



Quelle: Ursula Siemes, 2010

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

Inhaltsverzeichnis

1	Einleitung und Problemstellung	2
2	Allgemeine Grundsätze für OFE	4
2.1	Ökologischer Aussagebereich von OFE	4
2.2	Prinzipien der Versuchsplanung und ihre Umsetzung	6
2.3	Versuchsanlagen für OFE	8
2.4	Modellbildung für OFE	10
2.5	Auswertung von OFE ohne Georeferenzierung	11
2.5.1	Auswertung von Einzelversuchen ohne Georeferenzierung	11
2.5.2	Auswertung von Versuchsserien	12
3	Besonderheiten von OFE mit georeferenzierten Daten	13
3.1	Versuchsfrage und daraus abzuleitende Anforderung an die Versuchsfläche	13
3.2	Grundsätze der Versuchsplanung und einige Konsequenzen für die Auswertung bei georeferenzierten Daten	13
3.3	Datenaufbereitung	15
3.4	Aggregation und Zusammenführung von Merkmalsdaten	15
3.5	Weitere Spezifika der Auswertung	16
3.6	Hinweise zur Interpretation	18
4	Literatur (zitierte Literatur und Auswahl methodisch orientierter Bücher)	18

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 Ertrags erfassung 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.

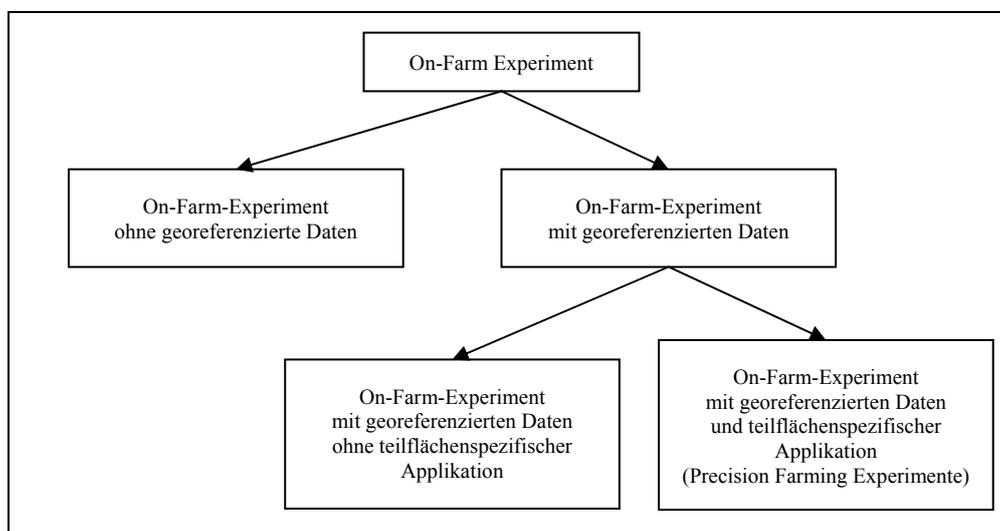


Abbildung 1: Übersicht zur Begriffsbildung und Klassifikation von On-Farm-Experimenten

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 Versuchstationen 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 Ver- suchs	ökologischer Aussagebereich	
Einzel- versuch	ein Ort ¹⁾	ein Jahr	Ackerfläche(n) des Schläges (der Schläge) und die Witterung des Versuchsjahres an dem Ort	
	ein Ort ^{1) 2)}	mehrere Jahre ³⁾	Ackerflächen des Versuchsortes (Schläges) ²⁾ und seine klimatischen Bedingungen ³⁾	
Versuchs- serie	meh- rere Orte	fest aus- gewählt	ein Jahr	Ackerflächen der Versuchsorte (Schläge) und die Witterung des Versuchsjahres an diesen Orten
		zufällig ausge- wählt		Anbaugebiet und Witterung des Versuchsjahres
	fest aus- gewählt zufällig ausge- wählt	mehrere Jahre ³⁾		Ackerflächen der Versuchsorte und ihr Klima ³⁾ Anbaugebiet 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 einheitlichen 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 versuchsmethodische 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 Anlage-

schema 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 Anbaugesbietes ausgewählt (Anbaugesbiet/Ort) und je Ort liegt ein Schlag mit einem vollständigen Block vor (Tabelle 2 Situation a erweitert auf mehrere Anbaugesbietes), so ist bei der Serienauswertung die Hauptwirkung der Anbaugesbietes und die Wechselwirkung Prüfglied x Anbaugesbiet 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 Anbaugbiet/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 Prüfglied-differenzen 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 Gesichtpunktes 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 Randomisationsplanes 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 restrinierten 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.

- 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 unterschied-

liche 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 teilflächenspezifischer 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 geeignet 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 (‚Basismerkmal‘) mit seinen Positionen ausgewählt (in der Regel das Merkmal mit der niedrigsten Erfassungsintensität). Für alle anderen Merkmale werden deren Werte auf den durch das Basismerkmal 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 Ran-

domisation 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 weitergehend 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 Randomisationsein-

heiten (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 Modellselektion (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 merkm=variante block kv1 kv2
Ansatz 1: REPEATED/TYPE=sp(modell) (x y) SUBJECT=parzelle
Ansatz 2: REPEATED/TYPE=sp(modell) (x y) SUBJECT=block
Ansatz 3: REPEATED/TYPE=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 Ver-

suchsplanung 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 Stichprobenumfangsplanung 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).

4 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): 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.
- SPIPKE, J.; RICHTER, C.; PIEPHO, H. P. (2010): Model selection and its consequences for different split-plot designs with spatial covariance and trend. Plant Breeding (in press).
- 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.