Analysen zur Anwendung der ‚Hohenheim - Gülzower - Serienauswertung' im regionalisierten Sortenversuchswesen in Mecklenburg - Vorpommern

D i s s e r t a t i o n

zur Erlangung des akademischen Grades
Doctor rerum agriculturarum
(Dr. rer. agr.)

eingereicht an der
Lebenswissenschaftlichen Fakultät
der Humboldt-Universität zu Berlin

von
Dipl. Ing. agr. Volker Michel
geboren am 13. April 1963 in Rostock

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Lebenswissenschaftlichen Fakultät:
Prof. Dr. Richard Lucius

Gutachterin/ Gutachter:

1. **Prof. Dr. Christel Richter**
2. **Prof. Dr. Hans-Peter Piepho**
3. **Dr. Friedrich Laidig**

Tag der mündlichen Prüfung: 30. September 2015

Inhaltsverzeichnis

1 Einführung 5

2 Zielstellung 9

3 Versuchsserien - Theorie, Literatur 10

4 Material 24

4.1 Pflanzenarten 24

4.2 Merkmale und deren deskriptive Beschreibung 25

4.3 Faktoren und Faktorstufen 34

4.3.1 Faktoren bzw. Klassifizierungsgrößen in Einzelversuchen 36

4.3.2 Zusätzliche Faktoren bzw. Klassifizierungsgrößen der Serien / Umwelten 45

4.4 Dem Sortenprüfsystem immanente systematische Unbalanciertheit 49

4.5 Differenzierte Präzision der Sortenmittelwerte aus Einzelversuchen 54

5 Methoden 56

5.1 Regionalisierung 56

5.1.1 Boden-Klima-Räume für alle Pflanzenarten als Bausteine verschiedener Regionalisierungsziele 56

5.1.2 Pflanzenartspezifische Anbaugebiete für die Sortenempfehlung 59

5.1.3 Konzept von Zielgebiet und Einbeziehung von Nachbargebieten 61

5.2 Transformation 63

5.2.1 Box-Cox-Transformation 64

5.2.2 Relativzahlen und die ‚folded exponential transformation' 66

5.3 Das Auswertungsmodell 69

5.3.1 Das allgemeine gemischte Modell und die verwendete Syntax für die Formulierung der Modellgleichungen 69

5.3.2 Herleitung des Auswertungsmodells für die Varianzkomponentenschätzung 70

5.3.3 Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen ... 74
5.3.4 Modellanpassungen zur Berücksichtigung der genetischen Korreliertheit der Sorteneffekte zwischen Anbaugebieten ... 77
5.3.5 Gewichtung von Zielgebiet und Nachbargebieten bei der Bestimmung der Sortenmittelwerte und ihrer Genauigkeit - Einführung ... 80
5.3.6 Vorhersage gewichteter Sortenmittelwerte bei Betrachtung des Faktors Sorte als zufällig - in einem Modellschritt ... 82
5.3.7 Vorhersage gewichteter Sortenmittelwerte bei Betrachtung des Faktors Sorte als zufällig - in zwei Modellschritten ... 83
5.3.8 Gewichtete Sortenmittelwerte ohne Schrumpfungsschätzung ... 86
5.3.9 Schätzung spezifischer Anbaugebietsgewichte und Präzisionsmaße für jede Sorte ... 88
5.4 Integration der Modelle in die SAS®-basierte Anwendersoftware PIAFStat ... 90
5.5 Definition von Szenarien ... 96

6 Ergebnisse und Diskussion ... 99
6.1 Modellvoraussetzungen, Datenverhalten und Transformation ... 99
6.1.1 Typische Muster beim Datenverhalten ... 99
6.1.2 Effekte der Transformation auf die Mittelwerte in der Serienauswertung ... 114
6.1.3 Diskussion ... 116
6.2 Varianzkomponenten- und Mittelwertschätzung ... 124
6.2.1 Typische Muster in der Struktur der Varianzkomponenten ... 124
6.2.2 Effekte der langjährigen Auswertung und der Einbeziehung vorgelagerter Versuche auf Mittelwerte und ihre Genauigkeit ... 131
6.2.3 Effekte einer Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen auf Serien-Mittelwerte und ihre Genauigkeit ... 138
6.2.4 Diskussion ... 140

6.3 Gewichtung von Zielgebiet und Nachbargebieten 146
6.3.1 Typische Muster in der Struktur der genetischen Korrelationen 146
6.3.2 Typische Muster in der Gewichtung von Nachbargebieten 148
6.3.3 Effekte der Einbeziehung von Nachbargebieten auf Mittelwerte im Zielgebiet und ihre Genauigkeit 151
6.3.4 Diskussion 156
6.4 Diskussion der Problematik: Faktor ‚Sorte' als fix oder als zufällig definieren? 161
6.4.1 BLUE- und BLUP- Ansatz für Sorteneffekte 161
6.4.2 Differenzierte Anforderungen in den Phasen der züchterischen Selektion und der nachfolgenden amtlichen Sortenprüfung 162
6.4.3 Verzerrungspotenzial durch Schrumpfungsschätzung bei unterlassener Berücksichtigung von mit dem Zielmerkmal korrelierten wertbestimmenden Merkmalen 165
6.5 Validierungsansätze 169
6.5.1 Stabilität der Varianzkomponentenschätzung 169
6.5.2 Auswirkung variierender Varianzkomponentenschätzungen auf die Schätzung der Sorteneffekte 170
6.5.3 Güte der Schätzung von Sorteneffekten nach einer sechsjährigen Auswertung als Vorhersage für das Folgejahr 172
6.5.4 Diskussion 175
7 Schlussfolgerungen 177
Zusammenfassung 184
Summary 186
Abkürzungsverzeichnis 188
Literaturverzeichnis 191
Tabellenverzeichnis 196
Abbildungsverzeichnis 198
Anlagenverzeichnis 200

Anlagen 201
Erklärung 213
Danksagung 214

1 Einführung

In den letzten drei Jahrzehnten war bei den meisten landwirtschaftlichen Kulturpflanzenarten ein zwar von Ertragsschwankungen begleiteter, aber doch anhaltender Ertragsfortschritt zu verzeichnen, der überwiegend in einem bislang nicht abgeschwächten Züchtungsfortschritt begründet ist (Laidig et al., 2014). Nach der züchterischen Selektionsphase und dem Sortenzulassungsverfahren beim Bundessortenamt schließt sich eine Phase der Sortenprüfung mit dem Ziel der regionalen Einschätzung von zugelassenen, vertriebsfähigen Sorten für die landwirtschaftliche Praxis an. Diese Regionalprüfung erfolgt in Deutschland durch die Länderdienststellen für das Sortenversuchswesen (LDS) im Wesentlichen auf der Basis von Landessortenversuchen sowie vorgelagerter Wertprüfungen und EU-Sortenversuche. Basis der Bewertung ist die Gesamtheit der wertbestimmenden Eigenschaften der Sorten unter regionalem Aspekt. Große Bedeutung kommt dabei der frühzeitigen, möglichst schätzgenauen, gut reproduzierbaren Auswertung jeder einzelnen wertbestimmenden Eigenschaft zu. Während in der Züchtungsphase Selektion und damit aus biometrischer Sicht die Bestimmung von Rangfolgen im Vordergrund steht, kommt es in der Regionalprüfung von Sorten, die sich mit der Zulassung bereits als Teil der leistungsfähigsten Spitze erwiesen haben, weniger auf Rangfolgen als auf die präzise Schätzung der Differenzen zwischen Sorten an. Unter Einbeziehung aller meist univariat ausgewerteten Merkmale können auf dieser Basis nicht nur Gesamtbewertungen erarbeitet werden, sondern darüber hinaus auch sortenspezifische Hinweise zur Einordnung einer Sorte in konkrete Anbausituationen (Fruchtfolge, Bodenbearbeitung, Saatzeit, Standortbesonderheiten u.v.m.).

Aufgrund der begrenzten, tendenziell gesunkenen Kapazitäten auf regionaler Ebene und der gleichzeitigen Forderung nach schneller Überleitung des Züchtungsfortschrittes in die Praxis gilt es dabei, die potenzielle Datenbasis voll auszuschöpfen. Probleme bereitet hierbei die dem integrierten Sortenprüfsystem immanente stark ausgeprägte Unbalanciertheit / Lückigkeit der Daten. Um Sortenunterschiede trotzdem mit möglichst geringem Fehler zu schätzen, sind der Komplexität der Datenstruktur und den Sorte $\times$ Umwelt - Interaktionen angemessene Auswertungsmodelle erforderlich. Hier haben sich gegenüber einem Stand der Versuchsserienauswertung, wie ihn z.B. Bätz (1984) beschreibt, erhebliche methodische Innovationen aufgetan, die der Problematik entgegenkommen.

Allerdings haben diese Neuerungen bislang kaum Einzug in die Praxis des angewandten Versuchswesens vieler Länderdienststellen gehalten. Die hergebrachte regionale Ertragsauswertung bezog i.d.R. nur Versuchsergebnisse der Landessortenversuche (LSV) der jeweiligen Länderdienststelle der letzten drei Jahre ein. Diese bislang übliche Beschränkung auf LSV, auf Sortenvergleiche innerhalb immer nur genau eines LSV-Jahrganges (also innerhalb vollständig orthogonal besetzter Datenstrukturen) sowie auf eine Regelprüfzeit von drei Jahren ermöglicht zwar eine traditionelle Versuchsserienauswertung mit orthogonaler Datenstruktur, wie sie bereits Mudra (1952) beschreibt, lässt aber einen erheblichen Umfang an vorliegenden Daten ungenutzt und ist damit im Sinne der o.g. Zielstellung nicht prinzipiell begründbar (Piepho und Michel, 2001). Im Interesse der Frühzeitigkeit und Genauigkeit ist es sinnvoll, die Datenbasis weiter zu fassen und voll auszuschöpfen. Die Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern hat dazu die Auswertung in folgender Weise verändert:

Seit 1998:

In der Region angelegte, den Landessortenversuchen zeitlich vorgelagerte amtliche Sortenprüfungen (Wertprüfungen, EU-Sortenversuche, Orientierungsversuche u.a.) wurden einbezogen. Dazu wurde auf ein Auswertungsmodell umgestellt, welches die Verrechnung unbalancierter Daten ermöglicht. Methodische Grundlage war zunächst die von Yates (1933) und Patterson (1997) vorgeschlagene sogenannte FITCON-Methode (method of fitting constants; s. Abschn. 3).

Seit 2000:

Das Auswertungsmodell wurde von FITCON auf gemischte Modelle umgestellt und der bis dahin verwendete Faktor *Umwelt* (dessen Stufen Einzelversuche kodieren) wurde untersetzt in die Faktoren *Jahr*, *Ort* und *Versuchstyp* an Ort × Jahr - Kombinationen. Einerseits war dies die Voraussetzung für die Ermittlung geeigneterer Maßzahlen für die Schätz- bzw. Vorhersagegenauigkeit (für künftigen Anbau in neuem Jahr an neuem Ort). Insbesondere konnten dadurch aber auch Verzerrungen in der Mittelwertschätzung abgebaut werden, wenn Sorten in Jahren, Orten oder Versuchstypen erheblich ungleichmäßig besetzt waren. Zusätzlich wurde die Option der Gewichtung von Mittelwerten aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen eingeführt.

Seit 2004:

Die regionalbezogene Sortenempfehlung erfolgt für standortkundlich definierte pflanzenartspezifische Anbaugebiete (nicht mehr für politisch definierte Gebiete wie Bundesland o.ä.). Die dafür erforderliche regionalisierte Versuchsauswertung basiert auf dem Konzept von Zielgebiet und gewichteter Einbeziehung von ähnlichen Nachbargebieten, wobei sich die Räume um unterschiedliche Zielgebiete überlappen können.

Diese Veränderungen mündeten in der *Hohenheim-Gülzower-Serienauswertung* (Michel et al., 2007a), die in der Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern (LFA) seit 2004 eingesetzt wird. Die Agrarministerkonferenz (2004) des Bundes und der Länder hat in ihrem *Beschluss zur Neuordnung des Sortenversuchswesens* die bundesweite Einführung der hier sogenannten *Hohenheimer Methodik* eingefordert. Mit dieser Methode können unter anderem folgende Besonderheiten des Sortenwesens berücksichtigt und adäquat verwertet werden:

Unbalanciertheit:

a) Ergebnisse der den LSV vorgelagerten Versuche (insbesondere Wertprüfungen des Bundessortenamtes (WP) und EU-Sortenversuche (EUSV)) können in die Versuchsserienauswertung einbezogen werden, obgleich ihre Sortimente (gemeinsam in einem Versuch geprüfte Sorten) sehr stark von denen der Landessortenversuche abweichen.
b) Trotz der bei langjährigen, aus einem Selektionsprozess gewonnenen Daten bestehenden großen Unbalanciertheit können viele (alle fachlich noch relevanten) Versuchsjahre einbezogen werden.
c) Bei differenzierter Datenbesetzung erfolgt eine Gewichtung der unterschiedlich besetzten Faktorstufen. Die Gewichtungsabstufung erfolgt durch das verwen-

dete gemischte Modell, sie basiert auf der Relation der Varianzkomponenten zueinander. Dies ist die Voraussetzung dafür, dass z.B. bezüglich einer Sorte schwach besetzte Jahre (i.d.R. Wertprüfungsjahre) in die Gesamtauswertung ohne das Risiko einer Über- oder Untergewichtung einfließen können. Dieses Risiko einer Fehlgewichtung besteht in der verbreiteten Herangehensweise von Versuchsanstellern auch in den Länderdienststellen (LDS), welche die Mittelwertberechnung über eine Versuchsserie traditionell mit einfachen Methoden der Tabellenkalkulation ausführen. So erfolgt in diesen LDS häufig in einem ersten Schritt der Zusammenfassung von Versuchen die Bildung arithmetischer Sortenmittelwerte über die Versuche in einem Jahr zu einem ‚Jahresmittelwert' und im zweiten Schritt die Bildung arithmetischer Mittelwerte von Jahresmittelwerten zu einem zwei- oder dreijährigen Mittelwert. Dieser Algorithmus bedingt eine nur suboptimale Gewichtung von Versuchsergebnissen, wenn zwischen den Jahren die Anzahl einbezogener Versuche unterschiedlich ist.

d) Die Gesamtauswertung erfordert keine absolut durchgängigen Verrechnungssorten oder orthogonalen Kerne.

Regionalität:

Ergebnisse aus benachbarten Regionen im standortkundlichen Überlappungsbereich können mit objektiviert abgestuftem Gewicht in die Auswertungen für ein Zielgebiet einfließen. Wesentliche Einflussgrößen für die Gewichtung sind die genetische Korrelation zwischen den Anbaugebieten (basierend auf den Varianz-Kovarianz-Strukturen) sowie auch die differenzierte Datenbesetzung der Anbaugebiete.

Differenzierte Versuchspräzision:

Unterschiedliche Fehlervarianzen der Mittelwerte in Einzelversuchen werden entsprechend den Vorschlägen von Möhring und Piepho (2009) bei der Gewichtung in der Versuchsserienauswertung berücksichtigt. Auch diese Gewichtsabstufung erfolgt in Abhängigkeit von der Relation der Varianzkomponenten durch das verwendete gemischte Modell. Relevante Ursachen für unterschiedliche Standardfehler sind:

- unterschiedliche Restvarianz der Einzelversuche,
- unterschiedliche Wiederholungsanzahl in den Einzelversuchen,
- unterschiedliche Wiederholungsanzahl von Prüfgliedern innerhalb von Einzelversuchen,
- Prüfglieder mit unterschiedlicher Fehlervarianz im Einzelversuch aufgrund unvollständiger Blocks oder räumlicher Modelle.

Verletzung der Modellvoraussetzungen / Nichtadditives Datenverhalten:

kann durch eine entsprechende Transformation der Daten berücksichtigt werden.

Die *Hohenheim-Gülzower-Serienauswertung* wurde in der Phase der Modellierung theoretisch hergeleitet und nur an wenigen Beispieldatensätzen getestet. Im Zuge der Einführung der hier vorgestellten Methoden wurden seitdem in der LFA für den *Großraum* diluvialer Standorte in Nordostdeutschland ausgesprochen große, langjährige Datenbestände für alle bearbeiteten landwirtschaftlichen Pflanzenarten und deren varianzanalytisch auswertbaren Merkmale zusammengeführt. Nach nunmehr mehrjähriger Einführung der *Hohenheim-Gülzower-Serienauswertung* und

Umsetzung für das o.g. Regionalisierungskonzept sollen mit dieser Arbeit verallgemeinerungswürdige Erfahrungen und Ergebnisse abgeleitet werden. Die sekundäre Auswertung der Datensätze von insgesamt 34 Variablen lässt insbesondere Aussagen hinsichtlich der Bedeutung unterschiedlicher Variationsursachen, des Datenverhaltens (Transformationserfordernis) und der bestmöglichen Nutzung von Daten aus Nachbargebieten erwarten. Vergleichbar große Datenbestände für derartige Analysen liegen außerhalb des Sortenwesens (aus üblichen geplanten Versuchsserien im pflanzenbaulichen Versuchswesen) nicht vor.

Die Datenstrukturen sind kompliziert und müssen in enger Wechselbeziehung zu den Sorte $\times$ Umwelt - Wechselwirkungen gesehen werden. Diese in integrierten Sortenprüfsystemen immanente Problematik hat die Entwicklung der *Hohenheim-Gülzower-Serienauswertung* motiviert und zunehmend erforderlich gemacht. Daher fällt die Beschreibung der Datenstrukturen und der systematischen Verflechtung zwischen den Klassifikationsfaktoren einerseits und der Lückigkeit anderseits in der vorliegenden Arbeit relativ umfangreich aus.

Unter den methodischen Grundlagen wurde das Auswertungsverfahren der *Hohenheim-Gülzower-Serienauswertung* und das Regionalisierungskonzept (Roßberg et al., 2007) in unmittelbarem Bezug zu dieser Arbeit unter Mitwirkung des Autors dieser Arbeit entwickelt. Daher werden auch diese Aspekte unter *Methoden* umfassend beschrieben.

2 Zielstellung

Die in der Einleitung beschriebenen eingeführten Innovationen haben die beratungsrelevante Datenbasis für die Auswertung von Sortenversuchen in Mecklenburg-Vorpommern erheblich erweitert. Dies geht aber einher mit einer z.T. extremen Zunahme der Unbalanciertheit der auszuwertenden Daten und mit einer großen Anzahl zu berücksichtigender Effekte in dem der Versuchsserienauswertung zugrunde liegenden mathematischen Modell.

Vor diesem Hintergrund dient diese Arbeit folgenden Aufgaben und Zielen:

- Schätzung (nicht Testung!) von Sortenunterschieden in einem Anbaugebiet unter Minimierung des Vorhersagefehlers von Sortendifferenzen als Grundlage für Sortenempfehlungen;
- die Schätzung des Vorhersagefehlers von Sortendifferenzen und -mittelwerten soll folgende Anforderungen beinhalten:
  - das Optimierungskriterium ist die Minimierung der mittleren Fehlervarianz der Vorhersage einer Differenz von Sortenmittelwerten (MSEP);
  - in diesen MSEP soll die Verzerrung einfließen, die sich aus der Einbeziehung von Nachbargebieten in die Schätzungen für ein Zielgebiet ergibt; die Minimierung des MSEP hat also Vorrang vor Erwartungstreue der geschätzten Mittelwertdifferenzen;
  - der je Sorte auszuweisende Vorhersagefehler des *Mittelwertes* soll sich - vor dem Hintergrund der Unbalanciertheit - aus dem Vorhersagefehler der Differenz zu einer Sorte mit identischer Varianz-Kovarianz-Matrix berechnen; der Vorhersagefehler des Mittelwertes soll unabhängig vom Wertelevel konkreter Umwelten sein, es sollen also keine reinen, nicht mit dem Faktor Sorte gekreuzten Umwelteffekte einfließen;

  → dieses Herangehen entspricht dem Anspruch der regionalen Sortenprüfung, den Landwirten Sorten für den *künftigen* Anbau auf *neuen* Schlägen im Anbaugebiet zu empfehlen;
- Erarbeitung verallgemeinerungsfähiger Aussagen für unterschiedliche landwirtschaftliche Pflanzenarten sowie unterschiedliche Merkmalstypen zu:
  - Einhaltung von Modellvoraussetzungen, Datenverhalten und Transformationserfordernis;
  - Varianzkomponenten (insbesondere der Sorte × Umwelt - Interaktionen);
  - genetischen Korrelationen zwischen benachbarten Anbaugebieten auf diluvialen Standorten im Nordosten Deutschlands;
- Einschätzung der Wirksamkeit der eingeleiteten Veränderungen hinsichtlich der Vorhersagegenauigkeit und Relevanz für die Sortenbewertung;
- Empfehlungen für den praktischen Einsatz der Auswertungsmethode unter Nutzung der Anwendersoftware PIAFStat und für die Regionalisierungsstrategien bei der Auswertung.

