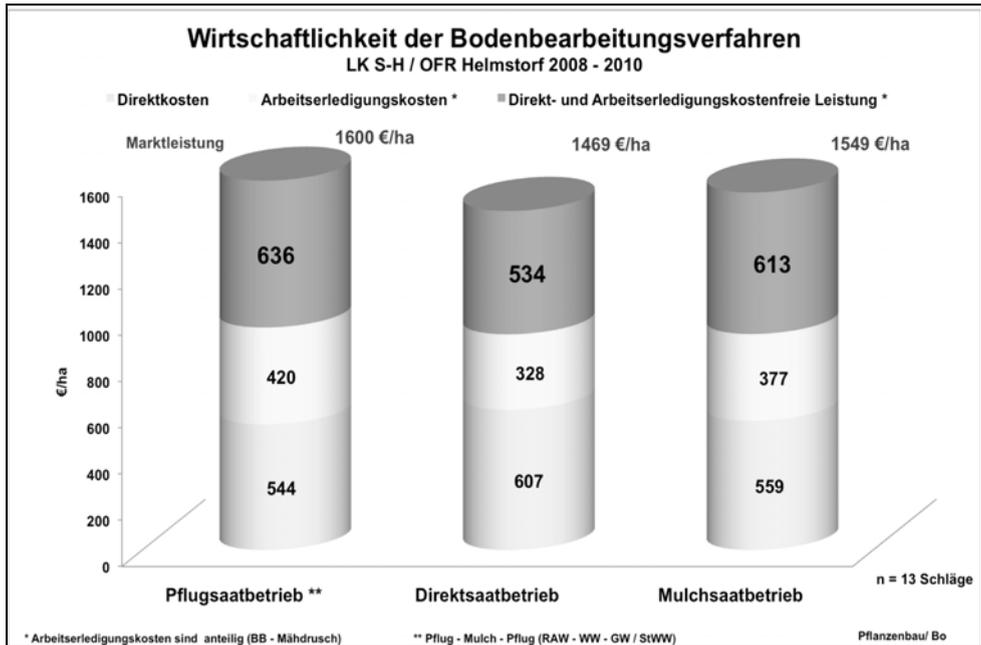


Abbildung 15: Wirtschaftlichkeit differenzierter Bodenbearbeitungsstrategien im Projekt On Farm Research – Betriebsmittel Raps - Weizen - Gerste

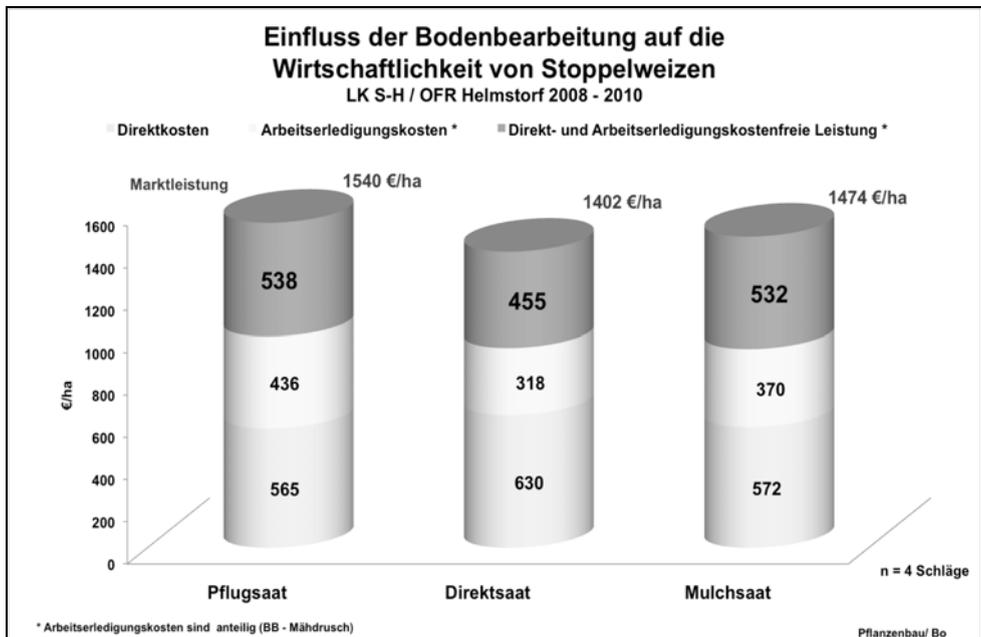


Abbildung 16: Wirtschaftlichkeit differenzierter Bodenbearbeitungsstrategien im Projekt On Farm Research – Stoppelweizen