3 Versuchsserien - Theorie, Literatur

Landwirtschaftliche Feldversuche werden als Versuchstypus eingeordnet, der für die Prüfung unter praxisnahen Bedingungen steht. Damit ist der geplante Aussagebereich a priori sehr viel weiter gefasst, als es bei Modellversuchen, Gefäßversuchen, in Klimakammern u.ä. der Fall ist. Gegenüber letzteren nimmt im Feldversuch die Anzahl steuerbarer Konstantfaktoren ab und die Repräsentativität für die Praxis nimmt zu (Rasch, 1987).

Hierbei steht insbesondere die Unvorhersagbarkeit der Witterung oder unter vielem anderen z.B. auch des Schaderregerdrucks während des Anbaus eines Feldversuches zu Buche. Insofern impliziert die Anforderung an Repräsentativität für die Praxis geradezu die Notwendigkeit der Anlage von Feldversuchen in mehreren Jahren. Bereits Kuckuck und Mudra (1950) gehen für die Pflanzenzüchtung davon aus, dass Neuzüchtungen nur auf Grund mehrjähriger Versuchsergebnisse zu beurteilen sind.

Auch die Repräsentativität einer einortigen Versuchsanstellung wird häufig den Erfordernissen der Praxis nicht gerecht. In aller Regel sollen abgeleitete Aussagen nicht nur für genau diesen Schlag gelten, sondern für ein mehr oder weniger breit definiertes Anbaugebiet. Selbst in relativ kleinräumig definierten Regionen sollte man nicht von Homogenität des Einflussfaktors *Standort* ausgehen. Insofern erfordert auch die Frage, ob die Behandlungen / die Prüfglieder in einer Region einheitlich reagieren bzw. wie bedeutsam Interaktionen sind, häufig nicht nur eine *mehrjährige*, sondern gleichzeitig auch *mehrortige* Versuchsanstellung.

Auch Cochran und Cox (1957) unterstellten, dass viele landwirtschaftliche Versuchsanstellungen in der Hoffnung durchgeführt werden, dass die Ergebnisse nicht nur für die Umweltbedingungen in den jeweiligen Einzelversuchen Gültigkeit haben, sondern auf die praktische Landwirtschaft übertragbar sind. Die Gesamtfragestellung erfordert somit i.d.R. eine Versuchsserie, in der der einjährige Feldversuch an einem Standort, hier im Weiteren als Einzelversuch bezeichnet, *nur* ein Element ist. Nach Mudra (1952) sollte es die Regel sein, Versuche über allgemein interessierende Fragen nicht als Einzelversuche, sondern als Versuchsserie zu planen und auszuwerten. Der induktive Schluss auf die Grundgesamtheit *Anbaugebiet und dessen Klima* setzt eine ausreichend repräsentative Stichprobe von Einzelversuchen voraus (Bätz, 1984). Ebenso argumentiert Rasch (1987), dass bei biologischen Problemen aufgrund der großen Anzahl nicht kontrollierbarer Umweltbedingungen die Antwort auf eine Versuchsfrage i.d.R. nur durch eine Versuchsserie möglich ist. Er definiert eine Versuchsserie als eine Serie von Einzelversuchen, die mit der gleichen Fragestellung, mit den gleichen Prüfgliedern räumlich oder / und zeitlich getrennt durchgeführt wird.

Methodische Grundlagen für die Auswertung von Versuchsserien wurden bereits von Yates und Cochran (1938) veröffentlicht. Mudra (1949) stellt die ‚Differenzmethode' in Einzelversuchen mit Randomisationseinschränkungen wie auch in so genannten Streuversuchen (viele Orte ohne Wiederholungen) einer varianzanalytischen Auswertung gegenüber. Ziel der *Differenzmethode* war die Ausschaltung der von Mudra beobachteten häufigen Korreliertheit von Prüfgliedeffekten in Blöcken oder an Orten mit einem damals erforderlichen einfachen mathematischen Ansatz. Dieser Ansatz lag in der Bestimmung der Fehlervarianz einer Differenz aus den Einzeldifferenzen von Prüfgliedpaaren anstelle der Bestimmung aus der Fehlervarianz eines Mittelwertes. Einen solchen Ansatz beschrieb bereits Student (1923). Patterson

(1997) beschreibt dies als ‚Methode der direkten Differenzen'. Auch Forkman (2013) greift wieder auf diese Ansätze zurück. Er verfolgt damit vor allem das Ziel, in unbalancierten Datensätzen ausschließlich direkt (in jeweils einem Versuch) gemessene Differenzen zu einer oder zu mehreren durchgängig geprüften Referenzsorten in der Auswertung zu verwenden.

Umfassende Zusammenfassungen zum internationalen Stand der Versuchsserienauswertung erfolgten u.a. durch Cochran und Cox (1957), Bätz (1984) und Patterson (1997). Auf die besondere Bedeutung der Entwicklung des REML-Algorithmus (restricted bzw. residual maximum likelihood) durch Patterson und Thompson (1971) wird im Zusammenhang mit Unbalanciertheit in Versuchsserien eingegangen (s.u.).

Cochran und Cox (1957) demonstrieren an Beispieldaten Probleme und Auswertungsansätze für Versuchsserien u.a. unter Berücksichtigung möglicher heterogener Prüffaktor × Umwelt - Interaktionen, heterogener Fehlervarianzen in Einzelversuchen oder ungleicher Strukturen in den Einzelversuchen.

Bätz (1984) formuliert folgende Schwerpunkte für die Auswertung einer Versuchsserie:

1. Analyse der Ursachen der Prüfglied/Umwelt-Wechselwirkungen,
2. Beurteilung der Verwendbarkeit und des Informationswertes von Versuchen, Versuchsorten und Versuchsjahren,
3. Beurteilung der Ökostabilität (Ertragssicherheit) von Prüfgliedern,
4. Qualifizierung der Entscheidung für einzelne Prüfglieder,
5. Berücksichtigung der Wechselwirkung Prüfglieder/Umwelt bei der Ableitung von Anbauempfehlungen.

In den 80'er Jahren erfolgte die Versuchsserienauswertung vorrangig durch Varianzanalyse mit anschließendem Mittelwertvergleich. Die Beurteilung der Prüfglieder und deren Wechselwirkung wurden als fachlicher Schwerpunkt betrachtet, während die Beurteilung der Orts- oder Jahresunterschiede eine untergeordnete Rolle spielte (Autorenkollektiv, 1987). Unabhängig von der Hypothesenwahl ‚*fix*' oder ‚*zufällig*' für einzelne Faktoren wurde die Frage der Varianzkomponentenschätzung gegenüber Tests zurückgestellt. Umfassend wurde die Berechnung der MQ-Werte (mittlere Abweichungsquadrate) in Zusammenhang mit der Wahl des ‚richtigen F-Tests' auch für zufällige Effekte im Modell diskutiert.

Mudra (1952) beschreibt die Versuchsserien-Auswertung „mehrjähriger Versuche von einem Ort" sowie methodisch analog die von „einjährigen Versuchen aus verschiedenen Orten". Die zusammenfassende Auswertung von mehrjährigen und gleichzeitig mehrortigen Versuchen bezeichnet Mudra noch als äußerst kompliziert. Auch Bätz (1984) beschreibt, dass zu diesem Zeitpunkt mit den verfügbaren Auswertungsverfahren nur 2-fach klassifizierte Auswertungen erfolgen würden (z.B. Versuche und Prüfglieder). Die Auswertungen erfolgten überwiegend auf Basis der Prüfgliedmittelwerte aus den Einzelversuchen unter Berücksichtigung der zu ‚poolenden' Fehlervarianzen der Einzelversuche. Die Auswertung von Versuchsserien für einen Prüffaktor wird durch ein Autorenkollektiv (1987) für verschiedene Modellansätze beschrieben. Richter et al. (1999) beschreiben die Auswertung von Versuchsserien zweifaktorieller Einzelversuche auf der Basis der Einzelwerte.

Vor dem Hintergrund eingeschränkter Informationstechnologie war es insbesondere im Falle unbalancierter Daten über einen großen Zeitraum ein sehr erfolgreicher pragmatischer Ansatz, auch Einzelversuche einer mehrjährigen mehrortigen Serie als statistisch unabhängig zu betrachten. Dabei werden die Variablen *Ort* und *Jahr* in einer einzigen Variablen *Versuch* bzw. *Umwelt* zusammengefasst. Die Auswertung

erfolgte dann häufig nach der von Yates (1933) vorgeschlagenen und von Patterson (1978) vertieften FITCON-Methode (method of fitting constants). Patterson (1997) zählt dazu im weiteren Sinne alle für diesen Zweck angepassten linearen Modelle mit einem einzelnen Fehlerterm. Im Sortenwesen erfolgt dies i.d.R. durch eine Mittelwert basierte Serienauswertung über Umwelten (Versuche), wobei die Haupteffekte *Prüfglied* und formal auch *Versuch* fix gesetzt werden und nur deren Wechselwirkung als Fehlerterm zufällig ist. Bereits 1978 beschreibt Silvey, dass diese Methode in der offiziellen Sortenprüfung Großbritanniens routinemäßig eingesetzt wird. Auch vom Bundessortenamt wird dieses Verfahren bereits langjährig für die Erstellung der deutschen Beschreibenden Sortenlisten verwendet (Laidig, 2013). Mit diesem vereinfachten Modellierungs-Ansatz wurden hinsichtlich der begrenzenden Rechnerleistungen Auswertungen komplex strukturierter großer unbalancierter Datensätze rein technisch ermöglicht, die ohne diese Vereinfachungen oft nicht erfolgreich durchlaufen würden. Patterson (1997) beschreibt erheblich kürzere Computer-Rechenzeiten bei FITCON gegenüber REML. Allerdings stellten Piepho und Michel (2001) an einem langjährigen Datensatz mit Rapssorten auch fest, dass die Annahme der Unabhängigkeit von Versuchen nicht tatsächlich erfüllt war, sondern dass Versuche im gleichen Jahr bzw. am gleichen Ort korrelierte Effekte aufweisen. Sie diskutierten - mit Verbesserung der Rechnerleistungen nun mögliche - Verbesserungen in der der Sachlage adäquaten Modellbildung, wobei auch hier noch von relativ einfachen Varianz-Kovarianz-Strukturen, insbesondere von homogenen Varianzen sämtlicher Zufallseffekte ausgegangen wurde. Auswertungsansätze auf Basis faktoranalytischer Varianz-Kovarianz-Strukturen diskutieren Denis et al. (1997), Gilmour et al. (1998), Piepho und van Eeuwijk (1999) und Piepho (1999). Kelly et al. (2007) beschreiben faktoranalytische Modelle als einfache Form zur Approximation der komplett unstrukturierten Form der genetischen Varianz-Kovarianz-Matrix, die zur besten Modellanpassung führen und in Züchtungsprogrammen auch das Ziel ‚Selektion der besten Genotypen' am besten erreichen.

Die Auswertung von Versuchsserien in der Pflanzenzüchtung diskutiert van Eeuwijk (2007) für Datenstrukturen von Genotyp $\times$ Umwelt - Mittelwerten. Dabei werden insbesondere folgende Problemkreise angerissen: Definition von Effekten im Modell als fix oder zufällig, Plausibilität additiver Modelle, Reaktionsnormen der Genotyp $\times$ Umwelt - Wechselwirkungen, multiplikative Modelle für die Wechselwirkung, Testung, Modelle zur Berücksichtigung von Varianzheterogenität.

Aufgrund der Bedeutung von Versuchsserien, in denen Einzelversuche *nur* Element der Gesamtauswertung sind, bemängeln bereits Nelder (1986) wie auch Richter et al. (1999), dass Literatur und Software-Lösungen überwiegend für Einzelversuche bereit stehen, für Versuchsserien aber nur bruchstückhaft vorhanden seien. Richter et al. (1999) leiten Auswertungs-Algorithmen für mehrortige und/oder mehrjährige Versuchsserien mit zwei Prüffaktoren (im Einzelversuch) detailliert her, wobei bezüglich der Faktoren Orte und Jahre alle Konstellationen für die Definition *fix* versus *zufällig* Berücksichtigung finden. Über alle Einzelversuche wird hierbei eine einheitliche Anlage in vollständigen Blocks, Varianzhomogenität sowie insgesamt vollständige Balanciertheit der Daten angenommen. Insofern greifen diese Methoden eher im Bereich der Forschung und anbautechnischer, befristeter Versuchsprojekte. In Sortenprüfsystemen können diese Voraussetzungen i.d.R. als nicht erfüllt gelten.

Einschritt- oder Zweischrittanalyse

Grundsätzlich können Versuchsserien in Form einer Einschritt- oder einer Zwei- bzw. Mehr-Schritt-Analyse ausgewertet werden (Piepho et al., 2012). Die Einschrittanalyse

setzt auf den Daten der Parzellen in den Einzelversuchen auf. Aus theoretischer Sicht liefert sie unter Annahme des Einschritt-Analysemodells exakte Schätzwerte für alle fixen Effekte (BLUE) und für alle zufälligen Effekte (BLUP). Smith et al. (2001a) bezeichnen die Einschrittanalyse als den ‚*goldenen Standard*'. Dies wird durch Simulationsergebnisse von Welham et al. (2010) bestätigt.

Bei der Zweischrittanalyse werden zunächst die Einzelversuche entsprechend ihren jeweiligen Besonderheiten ausgewertet und Sortenmittelwerte mit ihren Standardfehlern berechnet und diese dann im zweiten Schritt über die gesamte Versuchsserie verrechnet. Die Einzelversuche können sich bereits planungsseitig, z.B. in der Wahl der Versuchsanlage oder sogar der Anzahl Prüffaktoren, unterscheiden. Hinzu kommen u.U. weitere Unterschiede im Zuge der Auswertung der Einzelversuche, z.B. in Folge einer versuchsspezifischen Modellselektion oder der Nutzung einer räumlichen Modellierung des Fehlers.

Die Güte einer Zweischrittanalyse muss sich an der Einschrittanalyse bemessen, sofern die Einschrittanalyse alle Besonderheiten der Einzelversuche (z.B. die Unterschiedlichkeit von Fehler- und Blockvarianzen zwischen den Versuchen) in der Modellbildung vollständig berücksichtigt. Letzteres erreicht aber bei großen unbalancierten Datensätzen und komplexer Modellstruktur oft seine Grenzen. Sehr lange Rechenzeiten oder auch Abbruch wegen zu geringer Speicherkapazität treten nicht selten auf, wenn die MIXED Prozedur von SAS® genutzt wird. Dann ist häufig die Nutzung einer Zweischrittanalyse eine bessere Alternative als ein reduziertes Einschritt-Modell, welches relevante Gegebenheiten unberücksichtigt lässt. Die Zweischrittanalyse kann also erforderlich werden, wenn die Versuchsanlagen und Auswertungsmodelle zwischen den Versuchen erheblich variieren, allerdings zu Lasten der Effizienz im Vergleich zu einer adäquaten Einschrittanalyse.

Frensham et al. (1997) untersuchen die Frage der Genotyp × Umwelt - Varianzheterogenität in einer Zweischrittanalyse. In den meisten Fällen sei es nötig, eine Versuchsniveau abhängige Genotyp × Umwelt - Interaktion zu berücksichtigen.

Es ist also wichtig, für die Zweischrittanalyse eine Methode zu finden, die die Einschrittanalyse möglichst gut reproduziert. Eine zentrale Rolle kommt hierbei der Methode zu, mit der die Mittelwerte aus dem ersten Schritt im zweiten Schritt gewichtet werden. Möglich ist hierbei als einfachste Variante die Arbeit ohne Gewichtung, was einer Gleichgewichtung aller Einzelversuchs-Sortenmittelwerte unter Annahme von Varianzhomogenität gleichkommt. Methodische Ansätze der differenzierten Gewichtung für unterschiedliche Situationen diskutieren u.a. Smith et al. (2001b), Möhring und Piepho (2009) und Welham et al. (2010). Piepho et al. (2012) stellen eine Methode der Mehr-Schritt-Analyse vor, in der die vollständigen Varianz-Kovarianz-Matrizen der adjustierten Mittelwerte der Einzel-Umwelten in der Serienauswertung einbezogen werden. Diese Methode kann die Ergebnisse der Einschrittanalyse vollständig reproduzieren, wenn man bekannte Varianzkomponenten annimmt.

Der Aspekt differenzierter Gewichtung der Sortenmittelwerte aus Einzelversuchen wird in dieser Arbeit vertieft und in seiner Wirkung auf die Serienmittelwerte an realen Daten untersucht.

Hypothese ‚fix' oder ‚zufällig' für Versuche bzw. Umwelten

Die Planung und Auswertung von Versuchsserien steht maßgeblich im Zusammenhang mit der Hypothese ‚*fix*' oder ‚*zufällig*' für die Versuche bzw. für Orte oder / und für Jahre (Bätz, 1984). Die Fragen der Repräsentanz, des Aussagebereiches bzw.

der Verallgemeinerungsfähigkeit hängen damit zusammen. Häufig, insbesondere bei Versuchsanstellungen mit Beratungsauftrag, deckt sich die von Bätz (1984) formulierte Definition des Aussagebereiches ‚*Anbaugebiet und dessen Klima*' mit dem Ziel der Versuchsanstellung. Auch Richter et al. (1999) gehen davon aus, dass sich im Zuge der Serienauswertung im Allgemeinen eine zusammenfassende Beurteilung über alle Umwelten anschließen soll. Dies impliziert dann bereits eine Zielstellung der Versuchsanstellung, für die i.d.R. die Hypothese ‚*zufällig*' für die Umwelten angestrebt werden sollte. Nach Buhtz und Bätz (1984) kann die Hypothese ‚*zufällig*' angenommen werden, wenn die Orte eine zufällige und repräsentative Stichprobe aus einem Anbaugebiet darstellen und die Jahre hinsichtlich ihres Witterungsverlaufes ebenfalls eine zufällige und repräsentative Stichprobe ergeben. Unterschiedliche Konstellationen werden von den Autoren am Beispiel der Pflanzenzüchtung und Sortenprüfung dargestellt und hinsichtlich des sich ergebenden Aussagebereiches systematisiert (Tab. 1). Auch vom Autor dieser Arbeit werden hier im Weiteren überwiegend Beispiele aus dem Sortenversuchswesen aufgegriffen, wobei aber synonym für den Faktor *Sorte* auch andere Faktoren stehen könnten.