6 Fazit und Ausblick

Im Projekt On Farm Research der Landwirtschaftskammer Schleswig-Holstein und der Gutsverwaltung Helmstorf werden in großflächigen Versuchen langjährig (Projektlaufzeit zehn Jahre) Verfahren des teilflächenspezifischen Acker- und Pflanzenbaus und differenzierte Bodenbearbeitungsverfahren erprobt und verglichen. Dabei hat sich gezeigt, dass die teilflächenspezifische Bewirtschaftung ein erhebliches Maß an technischen, personellen und monetären Voraussetzungen erfordert. Teilflächenspezifischer Acker- und Pflanzenbau kostet Geld!

Nicht in jedem Fall konnten im bisherigen Projektverlauf wirtschaftliche Vorteile teilflächenspezifischer Bewirtschaftungsstrategien nachgewiesen werden (z. B. teilflächenspezifische N-Düngung). Zukünftig wird eine Verbesserung des Verfahrens der teilflächenspezifischen N-Düngung über die Kombination von N-Sensor und Ertragspotenzialkarte angestrebt. Auch die Einmaldüngung wird in Zukunft nicht mehr konstant, sondern teilflächenspezifisch am Ertragspotenzial der Teilfläche orientiert, ausgebracht (Map Overlay).

Andere Verfahren, wie beispielsweise die teilflächenspezifische Aussaat lassen sich als fester Bestandteil des gesamten Produktionsverfahrens nicht mehr ohne Weiteres kurzfristig und monokausal im Hinblick auf ihre positiven Wirkungen bewerten. Wer teilflächenspezifische Grunddüngung auf heterogenen Standorten betreibt, muss sich hinreichend genau über die Bodenart seiner Teilflächen informieren (Schlamm-Analyse für Ton-Schluffgehalte oder Hofbodenkarte).

Der Erfolg extensiverer Bodenbearbeitungsverfahren hing entscheidend von der Bestandes-etablierung nach der Aussaat ab. Zur Direktsaat – im Projekt stehen jetzt im 5. Projektjahr hervorragende Direktsaatbestände im Feld – gehören vor allem Unterfußdüngung mit einer geeigneten Düngerform und ein erfolgreiches Strohmanagement zur Vermeidung von Acker-schnecken und Mäusen. Für eine teilflächenspezifische Mulchsaat fehlten bisher noch die technischen Voraussetzungen der Ansteuerung des Grubbers in seiner Bearbeitungstiefe in Abhängigkeit von Bodenart, Relief und Strohanfall. Ab Herbst 2012 wird dazu im Projekt eine technische Lösung erprobt.

Die inferenzstatistische Bewertung der Ergebnisse zum Vergleich teilflächenspezifischer Bewirtschaftungsverfahren steht im Projekt noch aus. Dies ist bei dem multivariaten Ansatz „teilflächenspezifisches Produktionsverfahren“ und dem sich daraus ableitenden Erfordernis, die Effekte solcher Einflussgrößen wie zum Beispiel Bodenunterschiede, differenzierte Saatmengen, teilflächenspezifische Grunddüngung u. a. im „System“ zu quantifizieren, nicht einfach.

Der vorliegende Beitrag soll auf die Probleme solcher komplexer Lösungsansätze, wie sie in der Praxis bei teilflächenspezifischer Bewirtschaftung stattfinden, aufmerksam machen. Hier geht es nicht mehr um den „Erfolg“ monokausaler Ansätze, sondern um den Versuch der Bewertung komplexer Produktionssysteme, die in ihren Einzelbestandteilen auf erfolgreichen Erkenntnissen und „Bestvarianten“ aus Exaktversuchen beruhen.

Der Projektbetrieb und die Projektergebnisse sind zur umfassenden Information für alle offen. Detailergebnisse sind im Internet unter www.lksh.de/landwirtschaft/pflanze/ unter "On Farm Research" veröffentlicht und werden fortlaufend ergänzt.

Arbeitsergebnisse des gemeinsamen Workshops "On-Farm-Experimente"

des Arbeitskreises „Koordinierung im Versuchswesen“ des Verbandes der Landwirtschaftskammern und der Arbeitsgruppe "Landwirtschaftliches Versuchswesen" der Biometrischen Gesellschaft am 23. und 24. November 2011 in Kassel

Der Workshop hatte die Planung, Durchführung und Auswertung von OFE zum Inhalt. Ergänzt wurden diese Inhalte um Probleme der Ertragerfassung unter Produktionsbedingungen bei Druschfrüchten. Einen weiteren Schwerpunkt bildete die Besprechung von Anwendungsbeispielen. Nachfolgend sind die Arbeitsergebnisse des gemeinsamen Workshops zusammengestellt. Diese Ergebnisse stellen den **gemeinsamen Standpunkt der Workshop-Teilnehmer** dar und sind **Ausgangspunkt für künftige Beratungen zu OFE**.

Unter OFE wird ein Versuch verstanden, der mit dem Ziel der Beantwortung einer wissenschaftlichen Fragestellung durchgeführt wird und verallgemeinerungsfähige Aussagen liefern soll. Entsprechend muss es in einem OFE möglich sein, eine erwartungstreue Schätzung für die unbekannt Parameter (z. B. Erwartungswerte und Differenzen von Sorten, Düngungsstufen, Bodenbearbeitungsvarianten) und für ihre Präzision anzugeben. Davon abgeleitet ist die Einhaltung der allgemeinen Prinzipien zur Planung, Durchführung und Auswertung von landwirtschaftlichen Versuchen zwingend und unverzichtbar.

In Rahmen des Workshops wurde klar herausgearbeitet, dass Randomisation, also die zufällige Zuordnung der Prüfglieder zu den Randomisationseinheiten (Parzellen, Schläge, Orte), und die Existenz echter Wiederholungen der Prüfglieder auf der Ebene der Randomisationseinheiten stets eine valide Versuchsauswertung ermöglichen und den Lackmus-Test für die Auswertbarkeit eines OFE im Sinne eines biometrischen Versuchs darstellen. So sind Versuchsanlagen mit systematischer Abfolge der Prüfglieder abzulehnen. Beispielsweise stellt die alternierende Folge A-B-A-B-A-B keine valide Versuchsanlage dar. Die gewachsenen technischen Möglichkeiten einer georeferenzierten und sensorbasierten Datenerfassung ersetzen nicht die Notwendigkeit, echte Wiederholungen zu erzeugen. Auch ein großer Datenumfang sichert per se nicht die Einhaltung dieser Forderung. Daher muss streng gefordert werden, dass jedes OFE auch bei Nutzung der Mittelwerte auf Parzellenebene, also der Kombination Prüfglied*Randomisationseinheit, ausgewertet werden kann. Die Möglichkeit einer geostatistischen Auswertung ist als zusätzliches methodisches Instrumentarium zu sehen, dessen Anwendung zu einer Verbesserung der Treffgenauigkeit und Präzision führen kann. Keinesfalls jedoch kann eine geostatistische Auswertung fehlende Randomisation und fehlende Wiederholungen ersetzen.

Die Einhaltung der Prinzipien Randomisation und Wiederholung erfordert eine entsprechende Versuchsplanung, die auch über mehrere Schläge, Betriebe (Orte) und Jahre gehen kann, oftmals auch gehen muss. Es konnte herausgearbeitet werden, dass der Planungsansatz vollständiger und unvollständiger Blockanlagen eine hohe Flexibilität bietet, um die Spezifik der interessierenden Prüffaktoren (bspw. Bodenbearbeitung) mit den einzuhaltenden Prinzipien zu verknüpfen. Gleichzeitig sichert das Prinzip der Blockbildung die Möglichkeit einer Eliminierung systematischer umwelt- und managementbedingter Störgrößen.