Tab. 1: Aussagebereich von Feldversuchen nach Buhtz und Bätz (1984)

Modell Nr.	**Stichprobe (Einzelversuch)**	**Hypothese für Orte**	**Hypothese für Jahre**	**Aussagebereich (Grundgesamtheit)**
1.	Orte in einem Jahr	zufällig	(fix)	Anbaugebiet mit Witterungsbedingungen, die denen des Versuchsjahres entsprechen
2.	Orte in einem Jahr	fix	(fix)	Mittlere Anbaubedingungen der Orte bei Witterungsbedingungen, die denen des Versuchsjahres entsprechen
3.	Ort in mehreren Jahren	(fix)	zufällig	Ort und dessen Klima
4.	Ort in mehreren Jahren	(fix)	fix	Ort mit Witterungsverhältnissen, die dem „Mittel" der Versuchsjahre entsprechen
5.	Orte und Jahre	zufällig	zufällig	Anbaugebiet und dessen Klima
6.	Orte und Jahre	fix	fix	Mittlere Anbaubedingungen der Orte mit Witterungsverhältnissen, die dem „Mittel" der Versuchsjahre entsprechen
7.	Orte und Jahre	zufällig	fix	Anbaugebiet mit Witterungsverhältnissen, die dem „Mittel" der Versuchsjahre entsprechen
8.	Orte und Jahre	fix	zufällig	Mittlere Anbaubedingungen der Orte und deren Klima

( ) ‚Jahr' bzw. ‚Ort' in nur einer Stufe, also in der Modellgleichung vernachlässigbar

Die von Bätz (1984) formulierte Zielstellung ‚*Schluss auf die Grundgesamtheit Anbaugebiet und dessen Klima'* (Zeilennummer ‚5' in Tab. 1) trifft uneingeschränkt auf das in dieser Arbeit diskutierte Sortenprüfsystem mit dem Ziel der Sortenberatung zu. Die anderen in Tab. 1 aufgeführten Ansätze passen in einigen Fällen für geplante Versuchsserien z.B. zu anbautechnischen Fragen, bei denen im Gegensatz zum Sortenversuchswesen die Fragestellung a priori zeitlich und räumlich eng gefasst wird oder bei denen die Versuchsorte bewusst so distinkt gewählt werden, dass die Reaktionen auf jeden Standort a priori im Zentrum der Fragestellung stehen und eine Mittelwertbildung über Orte selbst bei nicht signifikanter Interaktion eher nur eine Zusatzinformation darstellt.

Auch wenn häufig die Hypothese ‚*zufällig*' für Jahre und Orte anzustreben ist, so ist aber der Idealfall einer tatsächlich zufälligen Wahl der Orte im absoluten Sinne nicht realisierbar. Bestenfalls kann ein Standortnetz etabliert werden, das fachlich-subjektiv als *hinreichend* repräsentativ sowohl für das Mittel als auch für die Variation der Standortbedingungen eingeschätzt wird. Es wird bei der Hypothesenwahl also immer ein gewisser Widerspruch zwischen Anforderung aus der Zielstellung und tatsächlicher ‚*Natur*' eines Faktors hinsichtlich Auswahl / Randomisation der Umwelten bestehen, sobald ein Umwelt-Faktor, speziell aber der Faktor *Ort,* als zufällig eingestuft wird. Insofern verwendet bereits Bätz (1984) die relativierende, pragmatische Formulierung „…*ausreichend* repräsentative Stichprobe…". Es gibt keine rein statistischen Kriterien für diese Entscheidung, der Aussagebereich muss vorrangig durch sachlogische Erwägungen bestimmt werden (Autorenkollektiv, 1987).

In frühen Literaturquellen neigte man im Zweifelsfall eher zu einer formalen Anwendung der Hypothese ‚*fix*'. So argumentiert z.B. Bätz (1984), dass bei geringer Anzahl von Jahren (oder Orten) eher der Ansatz ‚*fix*' verwendet werden sollte. Für zufällige Effekte mit weniger als fünf bis zehn Stufen, welche einen zu testenden fixen Effekt nicht enthalten, kann es nach Piepho et al. (2003) vorteilhaft sein, diesen als fix zu definieren. Ursache ist die bei wenigen Stufen u.U. sehr geringe Genauigkeit der Varianzschätzungen, wie es in Analogie hierzu auch bei der Nutzung der Inter-Block-Information in Versuchsanlagen mit unvollständigen Blocks diskutiert wird. Bei geringer Anzahl von Stufen eines seiner Natur nach zufälligen Faktors oder Effektes im Modell wird also u.U. eine formale Umdefinition zu ‚fix' pragmatisch sinnvoll sein. Unabhängig von einer derartigen Umdefinition ist es für die Versuchsplanung ebenso wie für die Interpretation und Bewertung der Ergebnisse wichtig, dass sich der Versuchsansteller in allen Phasen darüber klar ist, ob die Zuordnung zu ‚fix' oder ‚zufällig' der ‚*Natur*' des Faktors entspricht. Andernfalls kann z.B. bei Definition von ‚Jahr' als fix die Genauigkeit der Mittelwertschätzung von Sorten nach nur zwei- oder dreijähriger Auswertung u.U. als hoch eingeschätzt werden, obgleich die (dann unbekannte) Vorhersagegenauigkeit für neue (zufällige) Jahre gering ist.

Um ein kleines Anbaugebiet zu repräsentieren, reichen häufig vier bis sechs Orte aus, die Anzahl Jahre sollte aber lt. Autorenkollektiv (1987) nicht vorab festgelegt, sondern durch ein sequentielles Prinzip bestimmt werden. Hierbei soll aufgrund der rückwirkenden Beschreibung der Umweltbedingungen beurteilt werden, ob die vorliegenden Versuche die Grundgesamtheit ausreichend repräsentieren oder ob weitere Versuche in zusätzlichen Jahren erforderlich sind. Allerdings wird dieses Prinzip als subjektiv-pragmatisch eingeschränkt, da es praktisch schwierig ist, eine Klassifikation der Witterungsbedingungen vorzunehmen. Nach Richter et al. (1999) kann die Repräsentativität von Jahren erst nach Ablauf der Versuchsdauer beantwortet werden.

Den Ansatz, Aussagen eher für vordefinierte (also fixe) Umwelten abzuleiten, zeigt ein durch Hamblin et al. (1980) beschriebenes Vorgehen, wonach aufgrund von Vorauswertungen solche Standorte auszuwählen seien, die für den *Durchschnitt* des Anbaugebiets repräsentative Ergebnisse liefern. Kienzl (1974) schlägt solche Vorauswahlen bei bayerischen Landessortenversuchen nicht nur für die Versuchsorte vor, sondern schließt auch ‚*untypische*' Jahre von der Auswertung aus. Nach Ansicht des Autor vorliegender Arbeit ist Repräsentativität von Umwelten aber letztlich nicht loszulösen von der Hypothese ‚*zufällig*', die Ausgrenzung z.B. von Jahren sollte nur in extremen Einzelfällen erfolgen. Die einbezogenen Umwelten sollten nicht nur den (vermeintlichen) Durchschnitt, sondern auch die Variabilität im Anbaugebiet und insbesondere die Reaktion der Prüfglieder auf diese uneingeschränkt repräsentieren. Auch Richter et al. (1999) stellen die Reaktion der Prüfglieder auf die durch die Standorte verursachte Variabilität in den Mittelpunkt des Interesses. Speziell auch das Verhalten z.B. von Sorten in besonderen Situationen, die u.U. vermeintlich ‚untypisch', aber doch Teil der Umweltvariabilität sind und die sich also in ähnlicher Weise wiederholen können, sind für die Nutzung in der Beratung besonders wichtig. Auch ein Landwirt will nicht nur in einem *typischen* Jahr die richtige Sortenwahl treffen, sondern im langjährigen Mittel, das auch besondere Jahre einschließt.

In diesem Sinne repräsentieren Versuchsjahre die klimatische Bandbreite und sollten m.E. als zufällig angesehen werden, solange nicht willkürlich Jahre ausgewählt oder ausgeschlossen werden, z.B. weil sie in der Rückschau als *typisch* oder *untypisch* klassifiziert werden. Da das Jahr mit seiner Witterung von Natur aus zufälligen Charakter trägt, kommt es vorrangig auf eine entsprechend große Anzahl Jahre an, um die Witterung repräsentativ zu erfassen (Autorenkollektiv, 1987). Dass wenige Versuchsjahre die klimatische Bandbreite nur sehr unvollkommen bzw. vage repräsentieren können, schlägt sich in den statistischen Maßzahlen dann angemessen nieder, wenn das Jahr als zufällig definiert ist. Der Witterungsverlauf ist in seinem Einfluss auf Pflanzenbestände derart mannigfaltig, dass sich - außer in Extremsituationen - nicht an einigen ausgewählten, oft über lange Zeiträume kumulierten Witterungskennzahlen festmachen lässt, ob die einbezogenen Jahre in ihrer Gesamtheit als repräsentativ oder Einzeljahre als ‚typisch' gelten können. Insofern können die Formulierungen „Witterungsbedingungen, die denen des Versuchsjahres entsprechen" oder „Witterungsverhältnissen, die dem Mittel der Versuchsjahre entsprechen" (Tab. 1) m.E. nur als sehr unscharf bzw. subjektiv angesehen werden.

In jüngerer Literatur gibt eher die Anforderung aus der Zielstellung das Primat für die Hypothesenwahl, sofern die Umweltfaktoren *hinreichend* als repräsentativ angesehen werden können. Smith et al. (2005) plädieren dafür, in der Pflanzenzüchtung und Sortenprüfung Genotypen als zufällig zu definieren, da sich so der ‚Selektionsfehler' minimieren lässt. Auch Van Eeuwijk (2007) unterstützt eine sehr weitgehende Wahl der Hypothese ‚*zufällig*' für den Faktor ‚*Genotyp*', sofern das Interesse nicht vorrangig auf die Ausprägung jedes einzelnen Genotyps (als Individuum), sondern auf das Gesamtsortiment gerichtet ist. Inzwischen ist auch in der Pflanzenzüchtung das Interesse an BLUP für Genotypen gestiegen, da sich die Präzision im Vergleich zu BLUE-Schätzungen erhöht (Piepho et al., 2008). Eher als in der Phase der Züchtung treten bei der amtlichen Sortenprüfung damit allerdings Zielkonflikte auf: aus Sicht der zu beratenden Landwirte kommt es vorrangig darauf an, aus der Gesamtheit der Sorten eine herausragende Leistungspitze zu finden, selbst wenn im Sinne dieses Gesamtzieles Individual-Sorten, die nicht als Teil einer ‚Population' von Sorten angesehen werden sollten, dabei im zufälligen Ansatz durch Schrumpfungsschätzung (siehe Abschnitt 6.4.3) ‚gestört' geschätzt werden könnten. Die den

Züchtungsfortschritt hervorbringenden Züchter erwarten dagegen von der offiziellen Sortenprüfung nachvollziehbarer Weise eine Schätzmethodik, die ihre jeweilige Sorte als Individuum, den Faktor Sorte also als fix, ansieht. Eine Schrumpfungsschätzung zum Sortenmittel findet bei ihnen kaum Akzeptanz. Richter et al. (1999) gehen davon aus, dass man üblicherweise die ‚*eigentlichen'* Prüffaktoren als fixe Faktoren behandelt. In diesem Zusammenhang soll festgestellt werden, dass eine Vielzahl der methodischen Arbeiten zur Versuchsserienauswertung aus dem Bereich der Züchtung- und Sortenprüfung stammt, dass hierbei wiederum oft die Arbeit des selektierenden Züchters in den Vordergrund gestellt wird - so auch bei van Eeuwijk (2007). Eine vertiefte Diskussion für den Bereich der amtlichen Sortenprüfung erfolgt im Abschnitt 6.4.

Während im Frühstadium eines züchterischen Selektionsprozesses häufig ein einzelnes Zielmerkmal im Fokus steht, erfolgt die Bewertung von ‚fertigen' Sorten im Sortenprüfsystem grundsätzlich auf Basis einer simultanen Gesamtsicht auf alle wertbestimmenden Eigenschaften (ggf. mittels einer Indexberechnung). Beim Winterweizen wird jede Sorte in der *Beschreibenden Sortenliste* (Bundessortenamt, 2014) durch 18 im Feld erfasste Merkmale und 12 Qualitätsmerkmale beschrieben. Für jedes dieser 30 Merkmale erfolgte dafür eine univariate Auswertung. Bei der Analyse jedes *Einzelmerkmals* kommt es auf eine möglichst präzise Schätzung / Vorhersage der Sortenunterschiede auf einer stetigen Skala an, damit in der Synthese die *Gesamtbewertung über alle Merkmale* bestmöglich fundiert ist. Sortenrangfolgen werden also erst in der multivariaten Sicht bedeutsam, sind je Einzelmerkmal aber i.d.R. wenig relevant. Ein statistischer Signifikanztest von Sorten gegeneinander ist auf der univariaten Ebene ebenso kaum hilfreich. Des Weiteren sind u.a. auch aus genannten Gründen für die univariaten Auswertungen parametrische Verfahren, welche Informationen auf einer stetigen Skala nutzen und Effekte auf einer stetigen Skala schätzen, prinzipiell zielführender als nichtparametrische Verfahren, welche auf Ranginformationen basieren - solange die Voraussetzungen für parametrische Verfahren hinreichend gegeben sind. In diesem Zusammenhang soll eine treffende Aussage von Stroup (2014) zitiert werden: „… they are focused only on testing, not on estimation. In most plant and soil science research, the question is not, 'Is there a treatment difference?'. Instead, it is, 'We know there is a difference. How big is it?' ".

Balanciertheit

Mudra (1952) geht davon aus, dass Versuche aus beliebig vielen Orten und Jahren zusammengefasst werden können, wenn die Zahl der Versuchsglieder und die Art der Anlage einheitlich sind. Damit deutet er bereits ein sehr häufiges Problem an, dass diese im besten Fall gegebene *Balanciertheit* der Daten nicht zwangsläufig gegeben sein muss - sei es durch Ausfälle von Prüfgliedern oder durch bewusst geplante Lücken z.B. im Zuge der Streichung bzw. Neuaufnahme von Prüfgliedern über die Jahre. 1949 formuliert Mudra es so, dass in der Serienauswertung nur Prüfglieder einbezogen werden können, die in allen Jahren und an allen Orten geprüft wurden, wobei es gleichgültig sei, ob in einzelnen Versuchen weitere Prüfglieder geprüft wurden. Die Abschlussformulierung beinhaltet das noch heute häufig verwendete *Schneiden orthogonaler Kerne* - es werden dabei Versuche oder Prüfglieder in der Weise von der Auswertung ausgeschlossen, dass die verbliebene Datenstruktur orthogonal / balanciert ist. Bei der Versuchsserienauswertung mit dem Softwareprodukt DAVEP in der DDR war das Bilden orthogonaler Kernstrukturen zwingende Voraussetzung für die Auswertung (Franko et al., 1982).

Möhring et al. (2004a) und Piepho und Möhring (2006) raten inzwischen bei Verwendung gemischter Modelle im Sortenversuchswesen vom willkürlichen Ausschluss von Prüfgliedern ab. Hintergrund sind ihre Analysen und Simulationen zur Fragestellung, inwieweit die durch Selektion hervorgerufene Unbalanciertheit zu Verzerrungen bei der Schätzung von Varianzkomponenten und Sorteneffekten führt. Da im Verlauf der Sortenprüfung in jedem Jahr auch Selektion stattfindet, kann von vollständig zufälligen Fehlstellen im mehrjährigen Datensatz nicht ausgegangen werden. Die Selektion findet zwar nicht nur auf Basis des einen gerade in der univariaten Auswertung betrachteten Merkmals statt, sondern auf Basis der Gesamtheit aller wertbestimmenden oder für das Sortenregister relevanten morphologischen u.a. Eigenschaften. Andererseits ist die Selektion häufig auch nicht völlig unabhängig von diesem Einzelmerkmal. Die ‚missing completely at random'-Annahme (MCAR) in der Definition nach Little und Rubin (2002) ist also nicht erfüllt. Allerdings kann die schwächere ‚missing at random'-Annahme (MAR) als erfüllt gelten, wenn zum einen ein separierbares Modell für den Fehlwertmechanismus sowie für die Merkmalsdaten angenommen werden kann und zum anderen die im Selektionsprozess involvierten Daten vollständig in die Auswertung einfließen. Piepho und Möhring (2006) kommen im Speziellen zu folgenden Aussagen: (1) Ein verzerrender Einfluss der Selektion kann hinreichend vernachlässigt werden, sofern alle im Gesamtzeitraum angefallenen Daten (auch die der herausgenommenen Sorten!) in die Auswertung einfließen. (2) Bei Unbalanciertheit im Zusammenhang mit Selektion sei der REML-Algorithmus dem ML-Algorithmus vorzuziehen und der BLUP-Ansatz (best linear unbiased prediction) sei dem BLUE-Ansatz (best linear unbiased estimation) vorzuziehen. Probleme, die in der Bestimmung von Sorteneffekten durch BLUP - also mit dem zufälligen Ansatz für Sorteneffekte - allerdings liegen können, werden in dieser Arbeit aufgezeigt und diskutiert (s. 6.4).

Die Problematik des Anstrebens absoluter Balanciertheit durch ‚Schneiden' orthogonaler Kerne soll an folgendem Szenarium verdeutlicht werden: wenn in einer langjährigen vielortigen Serie eine Sorte in einem einzelnen Versuch fehlt, so wäre dieser Versuch (oder je nach Modell sogar der Versuchsort über alle Jahre oder das Versuchsjahr über alle Orte) komplett auszuschließen oder alternativ müsste diese Sorte durchgehend aus dem Datensatz entfernt werden. In jedem Fall entstünde ein erheblicher Informationsverlust, welcher auch aus Gründen der wirtschaftlichen Effizienz kaum zu vertreten ist, wenn die Grundgesamtheit das Anbaugebiet und dessen Klima ist.

Insofern waren Auswertungsalgorithmen zur Abschwächung der von Mudra (1952) noch gestellten Balanciertheitsforderung gesucht. Patterson (1978) beschreibt Methoden, die bei unbalancierten Datenstrukturen ‚least squares'-Schätzungen (LS Means) für Sorten gestatten, wobei neben ‚direkten' auch ‚indirekte' Sortenvergleiche ermöglicht werden. Unter direkten Sortenvergleichen werden hier Vergleiche zwischen zwei Sorten verstanden, die in allen einbezogenen Versuchen gemeinsam geprüft wurden. Bei indirekten Vergleichen standen diese Sorten dagegen nicht in allen Versuchen und insbesondere wurden sie nicht in allen Versuchen *gemeinsam* geprüft. Indirekte Vergleiche setzten ‚Drittsorten' voraus, die quasi 'Brücken' zwischen Versuchen bilden, in denen das Sortenpaar nicht gemeinsam stand. Ein Mindestmaß an solchen Brückensorten ist in Sortenprüfsystemen i.d.R. automatisch vorhanden. Da allerdings der Umfang an solchen Brücken die Präzision von indirekten Sortenvergleichen maßgeblich mitbestimmt, sind für Deutschland die bundesweit abgestimmten Verrechnungs- und Vergleichssorten (s.a. 4.3) eine entscheidende Basis für indirekte Sortenvergleiche, wie sie z.B. die Einstufungen in der Beschreibenden Sortenliste (Bundessortenamt, 2014) liefern.

Da häufig mehr als ein zufälliger Effekt in einem linearen (additiven) Modell zu berücksichtigen ist, kommen zunehmend gemischte Modelle (mixed models) zum Einsatz. Die Schätzmethode für Varianzkomponenten nach dem ANOVA-Ansatz (analysis of variance, eingeführt durch Fisher (in Fisher und Mackenzie, 1923)) ist für Serien von Sortenversuchen aufgrund der erheblichen Unbalanciertheit nicht adäquat (Kempton, 1984). Im Fall unbalancierter Daten beeinflussen die Relationen zwischen den Varianzkomponenten der zufälligen Effekte der Modellgleichung z.T. erheblich die Mittelwertschätzung. Nach Patterson und Thompson (1971) ist unter Nutzung gemischter Modelle der Restricted Maximum Likelihood Algorithmus (REML) für die Auswertung unbalancierter Daten geeignet. Die Varianzkomponenten werden, Normalverteilung für zufällige Effekte zugrunde legend (in generalisierten linearen gemischten Modell (GLMM) auch andere Verteilungsfunktionen), mit dem REML - Algorithmus geschätzt (bei mehr als einer Varianzkomponente iterativ). Die REML - Schätzungen für Varianzen und Kovarianzen werden in die Mixed-Model-Gleichungen eingesetzt, wonach nach den unbekannten fixen Parametern aufgelöst und empirische gewichtete LS-Mittelwerte (eWLS-Schätzer; LS-Mittelwerte (LS Means) = kleinste Quadrat-Mittelwerte) geschätzt werden können. Im Zusammenhang mit modernen Softwarelösungen und leistungsfähiger Rechentechnik können diese Algorithmen nun zunehmend auch sehr große Datenumfänge mit unbalancierten Strukturen und mehreren fixen und zufälligen Effekten in der Modellgleichung verarbeiten. Bei der traditionellen Varianzanalyse unter Nutzung der Algorithmen der ANOVA (*analysis of variance)* erfüllt die Varianzschätzung zwar die Anforderung der Erwartungstreue (REML dagegen nicht absolut), die Mittelwertschätzung ist bei unbalancierten Daten aber fraglich. Robinson (1987) fasst die Literaturdiskussion zum Vergleich des ML-Ansatzes (maximum likelihood) mit REML zusammen. Als Nachteil von ML gegenüber REML stellt sie besonders heraus, dass im balancierten Fall die Varianzen von der ANOVA-Schätzung abweichen.

Van Eeuwijk (2007) gibt auf dem Internationalen Symposium ‚Agricultural Field Trials - Today and Tomorrow' in Stuttgart-Hohenheim einen Überblick zur Umsetzung der REML-Methodologie im Sortenwesen und verweist auf gängige Softwarelösungen für gemischte Modelle (z.B. SAS®, ASREML®, GenStat®).

Silvey (1978) stellt fest, dass die offizielle Sortenprüfung in Großbritannien fast ausnahmslos lückige Tabellen hervorbringt. Sie zeigt, dass der Sortenwechsel sich derart beschleunigte, dass eine ausnahmslos in allen Versuchen mitgeprüfte Kontrollsorte in langjährigen Datensätzen nicht mehr verfügbar ist. Daher wurde in Großbritannien zu dieser Zeit die Mittelwertbildung über Relativzahlen zu einer Kontrollvariante durch die Kleinst-Quadrat-Schätzung mittels der ‚fitting constants' - Methode (s.o.) abgelöst. Diese Problematik der systematischen Lückigkeit ist zum Beispiel bei der Planung der in dieser Arbeit diskutierten Landessortenversuche ein nicht zu umgehendes, dem Prüfsystem immanentes Charakteristikum: Landessortenversuche stellen keine zeitlich begrenzte Versuchsserie dar, sondern werden kontinuierlich über die Jahre fortgeführt. Dabei verlassen naturgemäß Sorten das System, während Neuzulassungen aufgenommen werden. Auch die Verweildauer der Sorten in der Serie ist sehr variabel. Bei der mehrjährigen Auswertung regionaler Sortenversuche müssten nach Analysen des Autors zur Erzwingung balancierter Reststrukturen häufig über 50% der vorhandenen Daten beratungsrelevanter Sorten ungenutzt bleiben. Die Problematik der Unbalanciertheit wird in dieser Arbeit anhand von Datenstrukturen im Sortenprüfsystem im Abschnitt 4.4 vertieft.

Forkman (2013) beschreibt Probleme in der Akzeptanz von verallgemeinerten Kleinst-Quadrat-Schätzungen, wenn der Datensatz unbalanciert ist. Es könne z.B.

zur Schätzung einer mittleren Differenz zwischen einer eingeschränkt geprüften Testsorte zu einer durchgängig geprüften Referenzsorte kommen, die außerhalb der in direkten Vergleichen gemessenen Differenzen liege (also nur in Versuchen, die beide Sorten enthalten). Eine derartige Beschreibung einer Sorte sei Akteuren mit besonderem Interesse an dieser konkreten Sorte kaum vermittelbar. Er stellt die von ihm so bezeichnete ‚reference treatment method' vor, die je Testsorte nur die direkten gemessenen Differenzen zu genau einer für alle Vergleiche definierten Referenzsorte auswertet und weitere Ergebnisse unberücksichtigt lässt. Alle Vergleiche von Testsorten untereinander erfolgen indirekt über die Einzelvergleiche zur Referenzsorte. Forkman schränkt ein, dass selbst für die Vergleiche zur Referenzsorte die Testeffizienz leidet, zugunsten der fachlichen Nachvollziehbarkeit für Nichtstatistiker. Insbesondere sinkt aber die Präzision der Vergleiche von Testsorten untereinander. Nach Einschätzung des Autors dieser Arbeit ist die Fokussierung auf Vergleiche zu einer Referenzsorte bzw. auf indirekte Vergleiche von Testsorten über eine Referenzsorte grundsätzlich suboptimal. Oft stellen über lange Zeiträume vorgesehene Referenzsorten im mehrjährigen Kontext nicht mehr den fachlichen Maßstab für neue Sorten dar, sondern sind methodisch motiviert (Brücke zwischen Versuchen, insbesondere Versuchen in verschiedenen Jahren).

Analyse der Prüfglied × Umwelt - Interaktionen

Nach Kuckuck und Mudra (1950) interessieren in der Züchtung häufig nicht nur Mittelwerte über die Orte bzw. Jahre sondern auch ihr Verhalten an den einzelnen Standorten bzw. in den Jahren.

Die erweiterte Auswertung von Versuchsserien insbesondere hinsichtlich der Prüfglied × Umwelt-Wechselwirkungen im Sortenwesen der DDR wurde maßgeblich durch Bätz (1984) geprägt. Bei Mittelwertvergleichen wurden Sorte × Umwelt-Wechselwirkungen bei der Durchführung der Tests im Kontext der Frage ‚*Wogegen ist zu testen?*' berücksichtigt. Bei Annahme ‚zufällig' für einen Effekt erfolgte ein Signifikanztest für diese Varianzkomponente, die Varianzkomponentenschätzung selbst war eher nachrangig. Die Möglichkeiten der Datenverdichtung, der Mittelwertbildung über Orte und/oder Jahre wurden von diesen Testergebnissen abhängig gemacht, d.h. bei Signifikanz jeweiliger Wechselwirkungen als ‚unzulässig' eingestuft. Obgleich dieser Entscheidungsprozess von Bätz (1984) so dargelegt ist, räumt er doch auch ein, dass auch bei signifikanter Wechselwirkung zusammenfassende Aussagen für die Grundgesamtheit von primärem Interesse sein können.

Dagegen werden in jüngeren Auswertungen mit gemischten Modellen alle im Modell als ‚*zufällig* ' definierten Effekte konsequent als solche behandelt. D.h. insbesondere, dass i.d.R. nicht vorrangig eine Testung dieser Effekte erfolgt, sondern eine Schätzung der Varianzkomponenten. Die Bewertung der Größenordnung der Varianzen und insbesondere ihrer Relationen zueinander nimmt z.B. bei Laidig et al. (2008) wie auch in der vorgelegten Arbeit umfassenden Raum ein. Eine hohe Sorte × Umwelt - Varianz schränkt die Vorhersagegenauigkeit ein bzw. erhöht den Standardfehler von Sortenmittelwerten bzw. Mittelwertdifferenzen. Eine Mittelwertbildung über die Umwelten erfolgt dessen ungeachtet. Allerdings sollte speziell bei hoher Sorte × Ort - Interaktion innerhalb eines Anbaugebietes über die Abgrenzung des Anbaugebietes, seine ausreichende innere Homogenität, sowie über die Repräsentativität der Versuchsorte für dieses Anbaugebiet nachgedacht werden.

Großes Interesse findet die Ausweisung von statistischen Kennziffern der Umweltstabilität / Ökostabilität von Prüfgliedern. Die Verwendung der Ökovalenz wird u.a.

durch Wricke (1967) und die der Ökoregression u.a. durch Utz (1972) diskutiert. Eine umfassende methodische Zusammenfassung erfolgte durch Bätz (1984). Er stellt am Beispiel des Merkmals Ertrag im Sortenwesen die Ertragsstabilität als ein besonderes Ziel der Sortenwahl heraus. Piepho (2005) diskutiert diese Stabilitätsmaße als Parameter eines gemischten Modells. Damit können mit Hilfe von Software für gemischte Modelle Stabilitätsanalysen auch bei unbalancierten Daten auf einfachem Wege durchgeführt werden. Mühleisen et al. (2014a) vergleichen unter Nutzung gemischter Modelle die Ertragsstabilität von Hybrid- und Liniensorten bei selbstbefruchtenden Getreidearten. Mühleisen et al. (2014b) fanden für Wertprüfungen mit Wintergerste, dass eine präzise Bewertung der Ertragsstabilität einzelner Sorten eine Prüfung in mindestens 40 Test-Umwelten erfordert.

Die Ökovalenz charakterisiert die Umweltstabilität, z.B. die Ertragssicherheit einer Sorte. In der vergleichenden Sortenbewertung gilt eine Sorte dann als stabil, wenn sich ihre Leistung proportional zum Ertragspotenzial der Umwelt, z.B. ausgedrückt im (ggf. adjustierten) Versuchsmittel, verhält. Dahinter steht das agronomische bzw. dynamische Konzept der Ökostabilität (Thomas, 2006). Dies kommt in vergleichsweise stabilen Sorte $\times$ Umwelt - Wechselwirkungseffekten bzw. in stabilen Rängen (bzw. bei Darstellung von Relativerträgen in stabilen Relativzahlen) einer Sorte über alle Versuche zum Ausdruck. Das statische oder biologische Konzept der Ökostabilität (Becker, 1981), nach dem eine Sorte unter allen Bedingungen gleiche Absoluterträge (o.ä.) realisiert, ist für die landwirtschaftliche Praxis nicht realistisch - eine derartige Form der absoluten Stabilität, z.B. durch absolute Trockentoleranz, ist kaum denkbar.

Die Ökoregression kann Hinweise darauf geben, warum Sorten u.U. vom agronomischen Konzept der Ertragsstabilität abwichen. Grundlage der Berechnung ist z.B. die sortenspezifische lineare Abhängigkeit der Leistung von den zugehörigen (ggf. adjustierten) Versuchsmitteln (Mittel aller Sorten je Einzelversuch). Sorten mit einem Anstieg über 1,0 weisen in ertragsstarken Versuchen tendenziell überdurchschnittliche Sorte $\times$ Umwelt - Wechselwirkungseffekte auf, sie konnten die günstigen Bedingungen überdurchschnittlich verwerten und werden häufig - mit gebotener Vorsicht - als ‚Intensivsorten' klassifiziert. Sorten mit einem Anstieg < 1,0 werden dagegen eher als ‚Extensivsorten' angesehen. Einschränkend muss gerade bezüglich der Begriffe *Extensivsorte* versus *Intensivsorte* betont werden, dass allein der Parameter der Ökoregression noch keine ursächliche Wirkung aufdeckt.

In der Bewertung der Eignung von Sorten für die Anbaugebiete in Mecklenburg-Vorpommern werden erste Ansätze genutzt, um Kennziffern der Umweltstabilität im Zuge der Anwendung der Hohenheim-Gülzower-Serienauswertung zu erzeugen (s. Anlage 9). Um das Risiko von Überinterpretationen zu verringern, wird die Ökovalenz - subjektiv erfahrungsbasiert - nur für Sorten ausgewiesen, von denen mindestens zwanzig Versuchsergebnisse vorliegen. Die Ökoregression wird nur für Sorten ausgewiesen, bei denen sich der Regressionskoeffizient signifikant von ‚1' unterscheidet. Diese Parameter dienen aber ausschließlich der Einbeziehung in die interne Gesamtbewertung. Aufgrund des erhöhten Abstraktionsniveaus werden diese Parameter im Beratungsmaterial für die breite Praxis nicht dargestellt. Die derzeitigen Ansätze erfordern nach Einschätzung des Autors dieser Arbeit aufgrund der Komplexität des Auswertungsmodells weitere methodische Untersuchungen, die nicht Teil dieser Arbeit werden sollen.

Prüfglied-Umwelt-Interaktionen stellen häufig, insbesondere auch im Sortenwesen, Wechselwirkungseffekte zwischen fixen und zufälligen Faktoren dar. Hierbei muss zwischen der *Modellformulierung mit unabhängigen* und der *Modellformulierung mit*

abhängigen Wechselwirkungen unterschieden werden (Searle et al., 1992; Richter et al., 1999). Bei unabhängigen Wechselwirkungen wird die Annahme gemacht, dass alle zufälligen Effekte (worin alle Kombinationseffekte zufälliger und fixer Faktoren eingeschlossen sind) unabhängig sind. Bei abhängigen Wechselwirkungen wird die Bedingung aufgenommen, dass die Summe über die zufälligen Wechselwirkungseffekte bei Summation über den Index eines fixen Faktors null ergibt. Damit sind die zufälligen Wechselwirkungseffekte auf verschiedenen Stufen einer fixen und gleichen Stufe eines zufälligen Faktors nicht mehr unabhängig. Beim Test der fixen Effekte gibt es zwischen beiden Ansätzen keinen Unterschied. Basford et al. (2004) weisen für züchterische Parameter darauf hin, dass zwar die Schätzung des Selektionsgewinns unabhängig von der Wahl der Formulierung ist, dass aber die genetische Korrelation und die Heritabilität unterschiedlich ausfällt - sie empfehlen für diese Zwecke das Modell mit abhängigen Wechselwirkungen. Die schätzbaren Funktionen sind bei dem Model mit abhängigen Wechselwirkungen durch den ‚Null-Summen-Ansatz' leichter nachvollziehbar. Die in dieser Arbeit verwendete Software SAS® nutzt aber grundsätzlich das Modell mit unabhängigen Wechselwirkungen. Nur schätzbare Funktionen, die unabhängig von der Annahme bezüglich der Wechselwirkung sind (Mittelwerte oder Differenzen), stellen interpretierbare Schätzungen dar.

Aus einer Vielzahl von Gesprächen und Diskussionen mit Versuchsanstellern, Sortenprüfern, mit der Versuchsauswertung betrauten Mitarbeitern und Biometrikern hat der Autor dieser Arbeit den Eindruck gewonnen, dass häufig nicht für jeden konkreten Einzelfall eine ergebnisoffene Abwägung z.B. dahingehend stattfindet, wie weit der Aussagebereich definiert werden soll, ob Testung oder Schätzung im Vordergrund stehen, welche Hypothesen für die Umweltfaktoren (z.B. Jahre und Orte) zu wählen sind oder welches Ausmaß der Unbalanciertheit hinnehmbar ist. Vielmehr zeichnen sich hierbei zwei subjektive Grundmuster im methodischen Herangehen ab, zu denen die jeweiligen Bearbeiter eher tendieren:

Die eine, tendenziell eher traditionelle ‚Schule' neigt dazu, überschaubare Datensätze in streng orthogonaler Struktur zu nutzen oder diese künstlich zu bilden. Dies erfolgt selbst dann, wenn wie im hier beschriebenen Fall des Sortenprüfsystems ca. 50% der beratungsrelevanten Daten ‚abgeschnitten' und per Definition nur die letzten drei Versuchsjahre genutzt werden. Letzteres hat dann u.a. wiederum zur Folge, dass Jahre eher wie Stufen eines fixen Faktors betrachtet werden. Ebenso werden Orte eher als fix betrachtet und häufig bereits so gewählt, dass jeder Ort auch innerhalb eines Anbaugebietes eine klar abgegrenzte Standortcharakteristik aufweist. Die Interpretation erfolgt dann vorrangig zurückgerichtet auf die kurze Versuchsperiode bzw. auf die Einzeljahresreaktionen und bezogen auf die konkreten Versuchsstandorte. Verallgemeinerungen für einen übergeordneten Aussagebereich (Anbaugebiet und dessen Klima) erfolgen mit Vorbehalt und sehr vorsichtig. Es besteht oft eine gewisse Scheu, sich von der erinnernden Beschreibung der selbst betreuten Versuche und ihrer sehr konkreten Umstände zu lösen und diese Versuche bzw. Orte und Jahre abstrahierend als eine Stichprobe der Grundgesamtheit einzuordnen. Den Aussagebereich entsprechend den Notwendigkeiten weiter zu fassen und z.B. Empfehlungen für andere Orte im Anbaugebiet in neuen Jahren abzuleiten, erscheint oft zu gewagt. Auch wird Testung häufig wie selbstverständlich als zentrale Aufgabe der biometrischen Auswertung betrachtet.

Das zweite beobachtete und auch in dieser Arbeit favorisierte Herangehen ist eher dadurch gekennzeichnet, die vorliegenden Daten umfassend zu nutzen, die gegebenenfalls systemimmanente Unbalanciertheit hinzunehmen und die immanen-

ten Gegebenheiten des Datensatzes durch Modellwahl, ggf. auch durch Transformation etc. bestmöglich zu berücksichtigen. Die Hypothesenwahl wird vorrangig von der ‚Natur' der Faktoren und dem Ziel der Versuchsanstellung abgeleitet und neigt für die Umweltfaktoren eher zum Ansatz ‚zufällig', da i.d.R. eine Aussage für ein Anbaugebiet und dessen Klima als Zielstellung gesehen wird. Demzufolge werden möglichst viele Jahre und Orte einbezogen und bereits bei der Wahl der Orte wird der Anspruch ‚*repräsentativ*' in den Vordergrund gestellt. Sofern nicht Grundsatzaussagen zur Wirkung von Faktoren und deren Stufen Zielstellung sind, sondern Beratungsaussagen, hat Schätzung im Falle immer wieder auftretender Zielkonflikte Vorrang vor Testung von Prüfgliedeffekten und -differenzen.

Die Modellbildung für hier diskutierte Sortenprüfsysteme wird - ausgehend von einfachsten Ansätzen und nachfolgender sukzessiver Aufnahme von relevanten Besonderheiten des Prüfsystems - im Abschnitt 5.3 dargelegt, wobei durchweg von Zweischrittanalysen und vom zufälligen Ansatz für Umwelten ausgegangen wird.

Die in dieser Gesamtschau auf die Versuchsserienauswertung aufgezeigten Aspekte tangieren alle den Gegenstand und die Zielstellungen (s. Abschnitt 2) der vorgelegten Arbeit bzw. der darin dargelegten methodischen Ansätze der Hohenheim-Gülzower-Serienauswertung. Die Problematik der systematischen Unbalanciertheit in Sortenprüfsystemen wird aufgezeigt und in ihrer Auswirkung auf die Schätzung von Sortenmittelwerten unter unterschiedlichen Szenarien untersucht. Bei Anwendung einer Zweischrittanalyse wird das Ausmaß der Differenziertheit der Präzision der Sortenmittelwerte aus Einzelversuchen gezeigt und Methodik und Nutzen einer Gewichtung mittels der Kehrwerte ihrer Fehlervarianzen diskutiert. Bezüglich der Einhaltung der Modellvoraussetzungen werden Ansätze zur Datentransformation diskutiert und in ihrer Auswirkung auf die Schätzung von Sorteneffekten untersucht. Bezüglich der Sorte $\times$ Umwelt - Wechselwirkungen werden zum einen unterschiedliche Regionalisierungsstrategien untersucht - insbesondere unter dem Aspekt, wie Nachbargebiete mit zu einem Zielgebiet korrelierten Sorteneffekten in die Auswertung für dieses Zielgebiet einfließen können. Zum anderen werden die Varianzkomponenten des Komplexes der Sorte $\times$ Umwelt - Wechselwirkungen für eine breite Palette von Pflanzenarten und Merkmalen in ihrer Bedeutung untersucht und interpretiert. Für das Sortenprüfsystem wird diskutiert, inwieweit eine BLUP-Schätzung, insbesondere der Aspekt der ‚Schrumpfung' für Sorteneffekte geeignet bzw. tragfähig ist. Letztlich wird hinterfragt, ob die mit dem Methodenkomplex der Hohenheim-Gülzower-Serienauswertung vorgenommen Schätzungen von Varianzen und Sorteneffekten hinreichend valide sind.

4 Material

4.1 Pflanzenarten

Der Begriff *Pflanzenart* wird hier im Sinne der Definition des Bundessortenamtes in der Beschreibenden Sortenliste (Bundessortenamt, 2009) verwendet. Er beinhaltet also neben der botanisch definierten Pflanzenart ggf. auch Unterteilungen in Sommer-/ Winterform oder in Nutzungsrichtungen. Die im Weiteren lt. Tab. 2 verwendeten Kurzbezeichnungen oder Kürzel enthalten Bezüge zur Sommer-/ Winterform und zur Nutzungsrichtung nur, wenn dies zur eindeutigen Unterscheidbarkeit erforderlich ist (z.B. reicht Raps, da in dieser Arbeit keine Sommerform einbezogen wird; es reicht Lupine, da nur Blaue Lupine geprüft wird etc.).

Die Analysen zu typischen Mustern im Datenverhalten (6.1.1), in der Struktur der Varianzkomponenten (6.2.1) und Muster der genetischen Korrelationen (6.3.1) wurden für alle hier bearbeiteten landwirtschaftlichen Pflanzenarten (Tab. 2) an den jeweiligen varianzanalytisch auswertbaren Merkmalen (s. a. 4.2) durchgeführt.