So kann es möglich sein, dass auf einem Schlag beispielsweise aus technischen Gründen trotz aller Anstrengungen keine echten Wiederholungen möglich sind. Dann muss dieser Schlag als ein Block aufgefasst werden. Valide Aussagen je Jahr und Ort können gewonnen werden, wenn im selben Jahr und Ort die notwendigen Wiederholungen durch Einbeziehung weiterer Schläge erzeugt werden, wobei die Anordnung der Prüfglieder auf jedem Schlag neu zu randomisieren ist. Fehlende Wiederholungen und/oder Randomisation erlauben keine verallgemeinerungsfähigen Aussagen mit wissenschaftlicher Zielstellung und haben bestenfalls Hypothesen generierenden Charakter. Die Möglichkeit einer Fehlinterpretation und systematischen Verzerrung würde dann bewusst in Kauf zu nehmen sein, die eingesetzten finanziellen Mittel würden zu keinem verwertbaren Forschungsergebnis führen.

Der Übergang zur Nutzung von Einzeldaten bei der Auswertung erfordert eine sorgfältige Modellwahl, da die über die Prüf- und Designfaktoren hinaus zu berücksichtigenden festen und zufälligen Einflussgrößen spezifisch für den Einzelversuch festzulegen sind. Die Nutzung der Einzelwerte kann zusätzliche Informationen über den auszuwertenden Versuch liefern. Für die Modellwahl bieten analytische Kriterien (z. B. Akaike Informationskriterium AIC, korrigierte Akaike Informationskriterium AICC) eine gute Möglichkeit der Eingrenzung der Modellvarianten. Die Kontrolle durch eine Residuenanalyse und der Vergleich von beobachteter und geschätzter Varianz-Kovarianzfunktion sind jedoch unverzichtbar und stellen eine wirksame Rückversicherung gegen die Gefahr dar, beispielsweise Störgrößen zu übersehen. Die Modellwahl kann nicht durch eine automatisierte Softwarelösung ersetzt werden. Hier ist oft eine Kooperation von landwirtschaftlich-fachlicher und biometrischer Expertise erforderlich.

Insgesamt zeigten die dargestellten Anwendungsbeispiele die Vielfalt der anzuwendenden Auswertungsmodelle. Das führt zwingend zu einer engen Verbindung zwischen fachlicher und biometrischer Kompetenz. Lineare gemischte Modelle haben sich als ein Modellansatz erwiesen, mit dem die spezifischen Anforderungen von OFE gut abgebildet werden können.

Die Analyse der praktischen Anwendungen zeigte auch bedeutsame Störgrößen, die sich aus der angewendeten Messtechnik ergaben. Die Diskussion möglicher Einflussgrößen bei der Datenerfassung zeigte die Gefahren, die sich aus einer Nutzung unbereinigter Daten ergeben können. Daher ist die Aussagekraft von OFE auch daran gebunden wie es gelingt, die Messprinzipien und Einflussgrößen auf das Messergebnis zu verbessern. Eine permanente Justierung der Messeinrichtungen und Überprüfung der erfassten Einzelwerte ist bei dem derzeitigen Stand der technischen Entwicklung unverzichtbar. Das erfordert eine Kontrollwägung je Parzelle und eine darauf basierende Korrektur der Einzelwerte vor einer biometrischen Analyse.

Die Diskussion führte zur Notwendigkeit der Erarbeitung einer Schrittfolge bei der Planung, Durchführung und Auswertung von OFE. Nach dem derzeitigen Kenntnisstand bietet sich der folgende Ablauf an:

- Planung entsprechend der Planungsschritte der statistischen Versuchsplanung (Festlegung der Prüf- und Designfaktoren sowie möglicher Störfaktoren (bspw. Mähdrescher), des Versuchsplanes, der Wiederholungen);
- Versuchsdurchführung (bei Beachtung von Kontrollwägungen etc.);
- Auswertung:
 - Datenkontrolle auf Plausibilität, Fehlwerte, Fehlmessungen;

- Auswertung entsprechend des bei der Planung festgelegten vorläufigen Auswertungsmodells, ggf. mit Erweiterung zum Planungszeitpunkt unbekannter Störfaktoren;
 - Residuenkontrolle auf Verzerrungen und Verteilung;
 - bei Verfügbarkeit georeferenzierter Daten Modellverfeinerung durch Beachtung weiterer fester (bspw. räumlicher Trend) und zufälliger Störgrößen (bspw. räumliche Kovarianzen);
 - Festlegung des endgültigen Auswertungsmodells;
- Auswertung und Interpretation.

Es wurde deutlich, dass in einem koordinierten Versuchswesen nur bestimmte Fragestellungen mit Hilfe von OFE beantwortet werden können, Parzellenversuche in Versuchsstationen damit aber keinesfalls ersetzbar sind. Im Rahmen des Workshops wurde herausgearbeitet, dass OFE nur dann die mit ihnen verbundenen Vorteile ausschöpfen und den oft erheblichen Mitteleinsatz rechtfertigen, wenn es zu einer stetigen Zusammenarbeit zwischen dem praktischen Versuchswesen, der biometrischen Methodenentwicklung und der landtechnischen Entwicklung kommt. Die Planung, Durchführung und Auswertung von OFE erfordert eine hohe Fachkompetenz. Den akademischen Bildungseinrichtungen im Agrarbereich kommt bei der Vermittlung der verschiedenen Aspekte von OFE eine besondere Verantwortung zu, der derzeit noch nicht in vollem Umfang Rechnung getragen wird.

Verzeichnis der Autoren

- BORCHARDT, IMKE: Landwirtschaftskammer Schleswig-Holstein, Grüner Kamp 15-17, 24768 Rendsburg; Email: iborchardt@lksh.de
- BÜCHSE, ANDREAS: Bahnhofstraße 8, 67269 Grünstadt; Email: andreas.buechse@t-online.de
- DICKE, DOMINIK: Regierungspräsidium Gießen, Pflanzenschutzdienst Hessen, Schanzenfeldstraße 8, 35578 Wetzlar; Email: dominik.dicke@rpgi.hessen.de
- KOCK, CARSTEN: Gutsverwaltung Helmstorf, Magnus von Buchwaldt, 24321 Helmstorf, Email: helmstorf@t-online.de
- LUBKOWITZ, CRISTOPH: Landwirtschaftskammer Schleswig-Holstein, Grüner Kamp 15-17, 24768 Rendsburg; Email: clubkowitz@lksh.de
- MÖHRING, JENS: Fachgruppe Bioinformatik, Institut für Kulturpflanzenwissenschaften , Universität Hohenheim, Fruwirthstraße 23, 70599 Stuttgart
Email: moehring@uni-hohenheim.de
- OBENAUF, ULFRIED: Landwirtschaftskammer Schleswig-Holstein, Abteilung Pflanzenbau, Pflanzenschutz und Umwelt, Referat Marktfruchtbau/Getreide, Grüner Kamp 15-17, 24768 Rendsburg; Email: uobenauf@lksh.de
- PIEPHO, HANS-PETER: Fachgruppe Bioinformatik, Institut für Kulturpflanzenwissenschaften , Universität Hohenheim, Fruwirthstraße 23, 70599 Stuttgart
Email: piepho@uni-hohenheim.de
- RICHTER, CHRISTEL: Fachgebiet Biometrie und Versuchswesen, Department Nutzpflanzen- und Tierwissenschaften, Landwirtschaftlich-Gärtnerische Fakultät, Humboldt-Universität zu Berlin, Invalidenstr. 42, 10115 Berlin
Email: ch.richter@agrار.hu-berlin.de
- SCHÖN, HANS-GEORG: IT-Center, Fakultät Agrarwissenschaften und Landschaftsarchitektur, Hochschule Osnabrück, Am Krümpel 31, 49090 Osnabrück
Email: h.schoen@hs-osnabrueck.de
- SPIPKE, JOACHIM: Arbeitsgruppe Biometrie und Agrarinformatik, Institut für Agrar- und Ernährungswissenschaften, Martin-Luther-Universität Halle-Wittenberg, Karl-Freiherr-von-Fritsch-Straße 4, 06120 Halle (Saale)
Email: joachim.spilke@landw.uni-halle.de
- THÖLE, HEINRICH: Stechlinweg 6, 14469 Potsdam, Email: heinrich.thoele@web.de
- WILLIAMS, EMLYN R.: Statistical Consulting Unit, The Australian National University, Canberra ACT 0200, Australia; Email: emlyn.williams@anu.edu.au