Tab. 2: In die Auswertungen einbezogene Pflanzenarten

Bezeichnung der Pflanzenarten lt. Beschreibender Sortenliste	**im Weiteren verwendete**	
	Kurzbezeichnung	**Kürzel**
Winterweichweizen	Winterweizen	WW
Sommerweichweizen	Sommerweizen	WS
Wintergerste	Wintergerste	GW
Sommergerste	Sommerfuttergerste	GS_F
	Sommerbraugerste	GS_B
Winterroggen	Roggen	RW
Wintertriticale	Triticale	TIW
Sommerhafer	Hafer	HA
Winterraps	Raps	RAW
Futtererbse	Erbse	EF
Blaue Lupine	Lupine	LUB
Mais, Körnernutzung	Körnermais	MK
Mais, Silonutzung	Silomais	MS
Kartoffeln, Wirtschaftssorten	Kartoffeln	KA

Weitergehende Analysen zu Effekten der Transformation (6.1.2), der Einbeziehung vorgelagerter Versuche (6.2.2), der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen (6.2.3) und der Einbeziehung von Nachbargebieten mit abgestufter Gewichtung (6.3.2) auf Mittelwerte und Genauigkeit erfolgten nur für einzelne Merkmale, schwerpunktmäßig für den Winterweizen am Merkmal Kornertrag. Für diese Variable liegt mit Abstand der größte langjährige Datensatz vor (N=9993 s. Tab. 6). Transformationsparameter, Varianz-

komponenten und genetische Korrelationen waren hier also besonders präzise schätzbar.

4.2 Merkmale und deren deskriptive Beschreibung

Ein Großteil der in den Landessortenversuchen in Mecklenburg-Vorpommern metrisch erfassten wertbestimmenden Leistungs- und Qualitätsmerkmale wurde für die Analysen herangezogen.

Die Datenhaltung für Landessortenversuche erfolgt bundesweit in Datenbanksystemen der Software *PIAF* (Planungs-, Informations- und Auswertungssystem für das Feldversuchswesen) (Michel et al., 2007b). Auch der Datenaustausch zwischen den Dienststellen der Bundesländer und dem Bundessortenamt erfolgt über die so genannte PIAF-Schnittstelle. Daher wird hier die datenbankgerechte Nomenklatur für die Merkmale im PIAF-System verwendet.

Die Merkmalscharakterisierung erfolgt in Tab. 3 und 4. Nachfolgend werden Anmerkungen und Erläuterungen zu dieser Merkmalscharakterisierung - Bezug nehmend auf die Spaltenüberschriften - gegeben.

ad Spalte ‚Merkmalskategorie:

Die ausgewerteten Merkmale wurden hier in die Kategorien *Ertrag*, *Minderertrag* (s.u. *Intensitätsstufe*), Qualitätsmerkmale in Prozentangabe (*Qualität [%]*) sowie sonstige Qualitätsmerkmale (*sonst. Qualität*) gruppiert. Veranlassung für diese Gruppierung war die subjektive Erwartung, dass diese Kategorien unterschiedliche Eigenschaftsmuster hinsichtlich der Erfüllung von Modellvoraussetzungen, des Transformationserfordernisses, in der Ausprägung der Varianzkomponenten u.ä. aufweisen und somit unterschiedlichen Schlussfolgerungen bei der Bearbeitung der Zielstellungen zulassen. Ebenso ist es von Interesse, ob innerhalb einer Kategorie Pflanzenart übergreifend vergleichbare Muster zu finden sind. Bei den Qualitätsmerkmalen in Prozentangabe handelt es sich in keinem Fall um diskrete Daten (z.B. gezählte Anteile), sondern ausschließlich um kontinuierliche Daten (Masseanteile; die Grenzwerte ‚0‘ und ‚1‘ (0 und 100%) wurden in keinem Fall erreicht).

ad Spalte ‚Intensitätsstufe’:

Bei Getreidearten und Raps kamen häufig zweifaktorielle Versuchsanlagen zum Einsatz, wobei zum Faktor ‚Sorte‘ dann der mit ‚Sorte‘ kreuzklassifizierte Faktor ‚Intensität‘ hinzukommt. Der Faktor ‚Intensität‘ wurde dann ausnahmslos in zwei Stufen (‚Stufe I = reduziert‘ und ‚Stufe II = ortsüblich‘) definiert. In der Versuchsserienauswertung ist in keinem Fall der zweifaktorielle Ansatz übernommen worden, sondern entweder Auswertungen auf einer Intensitätsstufe, Auswertungen im Stufenmittel je Sorte oder Auswertungen der Differenzen je Sorte zwischen den beiden Intensitäten. Eine tiefere Beschreibung der Intensitätsstufen bzw. Differenzen zwischen Intensitätsstufen (hier *Minderertrag geringe Intensität* = Stufe II - Stufe I) und der Auswertungsansätze bei den einzelnen Pflanzenarten erfolgt unter 4.3.1. An dieser Stelle sei auf die besondere Problematik der differenzierten Faktorstruktur in den Versuchsanlagen beim Raps verwiesen, die ebenfalls unter 4.3.1 diskutiert wird.

ad Spalten ‚Merkmalsbezeichnung im PIAF-System' und ‚Label kurz':

Das Merkmal ‚Minderertrag geringe Intensität' wird nicht in der PIAF-Datenbank verwaltet, sondern nur temporär durch Differenzbildung aus den Originaldaten erzeugt.

ad Spalte ‚Skalierung'

Alle mit den hier diskutierten Methoden analysierten Merkmale sind metrisch erfasst und verhältnis- oder intervallskaliert, so dass unter der Voraussetzung gegebener weiterer Modellvoraussetzungen parametrische, lineare Modelle prinzipiell zulässig sind.

Tab. 3: Einbezogene Merkmale; Teil 1 Kategorien ‚Ertrag' und ‚Minderertrag'

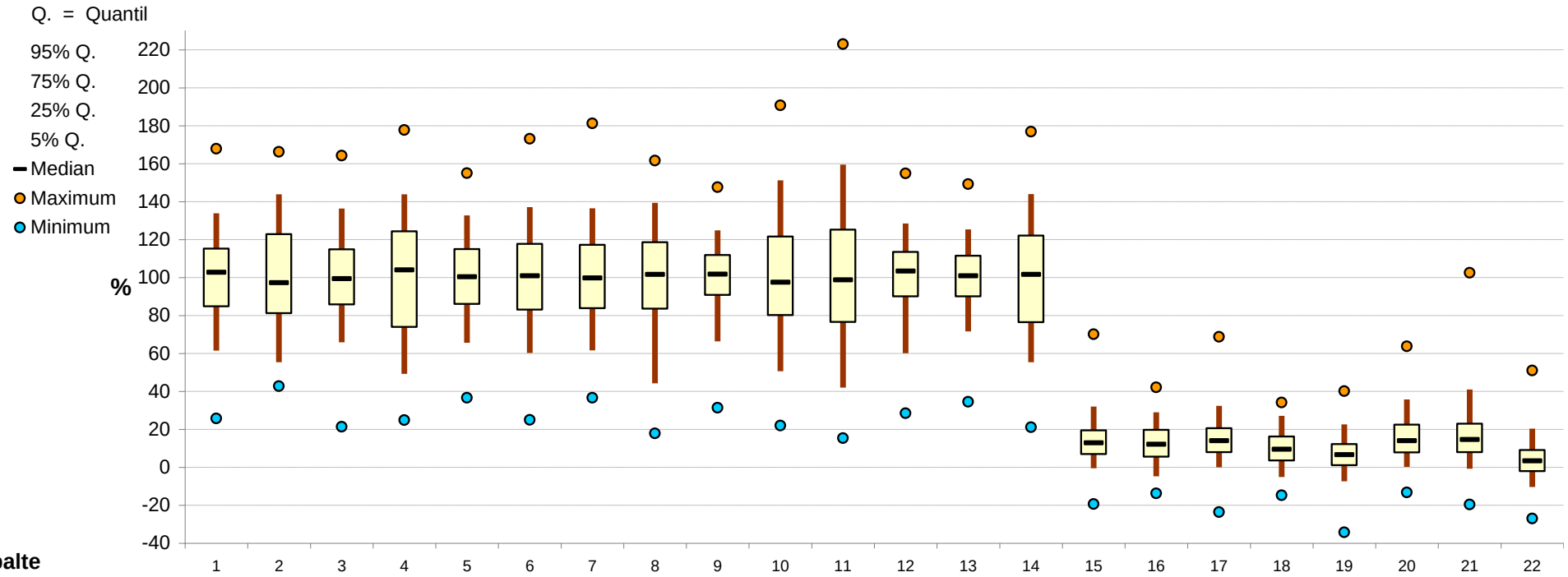
Merkmals-kategorie	**Pflanzenart (Kurzbezeichnung)**	**Intensitäts-Stufe** *	**Merkmalsbezeichnung im PIAF-System**	**Label kurz**	**Einheit**	**Skalierung**
Ertrag	Winterweizen	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Sommerweizen	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Wintergerste	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Sommerfuttergerste	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Sommerbraugerste	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Roggen	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Triticale	II	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Hafer	einfaktoriell	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Raps	einfaktoriell bzw. (II+I)/2	Kornertrag bei 91% TS dt/ha	ERTR91DT	dt/ha	Verhältnisskala
Ertrag	Erbse	einfaktoriell	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Lupine	einfaktoriell	Kornertrag bei 86% TS dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Körnermais	einfaktoriell	Ertrag (Körnermaiskorr.) dt/ha	ERTR86DT	dt/ha	Verhältnisskala
Ertrag	Silomais	einfaktoriell	Gesamttrockenmasse (Silomais NIRS) dt/ha	TMGEDTHA	dt/ha	Verhältnisskala
Ertrag	Kartoffeln	einfaktoriell	Knollenertrag dt/ha	KNO_ERDT	dt/ha	Verhältnisskala
Minderertrag	Winterweizen	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Sommerweizen	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Wintergerste	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Sommerfuttergerste	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Sommerbraugerste	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Roggen	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Triticale	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala
Minderertrag	Raps	II - I	Minderertrag geringe Intensität *	MinderE	dt/ha	Intervallskala

* Stufen I (ortsüblich) und II (reduziert) s. 4.3.1; Minderertrag geringe Intensität' wird nicht im PIAF-System verwaltet

Tab. 4: Einbezogene Merkmale; Teil 2 Kategorien *Qualität [%]* und *sonstige Qualität*

Merkmals-kategorie	**Pflanzenart (Kurzbezeichnung)**	**Intensitäts-Stufe** *	**Merkmalsbezeichnung im PIAF-System**	**Label kurz**	**Einheit (/ *Bezug*)**	**Skalierung**
Qualität [%]	Silomais	einfaktoriell	TS Gesamtpflanze (NIRS) %	TSNIGEPF	**%**	Verhältnisskala
Qualität [%]	Körnermais	einfaktoriell	Trockensubstanz (Korn) %	TS_KORN	**%**	Verhältnisskala
Qualität [%]	Raps	einfaktoriell	Glucosinolatgehalt bei 91 % TS (%)	GLUC91	**%**	Verhältnisskala
Qualität [%]	Raps	einfaktoriell	Öl bei 91 % TS (%)	FETT91	**%**	Verhältnisskala
Qualität [%]	Erbse	einfaktoriell	Rohprotein bei 86% TS	RP__86	**%**	Verhältnisskala
Qualität [%]	Lupine	einfaktoriell	Rohprotein bei 86% TS	RP__86	**%**	Verhältnisskala
Qualität [%]	Silomais	einfaktoriell	Stärkegehalt in Gesamtpflanze %	STAEGEH	**%**	Verhältnisskala
Qualität [%]	Kartoffeln	einfaktoriell	Stärkegehalt in Knolle %	STAEGEKN	**%**	Verhältnisskala
Qualität [%]	Sommerbraugerste	II	Sortierung > 2.5 mm (%)	SORTGR25	**%**	Verhältnisskala
Qualität [%]	Sommerbraugerste	II	Rohprotein (Korn/Kern) in TM (%)	RP__I_TM	**%**	Verhältnisskala
sonst. Qualität	Sommerbraugerste	II	Tausendkornmasse, lufttrocken g	TKG_LUTR	**g** */1000 K.*	Verhältnisskala
sonst. Qualität	Sommerbraugerste	II	Hektolitergewicht kg	HLG	**kg** */hl*	Verhältnisskala
sonst. Qualität	Lupine	einfaktoriell	Tausendkornmasse bei 86% TS (g)	TKG_86	**g** */1000 K.*	Verhältnisskala
sonst. Qualität	Silomais	einfaktoriell	NEL/kg TM (NIRS) (ELOS, NDForg, XL, XA)	NELKGX	**MJ** *NEL /kg*	Verhältnisskala

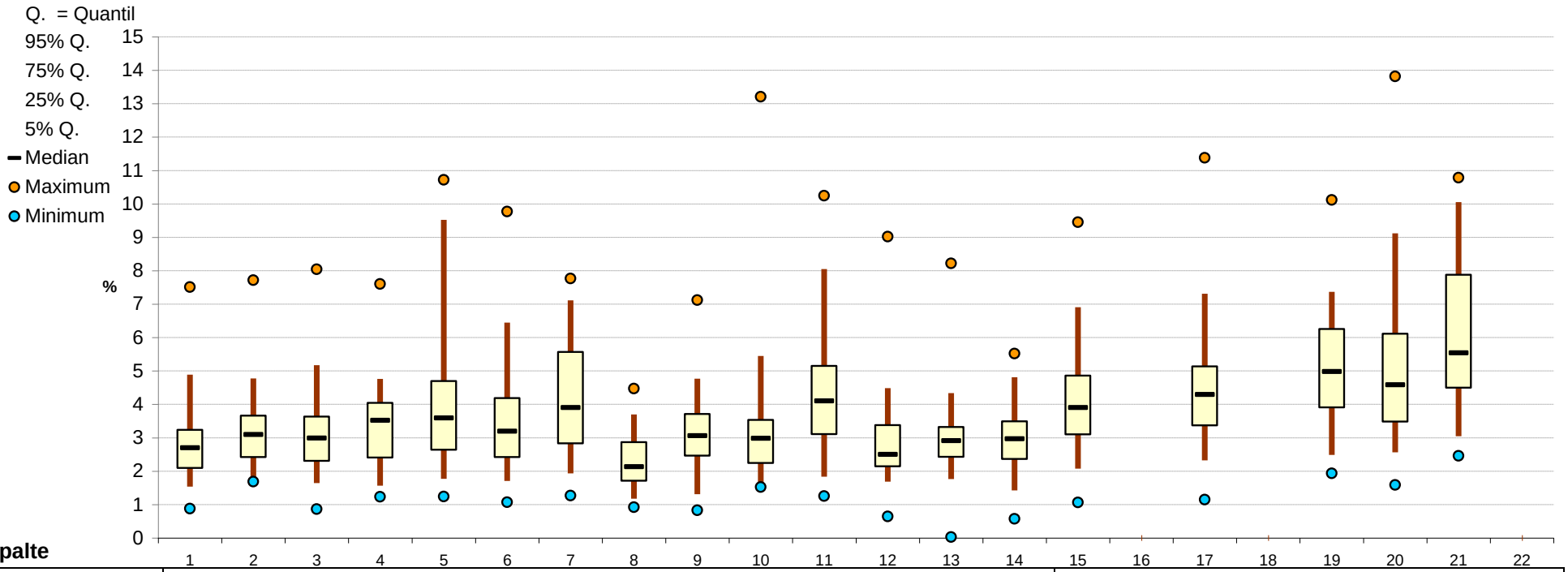
* Stufen I und II s. 4.3.1



Kategorie	**Ertrag** (dt/ha); in zweifaktoriellen Versuchen in der Intensitätsstufe lt. Tab. 3)														**Minderertrag** (dt/ha)							
Relativ zu	arithmetischem Mittel der Urwerte														arithmetischem Mittel in der Kategorie ‚Ertrag' der gleichen Pflanzenart (in Spalte 1-14) (Stufe II -Stufe I)							
Pflanzenart	WW	WS	Gw	GS_F	GS_B	RW	TlW	Ha	RAW	EF	LUB	MK	MS	KA	WW	WS	GW	GS-F	GS-B	RW	TlW	RAW
100 % entspricht	94,1	62,7	87,3	54,6	62,6	81,8	88,1	62,8	51,0	49,6	28,8	104	178	547	siehe Spalten-Nr. 1 bis 14 unter gleicher Pflanzenart							
N	9993	346	5973	384	2879	3098	3172	1343	6314	1591	805	1410	7179	1373	9104	338	5656	346	2778	3051	3110	1972

N = Anzahl Sortenmittelwerte aus Einzelversuchen

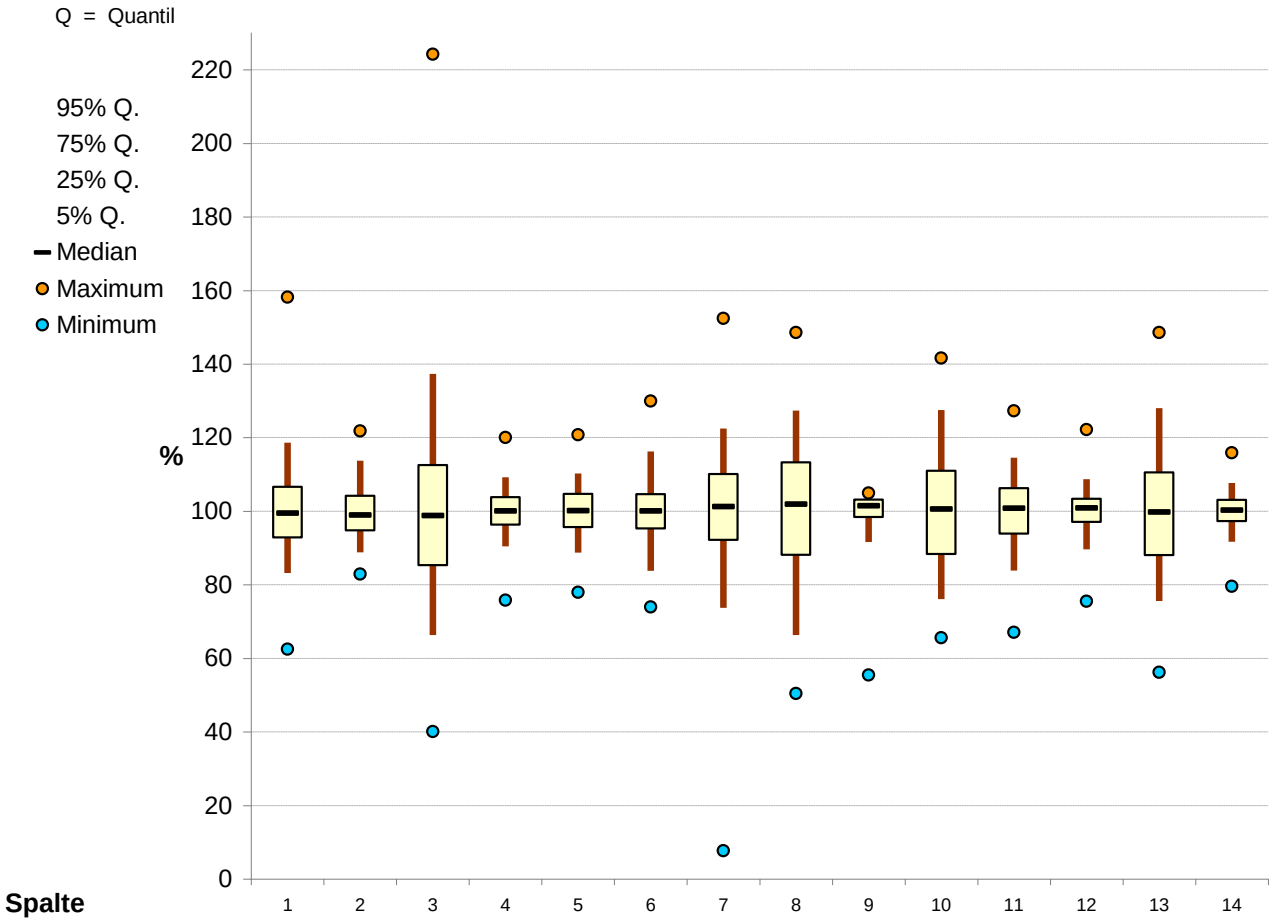
Abb. 1: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Ausgangsdaten; Merkmale der Kategorien ‚Ertrag' und ‚Minderertrag'



Kategorie	**Ertrag** (dt/ha); in zweifaktoriellen Versuchen in der Intensitätsstufe lt. Tab. 3)														**Minderertrag** (dt/ha)							
Relativ zu	arithmetischem Mittel der Urwerte														arithmetischem Mittel in der Kategorie ‚Ertrag' der gleichen Pflanzenart (in Spalte 1-14) (Stufe II -Stufe I)							
Pflanzenart	WW	WS	GW	GS_F	GS_B	RW	TlW	Ha	RAW	EF	LUB	MK	MS	KA	WW	WS	GW	GS-F	GS-B	RW	TlW	RAW
100% entspricht	94,1	62,7	87,3	54,6	62,6	81,8	88,1	62,8	51,0	49,6	28,8	104	178	547	siehe Spalten-Nr. 1 bis 14 unter gleicher Pflanzenart							
N	8874	336	4885	330	2583	2681	2296	1178	5786	1358	674	1200	6187	989	8181	**kein SE**	4622	**kein SE**	1017	2634	2234	**kein SE**

N = Anzahl Sortenmittelwerte aus Einzelversuchen

Abb. 2: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Variationskoeffizienten der Mittelwerte in Einzelversuchen ($SE_{Versuch}$%); Merkmale der Kategorien ‚Ertrag' und ‚Minderertrag'



Kategorie	**Qualität [%]**										**sonstige Qualität**			
Relativ zu	arithmetischem Mittel der Urwerte													
Pflanzenart Merkmal	MS TSNIGEPF	MK TS-KORN	RAW GLUC91	RAW FETT91	EF RP_86	LUB RP_86	MS STAEGEH	KA STAEGEKN	GS-B SORTGR25	GS-B RP_I_TM	GS-B TKG_LUTR	GS-B HLG	LUB TKG_86	MS NELKGX
100 % entspricht	**34,0** %	**72,6** %	**14,5** %	**42,5** %	**19,9** %	**30,0** %	**32,4** %	**18,8** %	**95,0** %	**11,4** %	**48,9** g	**70,1** kg /hl	**155** g	**6,47** MJ
	7281	1410	3384	4253	1124	582	7302	1373	775	764	1045	855	743	7147

N = Anzahl Sortenmittelwerte aus Einzelversuchen

Abb. 3: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Ausgangsdaten; Merkmale der Kategorien ‚Qualität [%]' und ‚sonstige Qualität'

Abb. 1 und 3 charakterisieren die Datensätze deskriptiv hinsichtlich des Datenumfangs, der relativen Variabilität und der Symmetrie der Urwerte für die einzelnen Merkmale und Merkmalskategorien. Nachfolgend erfolgt jeweils eine kurze Diskussion der erkennbaren Besonderheiten bei einzelnen Pflanzenarten und Merkmalen.

ad Abb. 1 ; Kategorie ‚Ertrag'

Die Erträge wurden auf das arithmetische Mittel aller Teilstückswerte (Parzellenwerte) der in Tab. 3 ausgewiesenen Intensitätsstufe relativiert, um eine Vergleichbarkeit zwischen den Pflanzenarten zu erzielen.

Es ist ersichtlich, dass die relative Variation bei den Pflanzenarten ähnlich groß, also relativ unabhängig vom Niveau des Wertebereiches der Versuchsdaten ist (Zeile ‚100%' reicht von 28,8 dt/ha bei Lupine bis 547 dt/ha bei Kartoffeln). Allerdings zeigt die Lupine die größte relative Schwankungsbreite, insbesondere gibt es Einzelwerte mit mehr als doppelten Erträgen gegenüber dem Mittelwert. Dies wird vom Autor ursächlich damit in Zusammenhang gebracht, dass Lupinen überwiegend auf ausgesprochen ertragsschwachen Standorten geprüft werden, aber auch einzelne Versuche mit erheblich besseren Bedingungen enthalten sind. Der Raps zeigt eine vergleichsweise geringe relative Schwankungsbreite. Als grobe Charakterisierung kann im Durchschnitt der Pflanzenarten eingeschätzt werden: im Bereich Mittelwert ± 20% liegen etwa 50% der Werte, im Bereich Mittelwert ± 40% liegen etwa 90% der Werte, im Bereich Mittelwert ± 70% liegen nahezu alle Werte. Das Maximum-Minimum - Verhältnis liegt damit häufig über 5:1, was u.a. hinsichtlich eines möglichen Transformationserfordernisses von Interesse sein wird (s. 6.1).

Die Symmetrie der Datenverteilung ist überwiegend relativ gut gegeben.

Der Stichprobenumfang ‚N' ist über die Pflanzenarten sehr differenziert. Dies muss u.a. bei der Interpretation der Minimum - Maximum - Spanne berücksichtigt werden, da die Spannweite bei gleicher Varianz tendenziell mit größer werdendem Datensatz steigt. Die größten Datensätze weisen Winterweizen (N=9993), Silomais (N=7179) und Raps (N=6314) auf. Dagegen liegen für Sommerfuttergerste (N=384) und Sommerweizen (N=346) nur deutlich geringere Datenumfänge vor.

Bezüglich der späteren Diskussion einer eventuell förderlichen Datentransformationen ist hier schon anzumerken, dass es sich bei den meisten Ertragsmerkmalen nicht um original erfasste Messwerte an der Einzelparzelle handelt, sondern dass durch multiplikative Verknüpfung einer Ertragsmessung mit dem ebenfalls ermittelten Trockensubstanzgehalt rechnerisch ein Flächenertrag bei einem definierten einheitlichen Trockensubstanzgehalt ermittelt wurde. Erst dieser wurde dann varianzanalytisch ausgewertet und in die Versuchsserienauswertung einbezogen. Dies trifft bei den hier ausgewerteten Daten für die Trockenmasse-Erträge bei Silomais und Körnermais zu, da hier neben dem Rohertrag auch der Trockensubstanzgehalt je Teilstück erfasst wurde. Bei Mähdruschfrüchten bestand diese Problematik noch nicht, da der Trockensubstanzgehalt aus Kapazitätsgründen nur an einer Mischprobe je Prüfglied erfasst wurde. Inzwischen werden aber auch bei Mähdruschfrüchten im Zuge der Einführung von Messungen direkt beim Mähdrusch Gewicht und Trockensubstanzgehalt je Teilstück erfasst, so dass künftig diese Problematik an Relevanz gewinnen wird. Mit der Problematik multiplikativ verknüpfter Ausgangsmerkmale befassen sich Piepho et al. (2014). Im Falle von Fehlwerten bei jeweils nur einem der Ausgangsmerkmale schlagen sie eine multivariate Auswertung vor, um die Information des jeweils anderen Ausgangsmerkmals nicht zu verlieren.

Dazu werden die Daten vorab logarithmisch transformiert, um das Produkt der Merkmalskomponenten in eine Summe zu wandeln und ein lineares Modell nutzen zu können.

ad Abb. 1 ; Kategorie ‚Minderertrag'

Mindererträge sind für Pflanzenarten mit zweifaktorieller Versuchsanlage ausgewiesen. Hier wurde neben dem Faktor *Sorte* der dazu kreuzklassifizierte Faktor *Intensität* mit zwei Stufen geprüft (Stufe I = reduziert; Stufe II = ortsüblich mit Fungiziden und Wachstumsreglern). Der Minderertrag ist als Differenz (Stufe II - Stufe I) definiert. Eine vertiefte Charakterisierung des Faktors *Intensität* erfolgt in Abschnitt 4.3.1.

Die Skala hat keinen absoluten Nullpunkt, die Daten sind intervallskaliert und Unterschiede zwischen Sorten können also nur in Differenzen, nicht in Verhältnissen ausgedrückt werden (Richter 2004). Daher werden die Mindererträge nicht auf das Sortenmittel des Minderertrages, sondern auf das arithmetische Mittel aller Urwerte des ebenfalls in der Abbildung ausgewiesenen *Ertrages* (linke Kategorie) relativiert. Die Relevanz liegt auch sachlich eher in Relation zum Ertragsniveau z.B. in Stufe II als zum Niveau des Minderertrages. Eine Ausnahme hierbei ist der Raps, da hier der Minderertrag nicht unmittelbar auf den dargestellten Ertrag bezogen werden kann (s. 4.3.1). Beim Raps ist damit zwar die Lage aber nicht die Streuung mit den anderen Arten unmittelbar vergleichbar. Die größten Einzelwerte beim Minderertrag wurden in den Pflanzenarten Triticale (90 dt/ha), Winterweizen (66 dt/ha) und Wintergerste (60 dt/ha) festgestellt. Mindererträge lagen z.T. unter 0, d.h. in der intensiveren Stufe II wurden in diesen Fällen geringere Erträge als in der reduzierten Stufe I festgestellt. Ob dies im Zufallsbereich lag oder im Einzelfall mit negativen Wirkungen von Pflanzenschutzmitteln im Zusammenhang steht, soll hier nicht vertieft werden. Der Hälftenspielraum (Box) lag etwa im Bereich 5% bis 20% des Ertrages (nicht des Minderertrages!) (Ausnahme Raps, nicht direkt vergleichbar).

Die Datenumfänge für den Minderertrag sind gegenüber dem Ertrag etwas kleiner, da einzelne Versuche nur einfaktoriell angelegt oder nur in einer Intensitätsstufe ausgewertet wurden (z.B. Winterweizen N=9104 entspricht 91% der ‚Ertrags'-Daten).

ad Abb. 2 ; Variationskoeffizienten der Mittelwerte aus Einzelversuchen

Die in der Einzelversuchsauswertung ermittelten Standardfehler der Sortenmittelwerte sind als Variationskoeffizienten dargestellt ($SE_{Versuch}$%). Auffällig ist die ausgeprägte Schiefe der Verteilung. Das Gros der Qualitätsmerkmale wurde in den Einzelversuchen nur an einer Mischprobe der Teilstücke je Prüfglied analysiert. Daher liegen hier keine Einzelversuchsfehler vor und die Darstellung in Abb. 2 entfällt für diese Merkmalskategorien. Die $SE_{Versuch}$ werden ggf. für eine Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen herangezogen (s. 5.3.3).

ad Abb. 3 ; Kategorien ‚Qualität [%]' und ‚sonstige Qualität'

Um den Vergleich zwischen den Arten zu ermöglichen, wurden die Qualitäten auf das arithmetische Mittel aller Urwerte relativiert. Für die Interpretation sei darauf hingewiesen, dass damit in der Kategorie ‚Qualität [%]' eine Relativierung von Relativzahlen erfolgte. Diese Darstellungsweise ist nicht für alle Betrachtungen geeignet und kann einen verzerrten Eindruck vermitteln. Die tatsächlich gemessenen Spannen auf der Prozent-Skala sind im Zusammenhang mit der Frage des

Transformationserfordernisses bei Relativzahlen in Abb. 24 (6.1.1) dargestellt. Ein auffällig niedriges relatives Minimum ist beim Stärkegehalt von Silomais zu beobachten. Ursache hierfür war die äußerst schlechte Auskörnung in einigen Silomaisversuchen unter extremer Trockenheit im Jahre 2006. Diese Versuche wurden trotzdem einbezogen, um Sortenreaktionen gerade unter Trockenheit einfließen zu lassen. Eine erhebliche Linksschiefe ist bei dem Merkmal *‚Sortierung > 2,5 mm (%)‘* bei Sommerbraugerste zu beobachten. Das arithmetische Mittel liegt mit 95% bereits nahe der Obergrenze des Parameterraumes (100%). Für dieses Merkmal ist daher die Frage der Transformation mit möglichst guter Modellanpassung von besonderem Interesse (s. 5.2; 6.1). Eine besonders geringe Variabilität der Originalwerte weist das Merkmal *Rohprotein bei 86% TS* bei der Erbse auf (s.a. Abb. 24).

4.3 Faktoren und Faktorstufen

Zunächst werden die Faktoren sowie weitere Klassifizierungsgrößen, welche nicht als Faktor in die Modellgleichungen zur Versuchsserien-Auswertung einfließen, aufgelistet. Die in Tab. 5 angegebenen Kürzel werden nachfolgend verwendetet.

Tab. 5: Faktoren, Klassifizierungsgrößen und verwendete Kürzel

	Faktoren im Auswertungsmodell der Versuchsserien		**weitere Klassifizierungsgrößen (nicht im Auswertungsmodell)**	
	Kürzel	**Bezeichnung**	**Kürzel**	**Bezeichnung**
Einzelversuch (s. 4.3.1)	S	Sorte	I	Intensität
	Gr	Sortengruppe	(B)	Block (nur Einzelversuch)
zusätzlich in **Versuchsserie** (s. 4.3.2)	J	Jahr	BKR	Boden-Klima-Raum
	O	Ort	L	Land
	T	Versuchstyp		
	AG	Anbaugebiet		

Die Datensatzbeschreibungen (Tab. 6) entsprechen dem Datenbestand und Auswertungsstand zum 31.12.2009. Nach diesem Stichtag erfolgte Ergänzungen des Datenbestandes oder neue Auswertungen sind in diese Arbeit nicht mehr eingeflossen. Lediglich im Rahmen einer Validierung erfolgte für den Kornertrag bei Winterweizen zusätzlich eine Analyse für das Folgejahr 2010 (s. 6.5.3).

Tab. 6: Anzahl Faktorstufen und Werte in den Datasets der Auswertungsserien

Pflanzenart / Merkmal	Anzahl Stufen der Faktoren / Klassifizierungsgrößen (Kürzel entsprechend Tab. 5)									Anzahl Werte
	S	Gr	I	J	O	T *	AG	BKR	L	
Winterweizen										
ERTR86DT	1118	7	2	9	39	4	5	11	5	9993
MinderE	1118	7	-	9	36	4	3	11	5	9104
Sommerweizen										
ERTR86DT	44	3	2	9	6	2	1	5	4	346
MinderE	44	3	-	9	6	2	1	5	4	338
Wintergerste										
ERTR86DT	751	4	2	9	31	6	3	11	5	5973
MinderE	751	4	-	9	30	6	3	11	5	5656
Sommerfuttergerste										
ERTR86DT	34	1	2	9	11	1	1	6	5	384
MinderE	34	1	II - I	9	10	1	1	6	5	346
Sommerbraugerste										
ERTR86DT	437	3	2	9	20	4	2	6	5	2879
MinderE	420	3	-	9	20	3	2	6	5	2778
SORTGR25	267	3	1	8	14	2	2	5	5	775
RP__I_TM	266	3	1	8	14	2	2	6	5	764
TKG_LUTR	279	3	1	8	15	3	2	5	5	1045
HLG	270	3	1	8	14	3	2	5	5	855
Roggen										
ERTR86DT	162	3	2	9	30	3	2	9	5	3098
MinderE	161	3	-	9	29	3	2	9	5	3051
Triticale										
ERTR86DT	315	1	2	9	30	4	2	10	5	3172
MinderE	315	1	-	9	30	4	2	10	5	3110
Hafer										
ERTR86DT	166	1	1		17	4	2	8	5	1343
Raps										
ERTR86DT	794	2	2	9	33	3	5	10	5	6314
MinderE	96	2	-	10	23	2	2	10	5	1972
GLUC91	154	2	1	9	25	3	2	9	4	3384
FETT91	449	2	1	9	29	3	2	10	5	4253
Erbse										
ERTR86DT	161	1	1	9	21	3	2	5	5	1591
RP__86	161	1	1	9	16	3	2	5	5	1124
Lupine										
ERTR86DT	42	2	1	9	16	2	1	5	6	805
RP__86	42	2	1	9	13	2	1	4	6	582
TKG_86	42	2	1	9	16	2	1	5	6	743
Körnermais										
ERTR86DT	197	9	1	9	20	3	2	9	6	1410
TS_KORN	197	9	1	9	20	3	2	9	6	1410
Silomais										
TMGEDTHA	726	11	1	8	38	10	3	11	5	7179
TSNIGEPF	747	11	1	8	38	10	3	11	5	7281
STAEGEH	735	11	1	8	38	10	3	11	5	7302
NELKGX	735	11	1	8	37	9	3	10	5	7147
Kartoffeln										
KNO_ERDT	208	3	1	6	8	3	1	4	4	1373
STAEGEKN	208	3	1	6	8	3	1	4	4	1373

* Anzahl Versuchstypen, die im Maximum an einer O × J - Kombination stand

Für das in dieser Arbeit besonders tiefgehend analysierte Merkmal Kornertrag beim Winterweizen werden in Tab. 7 die Faktorstufen (außer Sorte und Ort) im Einzelnen aufgeführt. Eine Erläuterung bzw. Decodierung der Faktorstufen erfolgt in 4.3.1. Die Struktur der zur Verrechnung eingesetzten Daten-Datei für diesen Datensatz ist in Anlage 1 auszugsweise dargestellt.

Tab. 7: Verwendete Label für Stufen der Faktoren und Klassifizierungsgrößen am Beispiel Winterweizen; Datensatz für Auswertung des Kornertrages für Zielgebiete in Mecklenburg-Vorpommern

Faktor	**Auflistung der Label der Stufen ***	**Verweis zur Erläuterung**	**Anzahl Stufen**
Faktoren im Auswertungsmodell			
S	... (N=1118)	4.3.1	1118
Gr **	A B BH C CH E CK	4.3.1	7
J	2001 2002 2003 2004 2005 2006 2007 2008 2009	-	9
T ***	BSV EU EUSV LSV LSVfr OV WP1 WP2 WP3	4.3.2	9
O	... (N=39)	4.3.2	39
AG	3 4 5 45 82	5.1.2	5
Klassifizierungsgrößen			
BKR	101 102 104 109 146 151 153 154 156 157 158	5.1.1	11
Land	BB MV NI SH ST	4.3.2	5

* bei Sorte und Ort nur Angabe der Anzahl Stufen (N)

** Qualitätsgruppen A, B, C, E und K lt. Beschreibender Sortenliste 2009 (Bundessortenamt, 2009) mit dem Anhang „H“ bei Hybridsorten

*** Versuchstypen s. Abb. 4; BSV, EU und LSVfr sind in Abb. 4 nicht aufgeführt, sie haben keine Bedeutung für das Prüfsystem und wurden nur im Einzelfall verwendet, um angelegte Sonderversuche zu kodieren und sie damit von Versuchen an gleicher O×J - Kombination abzugrenzen

BB - Brandenburg; MV - Mecklenburg-Vorpommern; NI - Niedersachsen; SH - Schleswig-Holstein; ST - Sachsen-Anhalt;

Schlüssel für Anbaugebiet s. Abb. 6; Schlüssel für Boden-Klima-Raum s. Abb. 7

4.3.1 Faktoren bzw. Klassifizierungsgrößen in Einzelversuchen

Sorte (S)

Der Faktor *Sorte* ist der zentrale Gegenstand der hier diskutierten Versuchsserienauswertungen. Diese Auswertungen haben letztlich das Ziel der Schätzung der mittleren Sortenleistung (bzw. der Sorteneffekte) in einem Anbaugebiet im Vergleich zu anderen Sorten, wobei der Vorhersagefehler für neue Schläge im Anbaugebiet in einem neuen Jahr minimiert werden soll.

Unter dem Faktor *Sorte* werden hier nicht nur Genotypen von landwirtschaftlichen Pflanzenarten verstanden, die bereits in Deutschland oder einem anderen EU-Land zugelassen und damit in Deutschland vertriebsfähig sind, sondern darüber hinaus auch oft so genannte *Stämme*, die bereits aufgrund eines Antrages auf Sortenzulassung in der deutschen Wertprüfung stehen, aber nicht zugelassen sind. Im

Weiteren wird entsprechend der Nomenklatur beim Bundessortenamt aber ausschließlich der Terminus *Sorte* verwendet.

Die Anzahl Faktorstufen reichte hier von 34 Sorten bei der Sommerfuttergerste bis 1118 Sorten beim Winterweizen (Tab. 6). Diese Bandbreite ist in der unterschiedlichen Anbaubedeutung und daraus folgend in der unterschiedlich intensiven züchterischen Bearbeitung der Pflanzenarten begründet. Hinzu kommt, dass z.B. bei der Sommerfuttergerste keine Ergebnisse aus Wertprüfungen einbezogen werden können, weil die Wertprüfung ausschließlich als Sommerbraugerste angelegt wird. Dagegen ist z.B. beim Winterweizen eine große Anzahl der Sorten noch nicht zugelassen, da jedes Jahr in der Regel deutlich mehr als 100 Sorten neu in die Wertprüfung gestellt werden, von denen nur die wenigsten zur Zulassung und in Landessortenversuche gelangen.

Für die einzelnen Merkmale innerhalb jeweils einer Pflanzenart lagen ebenfalls z.T. Ergebnisse von einer sehr unterschiedlichen Anzahl Sorten vor. So waren es z.B. bei Raps im Merkmal Ertrag (ERTR86DT) Ergebnisse von 794 Sorten, im Minderertrag (MinderE) bei Raps aber lediglich 96 Sorten (Tab. 6). Dies hängt bei diesem konkreten Beispiel damit zusammen, dass Mindererträge nur in zweifaktoriellen Sortenversuchen bestimmt werden können, Wertprüfungen mit Raps aber einfaktoriell sind. Entsprechend ist für nahezu alle nicht zugelassenen Rapssorten im Datensatz kein *Minderertrag* enthalten. Auch Qualitätsmerkmale liegen z.T. nur für einen eingeschränkten Datenbestand vor.

Das Ziel der bestmöglichen Schätzung bzw. Vorhersage konzentriert sich jeweils auf die Sorten, die aktuell vertriebsfähig sind und zumindest eine potenzielle Anbaubedeutung und Empfehlungschance besitzen. Dies sind Sorten, die im jeweiligen Auswertungsjahr im aktuellen Sortiment des Landessortenversuches stehen oder die zwar aufgrund genügender langjähriger Ergebnisse aus dem aktuellen Landessortenversuch genommen wurden, aber trotzdem noch empfohlen und dargestellt werden. Die Menge dieser Sorten aus der Obermenge aller Sorten (Gesamtsortiment) wird im Weiteren als Kernsortiment bezeichnet. Beim Winterweizen stellen bei einem Gesamtsortiment von 1118 ausgewerteten Sorten nur 44 Sorten (4%) das Kernsortiment dar (Tab. 8). Auf dieses Kernsortiment ist das Hauptinteresse der Auswertung gerichtet. Es stellt den Kern der Datenstruktur mit relativ guter Balanciertheit dar, es bildet aber keinen orthogonalen Kern. Dieses Kernsortiment wird im Folgenden in acht nach dem Prüfstatus unterschiedene Kategorien unterteilt (Tab. 8). Der Prüfstatus einer Sorte kennzeichnet ihre aktuell erreichte Stufe im Verlauf des Prüfsystems (Abb. 4). Er steht in Bezug zum Faktor ‚Versuchstyp' (Tab. 11), in dem diese Sorte aktuell geprüft wird und zur Dauer der Prüfung dieser Sorte (Faktor ‚Jahr'). Die Sorten des Kernsortimentes innerhalb einer Prüfstatus-Kategorie weisen eine weitgehend vergleichbare Prüfdauer und Datenstruktur auf.

Die sonstigen 1074 Sorten stehen noch nicht oder nicht mehr im LSV. Diese Sorten verursachen eine insgesamt starke Unbalanciertheit, beeinträchtigen die Güte der Mittelwertschätzungen im Kernsortiment nicht, sondern fördern diese durch ihren Beitrag zur Schätzung der Varianzkomponenten und des Transformationsparameters sowie zur Adjustierung (durch zusätzliche ‚Brücken', indirekte Vergleiche etc.).

Tab. 8: Gruppierung der Winterweizensorten nach Prüfstatus*

Beschreibung Prüfstatus* der Sorten (Definition LSV, OV, WP s.u. unter ‚Versuchstyp')	**Kürzel Prüfstatus**	**Anzahl Sorten**
Kernsortiment	∑	**44**
davon:		
empfohlen, langjährig geprüft, nicht mehr im aktuellen LSV	alte Sorten	7
seit fünf bis acht Jahren im LSV	LSV5-8	8
seit vier Jahren im LSV	LSV4	2
seit drei Jahren im LSV	LSV3	3
seit zwei Jahren im LSV	LSV2	6
erstes Jahr im LSV	LSV1	4
im OV, Zulassungskandidat aus WP	OV_WP	12
im OV, übernommen aus EUSV	OV_EU	2
sonstige Sorten	∑	**1074**

* Der Prüfstatus einer Sorte kennzeichnet ihre aktuell erreichte Stufe im Verlauf des Prüfsystems (Abb. 4). Er steht in Bezug zum Faktor ‚Versuchstyp' (Tab. 11), in dem diese Sorte aktuell geprüft wird und zur Dauer der Prüfung dieser Sorte (Faktor ‚Jahr').

WP Wertprüfung
LSV Landessortenversuch (alle Sorten der Prüfjahre 1 bis x in gemeinsamem Versuch)
EUSV EU-Sortenversuch
OV Orientierungsversuch

Auf die im Vergleich zu zeitlich begrenzten Versuchsserien hier außergewöhnlich stark ausgeprägte Unbalanciertheit hinsichtlich des Faktors Sorte in zeitlicher und räumlicher Hinsicht wird aufgrund der Bedeutung für Auswertung und Interpretation gesondert unter 4.4 eingegangen.

Eine Besonderheit stellen so genannte Verrechnungssorten (VRS) und Vergleichssorten (VGL) dar. Dies sind Sorten, die zwischen Bundessortenamt und den Länderdienststellen für das Sortenwesen jährlich beraten und vereinbart werden. VRS sollen im jeweiligen Jahr Versuchstyp übergreifend in allen Sortenversuchen stehen, VGL in allen Wertprüfungen, wobei VGL potenzielle Kandidaten für die spätere Aufnahme als VRS oder ausgeschiedene VRS sind. VRS und VGL werden von Jahr zu Jahr nur in kleinem Umfang, also gleitend, gewechselt. Die Präsenz der VRS und VGL leistet den wesentlichen Beitrag bei der Auswertung sehr unbalancierter Daten, indem sie zwischen Orten und Jahren, besonders aber zwischen den zeitlich versetzten Versuchstypen ‚Brücken' bilden. (Tab. 10; VRS und VGL markiert).

Sortengruppe (Gr):

Bei einigen Pflanzenarten wurden Sortengruppen gebildet, wobei ‚Sorte' dann vollständig in Gruppe geschachtelt ist (es existiert keine Sorte, die in zwei Gruppen vorkommt). Diese Gruppierung erfolgte immer dann, wenn Sorten, welche zu unterschiedlichen diskreten Kategorien gehören, zwar in Versuchen gemeinsam geprüft wurden, aber in der Interpretation separat innerhalb dieser Sortengruppen betrachtet werden. Grund dafür ist i.d.R., dass z.B. der Ertragsvergleich von Sorten unterschiedlicher Gruppen keine unmittelbare Aussagekraft besitzt, sondern z.B. erst

durch ökonomische Kalkulationen unter Einbeziehung von Saatgutpreis (Hybriden gegen Nicht-Hybriden) oder Erlös (Weizen-Qualitätsgruppen) etc. sinnvoll wird. Die Zuordnung zu Sortengruppen erfolgt bundesweit einheitlich durch das Bundessortenamt im Rahmen der Stammdatenpflege der PIAF-Software. Folgende Gruppierungsprinzipien bei den Pflanzenarten werden genutzt: Qualitätsgruppe bei Winter- und Sommerweizen, Zeiligkeit der Ähre bei Winter- und Sommergerste, Hybrid- versus Populations- oder Liniensorte bei Roggen, Raps, Winterweizen und Wintergerste, Wuchstyp bei Lupine, Reifezahl bei Silo- und Körnermais sowie Reifegruppe bei Kartoffeln. Im Maximum führt dies beim Silomais zu 11 Gruppen. Bei Erbsen, Hafer und Triticale erfolgte keine Gruppierung (Tab. 6).

Eine offiziell definierte Gruppenzuordnung liegt für nicht zugelassene Sorten noch nicht vor. Somit können diese Sorten bei der Varianzkomponentenschätzung nicht einfließen. Bei der Mittelwertschätzung können aber alle Sorten wieder einbezogen werden.

Für das Kernsortiment Winterweizen gibt Tab. 9 Informationen zur Zuordnung zur jeweiligen Sortengruppe sowie rein sachliche Informationen zum Prüfstatus, zur Zulassungssituation, zur Empfehlung durch die Landesforschungsanstalt in Mecklenburg-Vorpommern, zur Verwendung als Teil der Bezugsbasis bei der Darstellung von Relativzahlen in Mecklenburg-Vorpommern und zum Status als Verrechnungs- oder Vergleichssorte beim Bundessortenamt. Diese Informationen entsprechen dem Stand Ende 2009, also für das letzte Jahr der hier erfolgten Serienauswertungen. Nachträglich ergaben sich naturgemäß jährlich für die Auswertungen in dieser Arbeit nicht mehr relevante Änderungen.

In Tab. 10 ist für das Kernsortiment die Anzahl einbezogener Versuchsergebnisse je Jahr im Auswertungszeitraum 2001-2009 dargestellt. Sorten innerhalb eines Prüfstatus‘ sind durch gepunktete Rahmenlinien abgegrenzt, von Prüfstatus zu Prüfstatus durch durchgezogene Rahmenlinien. Zu ersehen ist für jede Sorte die sehr unterschiedliche Anzahl Versuche je Jahr, die im unmittelbaren Kontext zu dem Versuchstyp (Tab. 11, s. 4.3.2) steht. Von Prüfstatus zu Prüfstatus ist der zeitliche Versatz des Prüfungseinstieges und im Kontext zu Tab. 11 des Durchlaufens der Versuchstypen (s. 4.3.2) zu erkennen. Des Weiteren kann man die untypisch hohe Zahl an Versuchsergebnissen in den Jahren erkennen, in denen eine Sorte Verrechnungs- und Vergleichssorten war, insbesondere beim Vergleich zu Sorten des gleichen Prüfstatus‘.

Tab. 11 zeigt, in welchen Jahren die Sorten in welchem *Versuchstyp* geprüft wurden - die Erläuterung dieses Faktors erfolgt in 4.3.2. Aufgrund der o.g. Bezüge zu Tab. 10 bei gleicher Tabellenstruktur werden beide Tabellen hier gemeinsam eingeordnet.

Tab. 9: Kernsortiment Winterweizen - Statusinformationen im Anbaujahr 2008/09

Sorten-name	Gr	Prüfstatus (Definition s. Tab. 8)	Zulassung	Empfeh-lung LFA 2008/09	Bezugs-basis LFA 2009	VRS / VGL 2009
Akratos	A	alte Sorten	Juni 2004	beschränkt		
Boomer	A	alte Sorten	März 2005	auslaufend		
Impression	A	alte Sorten	März 2005	auslaufend		
Magister	E	alte Sorten	März 2005	auslaufend		
Schamane	A	alte Sorten	März 2005	auslaufend		
Skagen	E	alte Sorten	März 2006	auslaufend		
Skalmeje	C	alte Sorten	Juli 2006	auslaufend		
Cubus	A	LSV5-8	März 2002	voll		
Tommi	A	LSV5-8	März 2002	voll		VRS
Dekan	B	LSV5-8	April 1999	voll		
Akteur	E	LSV5-8	März 2003	voll	x	
Toras	A	LSV5-8	März 2004	voll		
Türkis	A	LSV5-8	März 2004	voll		VRS
Brilliant	A	LSV5-8	März 2005	voll	x	
Hermann	C	LSV5-8	März 2004	voll	x	
Mulan	B	LSV4	März 2006	voll	x	
Potenzial	A	LSV4	März 2006	voll	x	
Discus	A	LSV3	März 2007	voll	x	
Esket	A	LSV3	März 2007	vorläufig		VGL
Jenga	A	LSV3	März 2007	vorläufig		
Adler	E	LSV2	Juli 2008			
JB Asano	A	LSV2	April 2008	vorläufig	x	VGL
Julius	B	LSV2	April 2008		x	VRS
Pamier	A	LSV2	April 2008			
Sophytra	B	LSV2	April 2008			
Tabasco	Ck	LSV2	April 2008		x	VGL
Event	E	LSV1	März 2009			
Global	B	LSV1	Juli 2009			
Kredo	B	LSV1	März 2009			
Primus	B	LSV1	September 2009			
Agil		OV_WP	nicht zugelassen			
Elbar		OV_WP	nicht zugelassen			
Expo		OV_WP	nicht zugelassen			
KWS Adlon		OV_WP	nicht zugelassen			
Leo		OV_WP	nicht zugelassen			
Nathan		OV_WP	nicht zugelassen			
Ohlin		OV_WP	nicht zugelassen			
Quintus		OV_WP	nicht zugelassen			
Rufus		OV_WP	nicht zugelassen			
Smaragd		OV_WP	EU-Sorte *			
Tarkus	B	OV_WP	März 2009			
Zappa	C	OV_WP	März 2009			
Mercato	B	OV_EU	EU-Sorte *			
Philipp	E	OV_EU	EU-Sorte *			

* Sorte mit Zulassung in einen EU-Mitgliedsland, für die in Deutschland Saatgut gemäß §55 Abs. 2 SaatG (Rutz, 2006) zugelassen werden kann

VRS - Verrechnungssorte; VGL - Vergleichssorte

Tab. 10: Kernsortiment Winterweizen - Anzahl Versuche je Sorte im Auswertungszeitraum 2001-2009 (N=9) bzw. 2004-2009 (N=6)

Sortenname	Versuchsjahr 2001 - 09									Summe *	
	01	02	03	04	05	06	07	08	09	2001-09	2004-09
Akratos	2	2	4	6	23	22	7	7	8	81	73
Boomer		2	2	4	3	21	19	18	3	72	68
Impression		2	2	4	11	21	15	12		67	63
Magister		2	2	4	3	7	4	4		26	22
Schamane		2	2	4	6	21	16	8		59	55
Skagen			3	3	4	4	12	13	2	41	38
Skalmeje			3	3	4	11	17	18	3	59	56
Cubus	3	5	12	21	21	16	12	9	11	110	90
Tommi	3	7	36	37	38	37	35	25	25	243	197
Dekan	8	7	22	23	16	15	14	8	10	123	86
Akteur	2	2	4	10	13	15	11	14	15	86	78
Toras	2	3	2	3	21	21	17	16	11	96	89
Türkis	3		4	10	32	32	33	30	28	172	165
Brilliant		2	2	4	11	18	16	17	15	85	81
Hermann	2	3	4	15	31	31	26	29	23	164	155
Mulan			3	3	4	11	19	21	26	87	84
Potenzial			3	3	4	11	15	15	18	69	66
Discus				3	3	5	17	20	18	66	66
Esket				3	3	5	16	28	19	74	74
Jenga				3	3	5	17	18	19	65	65
Adler					2	3	5	9	7	26	26
JB Asano					2	3	5	18	33	61	61
Julius					2	3	5	20	38	68	68
Pamier					2	3	5	7	14	31	31
Sophytra					2	3	5	4	10	24	24
Tabasco					2	3	5	18	33	61	61
Event						2	2	4	9	17	17
Global						2	2	4	20	28	28
Kredo						2	2	4	14	22	22
Primus						2	2	4	7	15	15
Agil						2	2	4	2	10	10
Elbar						2	2	4	2	10	10
Expo						2	2	4	2	10	10
KWS Adlon						2	2	4	2	10	10
Leo						2	2	4	2	10	10
Nathan						2	2	4	2	10	10
Ohlin						2	2	4	2	10	10
Quintus						2	2	4	2	10	10
Rufus						2	2	4	2	10	10
Smaragd						2	2	4	8	16	16
Tarkus						2	2	4	9	17	17
Zappa						2	2	4	8	16	16
Mercato							3	3	3	9	9
Philipp							3	3	4	10	10

VGL Vergleichssorte VRS Verrechnungssorte

* 2001-09 für Schätzung von Transformationsparameter und Varianzen, 2004-09 für Mittelwerte

Tab. 11: Kernsortiment Winterweizen - Versuchstyp je Sorte und Jahr

Sorten-name	Versuchsjahr 2001 - 09								
	01	02	03	04	05	06	07	08	09
Akratos	WP1	WP1	WP3	LSV	LSV	LSV	LSV	LSV	(LSV)
Boomer		WP1	WP2	WP3	OV	LSV	LSV	LSV	(LSV)
Impression		WP1	WP2	WP3	OV	LSV	LSV	LSV	
Magister		WP1	WP2	WP3	OV	LSV	LSV	LSV	
Schamane		WP1	WP2	WP3	OV	LSV	LSV	LSV	
Skagen			WP1	WP2	WP3	LSV	LSV	LSV	(LSV)
Skalmeje			WP1	WP2	WP3	LSV	LSV	LSV	(LSV)
Cubus	WP3	OV	LSV	LSV	LSV	LSV	LSV	LSV	LSV
Tommi	WP3	OV	LSV	LSV	LSV	LSV	LSV	LSV	LSV
Dekan	LSV	LSV	LSV	LSV	LSV	LSV	LSV	LSV	LSV
Akteur	WP1	WP1	LSV	LSV	LSV	LSV	LSV	LSV	LSV
Toras	WP1	WP2	WP2	LSV	LSV	LSV	LSV	LSV	LSV
Türkis	WP2		WP3	LSV	LSV	LSV	LSV	LSV	LSV
Brilliant		WP1	WP2	WP3	OV	LSV	LSV	LSV	LSV
Hermann	WP1	WP2	WP3	LSV	LSV	LSV	LSV	LSV	LSV
Mulan			WP1	WP2	WP3	LSV	LSV	LSV	LSV
Potenzial			WP1	WP2	WP3	LSV	LSV	LSV	LSV
Discus				WP1	WP2	WP3	LSV	LSV	LSV
Esket				WP1	WP2	WP3	LSV	LSV	LSV
Jenga				WP1	WP2	WP3	LSV	LSV	LSV
Adler					WP1	WP2	WP3	LSV	LSV
JB Asano					WP1	WP2	WP3	LSV	LSV
Julius					WP1	WP2	WP3	LSV	LSV
Pamier					WP1	WP2	WP3	OV	LSV
Sophytra					WP1	WP2	WP3	OV	LSV
Tabasco					WP1	WP2	WP3	LSV	LSV
Event						WP1	WP2	WP3	LSV
Global						WP1	WP2	WP3	LSV
Kredo						WP1	WP2	WP3	LSV
Primus						WP1	WP2	WP3	LSV
Agil						WP1	WP2	WP3	OV
Elbar						WP1	WP2	WP3	OV
Expo						WP1	WP2	WP3	OV
KWS Adlon						WP1	WP2	WP3	OV
Leo						WP1	WP2	WP3	OV
Nathan						WP1	WP2	WP3	OV
Ohlin						WP1	WP2	WP3	OV
Quintus						WP1	WP2	WP3	OV
Rufus						WP1	WP2	WP3	OV
Smaragd						WP1	WP2	WP3	OV
Tarkus						WP1	WP2	WP3	OV
Zappa						WP1	WP2	WP3	OV
Mercato							EUSV1	EUSV2	OV
Philipp							EUSV1	EUSV2	OV

* Faktor ‚Versuchstyp' definiert in 4.3.2

(LSV) nicht im Zielgebiet, nur noch Daten aus Nachbar-Anbaugebieten

Intensität (I):

Die Einzelversuche der Getreidearten waren überwiegend als zweifaktorielle Spaltanlagen mit dem Großteilstücksfaktor *Intensität* und dem Kleinteilstücksfaktor *Sorte* angelegt. Nur beim Hafer wurde erst nach Abschluss der hier erfolgten Auswertungen von einem einfaktoriellen auf einen zweifaktoriellen Sortenversuch analog zu allen anderen sonstigen Getreidearten umgestellt.

Der Faktor Intensität beinhaltete in den Einzelversuchen überwiegend die zwei Intensitätsstufen:

Stufe II (ortsüblich; mit Fungiziden und Wachstumsreglern),

Stufe I (reduziert; ohne Fungizide und Wachstumsregler).

Bei der Merkmalskategorie *Ertrag* erfolgt sowohl in der derzeitigen praktischen Anwendung wie auch in dieser Arbeit eine Serienauswertung für die Stufe II als Praxis relevante (ortsübliche) Variante (Sortenvergleiche auf gleicher Intensitätsstufe II). Die Stufe I wurde nicht separat verrechnet, da der Informationsgehalt von Sortenvergleichen auf Intensitätsstufe I eher im Bereich der Bonitur der Krankheiten und der Standfestigkeit liegt.

Eine Umsetzung des Verfahrens mit beiden Einzelversuchsfaktoren *Sorte* und *Intensität* für das Gesamtmodell der Serienanalyse erfolgt nicht. Die dafür erforderlichen komplexen Serienauswertungsmodelle (Abschnitt 5.3.2) wären rechentechnisch nicht realisierbar - Testläufe an realen Datensätzen zeigten, dass bei Anwendung der in 5.3 beschriebenen komplexen Modellansätze ein erfolgreicher Programmdurchlauf an der Rechnerleistung scheiterte bzw. zu unakzeptablen Rechenzeiten führte (mehrtägig für einen Durchlauf). Stattdessen wurde der sortenspezifische Minderertrag der Stufe I gegenüber der Stufe II (Differenz Stufe II – Stufe I) quasi einfaktoriell ausgewertet (als Wechselwirkungseffekt Sorte $\times$ Intensität auf gleicher Stufe des Faktors Sorte). Bei dieser Differenzbildung fallen alle nicht mit der Intensität gekreuzten Effekte weg, so dass nur die Interaktionen mit der Intensität betrachtet werden (Möhring et al., 2004b). Der Minderertrag der geringen Intensität gibt praxisrelevante sortenspezifische Hinweise zur Ertragswirksamkeit des Komplexes Resistenz und Standfestigkeit. Ihre Auswertung erbringt eine ebenso geeignete sortenspezifische Aussage für diesen Wechselwirkungseffekt wie eine komplexe Versuchsserienauswertung mit beiden Einzelversuchsfaktoren, so dass sich letztere erübrigt.

Beim Raps ist die Definition der zwei Intensitätsstufen in den regionalen Landessortenversuchen analog zum Getreide. Wertprüfungen werden aber nur einfaktoriell in einer ohne Fungizide und Wachstumsregler definierten Intensität angelegt und entsprechen damit der LSV-Stufe I. Es entsprachen aber bei einigen Landes-, Bundes- und EU-Sortenversuchen nicht alle einfaktoriellen Versuche exakt der Definition der Stufe I. Nur 1972 von 6314 Einzeldaten stammen aus einem zweifaktoriellen Versuch. Ursache für diesen geringen Anteil ist die hohe Sortenanzahl in den einfaktoriellen Wertprüfungen. Die zweifaktorielle Struktur wurde daher ausschließlich für die Bestimmung des Minderertrages herangezogen. Die Auswertung für die weiteren Merkmalskategorien erfolgte einfaktoriell in der Form, dass je Versuch die Mittelwerte je Sorte in die Versuchsserienauswertung einfließen, unabhängig davon ob der Versuch einfaktoriell war und welcher Intensitätsstufe dieser dann entsprach oder ob die Mittelwerte aus zwei Intensitätsstufen gebildet wurden. Damit enthalten alle sortenbezogenen Varianzkomponenten auch nicht mehr quantifizierbaren intensitätsbezogenen Anteil, wodurch sich andererseits der Aussage-

bereich aber auch erweitert. Allerdings sind die Niveauunterschiede zwischen den Stufen (s. Abb. 1) und auch die Wechselwirkungen beim Raps erheblich geringer als beim Getreide, es gibt bislang selten signifikante Sorteneffekte beim Minderertrag.

Anmerkung zu Blockungsfaktoren und Wiederholungen im Einzelversuch

In die Auswertungsmodelle für die Einzelversuche flossen unterschiedliche Formen der Blockbildung ein. Bei der hier vorgestellten Versuchsseriensauwertung wird auf Mittelwerte aus den Einzelversuchen und ggf. deren Standardfehler zugegriffen. Diese Blockungsfaktoren gehen in die Modellbildung der *Serienanalyse* also nicht explizit, aber bei Gewichtung der Sortenmittelwerte mittels der Kehrwerte ihrer Fehlervarianzen (s. 5.3.3) implizit ein. Aufgrund der Vielzahl unterschiedlicher Blockbildungen in den Einzelversuchen wird an dieser Stelle darauf verzichtet, die Versuchsanlagen detailliert zu beschreiben, es wird nachfolgend aber auf die häufigsten Versuchsanlagen und Formen der Blockbildung hingewiesen.

Die Blockbildung wird hier also nur der Vollständigkeit halber angesprochen. Die Anzahl Blockungsfaktoren war nicht einheitlich. Es wurden sowohl vollständige wie auch unvollständige Blocks in Planung und Auswertung verwendet.

Versuche aus benachbarten Bundesländern, Wertprüfungen und EU-Sortenversuche waren fast ausschließlich *einfaktorielle randomisierte Blockanlagen* oder *zweifaktorielle Spaltanlagen* (Intensität als Groß-, Sorte als Kleinteilstücksfaktor) mit vollständigen Blocks, wie sie u.a. durch Thomas (2006) beschrieben sind. Einzelne Wertprüfungen beim Mais und Sommergerste waren auch als *Alpha-Gitter* angelegt. In den Landessortenversuchen der LFA in Mecklenburg-Vorpommern wurden zunehmend auch Zeilen-Spalten-Anlagen mit der Option zusätzlicher Blockbildung durch Lateinisierung nach Williams (2007) eingeführt. Hierbei werden außer ‚Zeilen' und ‚Spalten' ggf. zusätzliche Blockungsfaktoren, bei denen sich ein Block jeweils über mehr als eine Spalte und/oder Zeile erstreckt, in Planung und Auswertung berücksichtigt.

Die Versuche wurden bei zweifaktoriellen Anlagen überwiegend mit zwei bzw. bei einfaktoriellen Anlagen mit drei oder vier Wiederholungen angelegt. In Einzelfällen traten bis zu sechs Wiederholungen auf. Am Versuchsstandort Gülzow wurden in zweifaktoriellen Anlagen häufig zwei Wiederholungen in Intensitätsstufe I und drei Wiederholungen in Intensitätsstufe II angelegt.

4.3.2 Zusätzliche Faktoren bzw. Klassifizierungsgrößen der Serien / Umwelten

Jahr (J):

Im Regelfall wurden zur Schätzung des Transformationsparameters (6.1) und der Varianzkomponenten (6.2) neun Jahre herangezogen. Aufgrund des hier verwendeten Stichtages 31.12.2009 für die Datensatzbildung waren dies die Erntejahre 2001-09 oder bei einzelnen Merkmalen 2000-08. Lediglich für einen Ansatz zur Methoden-Validierung im Abschnitt 6.5.3 wurden nachträglich gewonnene Daten des Jahres 2010 für den Kornertrag bei Winterweizen noch herangezogen.

Aufgrund eines grundlegenden Wechsels im Prüfsystem wurden bei Kartoffeln nur sechs Jahre, aufgrund einer grundlegend neuen Qualitätsbestimmung beim Silomais nur acht Jahre und aufgrund von Problemen der Datenverfügbarkeit bei Sommerbraugerste-Qualitätsmerkmalen ebenfalls nur acht Jahre einbezogen. Der Mindererertrag beim Raps wurde aufgrund der im Vergleich zu den anderen Merkmalen sehr schwachen Daten-Besetzung und der vergleichsweise großen Schätzfehler der Varianzkomponenten zehnjährig verrechnet.

Zur Schätzung der Sorten-Mittelwerte bzw. Sorteneffekte wurden nur sechs Jahre einbezogen. Einerseits war diese Begrenzung aus Gründen der Rechenkapazität erforderlich, denn der erfolgreiche Durchlauf ohne ‚*Absturz'* der Routine war häufig erst nach erheblicher Reduktion des Datenbestandes möglich. Andererseits war die Begrenzung des Datenbestandes am unschädlichsten (im Sinne der präzisen Mittelwertschätzung für das Kernsortiment) durch Begrenzung der Jahre möglich, da für das Kernsortiment und hier speziell für die besonders interessierenden jungen Sorten nur ein sehr geringer Anteil zeitlich noch älterer Daten vorlag (Tab. 10, 11, 14).

Anbaugebiet (AG) und Boden-Klima-Raum (BKR)

Die Herleitung der Anbaugebiete (AG) und Boden-Klima-Räume (BKR) ist gesondert im Abschnitt 5.1 dargelegt.

Ort (O):

Der Standort-Region-Bezug lässt sich als zweistufige stratifizierte Stichprobe beschreiben, wobei jeder Ort eindeutig einem Boden-Klima-Raum (BKR) zugeordnet ist und jeder BKR wiederum eindeutig einem Anbaugebiet (AG) zugeordnet ist. In der Regel lagen mehrere Orte je BKR vor. Die Orte werden hier als Stichprobe des jeweiligen AG betrachtet.

Damit eine unverzerrte Schätzung erhalten werden kann, soll die Stichprobe repräsentativ sein. Mit gewissen Vorbehalten kann das Versuchsstellennetz im einbezogenen Gebiet als repräsentativ angenommen werden, eine tatsächliche jährliche Zufallsauswahl der Versuchskoordinaten ist allerdings faktisch nicht realisierbar.

Ob Orte in die Auswertung einbezogen werden, ist unabhängig davon, ob es sich um Offizial-Prüfstandorte der Länderdienststellen, des Bundessortenamtes, der Züchter oder anderer Versuchsansteller handelt. Die Bandbreite der Anzahl verschiedener Orte über den gesamten Prüfzeitraum und über alle Versuchstypen hinweg geht von nur sechs Orten beim Sommerweizen (aus fünf BKR, vier Ländern) bis zu 39 Orten (aus 11 BKR und fünf Ländern) beim Winterweizen.

In Tab. 12 ist für die Ertragsauswertung von Winterweizen die Belegung der Jahre mit Versuchen in den räumlichen Strukturen (Faktoren Anbaugebiet und Ort, weitere Klassifikationsfaktoren: Boden-Klima-Raum und Bundesland) dargestellt.

Tab. 12: Räumliche und zeitliche Datensatzstruktur; Kornertrag Winterweizen

räumliche Klassifizierungen				Versuchsjahr 2001 - 09								
Anbau-gebiet	Boden-Klima-Raum	Bundes-Land	Ort	01	02	03	04	05	06	07	08	09
				Anzahl Versuche je O × J								
3	154	SH	Futterkamp	1	2	4	4	4	4	4	4	1
			Kastorf (HL)	.	.	.	.	.	.	.	1	1
			HL-Niendorf	1	.	1	1	1	1	1	.	.
			Wolkenwehe	.	.	.	1	1	1	1	.	.
	157		Birkenmoor	.	.	1	1	1	1	1	1	.
			Harzhof	.	.	.	.	.	1	1	.	1
			Hohenlieth	.	.	.	.	1	.	.	.	.
			Loit	.	.	.	.	.	.	.	.	2
			Schuby/Tolk	1	.	1	1	1	1	.	.	.
4	101	MV	Boldebuck	.	1	1	1	1	.	1	1	.
			Groß Kiesow	1	1	1	1	1	1	1	1	1
			Gülzow	2	2	1	.	2	2	2	2	2
			Tützpatz	4	4	3	3	3	3	3	3	2
		BB	Prenzlau	3	3	3	3	3	3	2	2	1
	158	MV	Biestow	1	1	1	1	1	1	1	1	1
			Granskevitz	2	2	2	1	.	1	1	1	.
			Köchelstorf	1	1	1	1	1	1	1	1	1
5	104	BB	Dürrenhofe	1	1	1	1	1	.	.	1	.
			Güterfelde	1	1	1	1	1	1	1	1	1
			Nuhnen	1	1	1	1	1	1	1	1	1
			Thyrow	.	.	1	1	1	1	1	.	1
45	102	MV	Gülzow G	.	.	.	2	3	1	1	1	1
			Malchow/MÜR	.	.	.	.	.	.	.	.	1
			Neuhof	2	2	1	2	2	2	.	2	.
			Vipperow	1	1	1	2	3	2	1	1	2
		BB	Badingen	2	4	3	3	3	2	3	2	.
			Berge	1	1	1	1	1	1	1	.	1
			Pessin	.	.	.	.	.	.	.	.	1
		ST	Beetzendorf	1	1	1	1	1	2	2	2	1
	109	NI	Lüchow-Clenze	.	.	1	1	1	1	1	1	.
8	146		Groß Malchau	.	.	.	1	1	.	.	.	.
			Rieste	.	.	1	1	.	.	.	.	.
			Ubbendorf	.	.	.	.	.	.	1	.	1
	151		Holtorfsloh	.	.	1	1	1	1	1	1	1
			Ohrensen	.	.	.	.	.	.	1	1	1
			Sittensen	.	.	1	1	1	1	.	1	.
	153	SH	Gudendorf	.	.	1	1	1	1	.	.	.
			Süderhastedt	.	.	.	.	.	.	.	1	1
	156		Schuby	.	.	1	1	1	1	.	1	1

BB - Brandenburg; MV - Mecklenburg-Vorpommern; NI - Niedersachsen; SH - Schleswig-Holstein; ST - Sachsen-Anhalt; Schlüssel für Anbaugebiet s. Abb. 6; Schlüssel für Boden-Klima-Raum s. Abb. 7

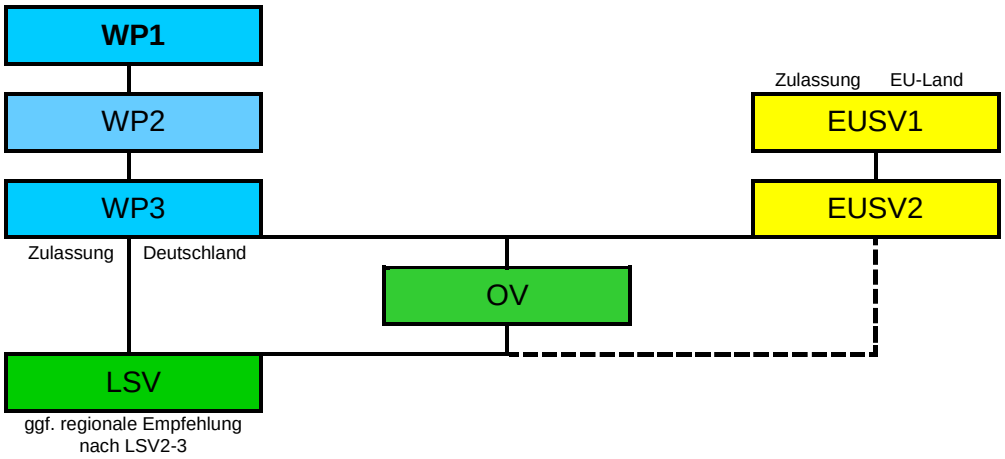
Versuchstyp (T) und das Sortenprüfsystem

Die Einbeziehung des Faktors *Versuchstyp* war Voraussetzung für die eindeutige Klassifizierung von unterschiedlichen Einzelversuchen an gleicher Jahr × Ort - Kombination. Der zentrale und häufigste Versuchstyp ist der Landessortenversuch (LSV), der mit dem Ziel der regionalen Sortenberatung für Anbaugebiete geplant und angelegt wird. Im LSV werden die Jahrgänge (LSV1 - LSVx) nicht in verschiedene Einzelversuche getrennt, sondern gemeinsam in einem Sortiment (LSV) geprüft. Neben LSV wurden auch alle anderen verfügbaren Offizialprüfungen herangezogen, soweit sie im Zielgebiet oder in einbezogenen Nachbargebieten standen. In diesem Fall waren es insbesondere Wertprüfungen (WP) des Bundessortenamtes im ersten bis dritten Prüfjahr (WP1 bis WP3), Orientierungsversuche (OV) im Übergang zwischen WP und LSV sowie EU-Sortenversuche (EUSV) im ersten und zweiten Prüfjahr (EUSV1 und EUSV2).

Nachfolgend wird in Abb. 4 der Prüfablauf im integrierten Prüfsystem nur am Beispiel Winterweizen erläutert (Stand 2009, MV). Im Falle der Sortenzulassung in Deutschland durch das Bundessortenamt durchläuft eine Sorte drei Jahre die WP. Bei Winterungen ist zum Zeitpunkt der Aussaat des LSV noch nicht über die Zulassung von WP3-Stämmen entschieden. Daher nehmen die Länder nur die ihres Erachtens aussichtsreichsten Kandidaten unmittelbar in den LSV auf und prüfen weitere Sorten ggf. im OV an einer meist eingeschränkten Standortanzahl. Sorten, die in einem anderen EU-Land zugelassen sind, können auf Antrag der Züchter/ Vertreiber die bundesweit organisierten EU-Sortenversuch (EUSV) durchlaufen. Diese Prüfung ist nur zweijährig, da die Sorten bereits eine Offizialprüfung durchlaufen haben. Nach erfolgreichem Durchlauf können Sorten aus dem EUSV in den OV oder direkt in den LSV aufsteigen.

Dieser Prüfablauf ist auch sortenspezifisch für das Kernsortiment Winterweizen in Tab. 11 dargestellt. Deutlich erkennbar ist der zeitliche Versatz zwischen Prüfstatus (Definition s. Tab. 8) und Versuchstyp. Es zeigen sich Ausnahmen wie die Sorte Toras mit zweifacher WP2 ohne WP3, Akratos mit zweifacher WP1 ohne WP2, Türkis mit einem Aussetzungsjahr zwischen WP2 und WP3 oder Akteur mit drei Jahren WP1 (2000, 2001, 2002; 2000 nicht in Tab. 11 aufgenommen).

In den Auswertungen wird ‚Versuchstyp' nicht als unabhängiger Faktor verwendet, da unterschiedliche Versuchstypen prinzipiell in gleicher Weise nach gleichen Richtlinien betreut werden - oft vom gleichen Versuchstechniker am gleichen Ort. Daher sind a priori keine Wechselwirkungen zwischen T und O × J bzw. deren Wechselwirkungen mit Sorte etc. zu erwarten. Der Faktor T wird daher im Auswertungsmodell lediglich als Faktor zur Stratifizierung innerhalb von O × J - Kombinationen benötigt, sofern an einer O × J - Kombination mindestens zwei eigenständige Versuche standen. Der in den Auswertungsmodellen (s. 5.3) zu berücksichtigende Effekt T(O•J) ist als zufällig definiert. In Tab. 12 wurden an denjenigen O × J - Kombinationen mehrere Typen parallel geprüft, an denen mehr als ein Versuch ausgewiesen ist.



Zuständigkeit:	vertreten durch:
Länder	Dienststellen der Bundesländer (hier LFA)
Bund	Bundessortenamt
Züchter	Bundesverband Deutscher Pflanzenzüchter

WP Wertprüfung (1, 2, 3 für 1., 2. und 3. Prüfjahr, jeweils als eigener Versuch angelegt)
LSV Landessortenversuch (alle Sorten der Prüfjahre 1 bis x in gemeinsamem Versuch)
EUSV EU-Sortenversuch (1, 2 für 1. und 2. Prüfjahr, in gemeinsamem Versuch)
OV Orientierungsversuch

Abb. 4: Regelprüfablauf im integrierten Sortenprüfsystem am Beispiel Wintergetreide; Stand 2009, MV

In Tab. 6 ist in Spalte ‚T' für Versuchstyp nicht die Anzahl der im gesamten Datensatz enthaltenen sachlich unterschiedenen Versuchstypen aufgeführt, sondern die Anzahl Versuchstypen die im Maximum an einer O × J - Kombination gestanden hat. Bei Mais waren dies bis zu 10 unterschiedliche Versuchstypen, also unterschiedliche Einzelversuche an einer O × J - Kombination. Ursache ist, dass hier außer den in Abb. 4 dargestellten Typen auch Reifegruppen zur Versuchs-Klassifikation herangezogen werden mussten. Dagegen gab es z.B. bei Sommerfuttergerste nur den einen Versuchstyp 'LSV' (in WP wird nur die Nutzungsrichtung Sommer*brau*gerste geprüft; die EUSV Sommerfuttergerste wurden grundsätzlich in LSV integriert und tauchen damit nicht als eigenständiger Versuchstyp auf). Die Integration von WP mit LSV und EUSV mit LSV in einen methodisch zusammengehörigen Versuch wurde überall dort praktiziert, wo die kombinierten Sortimentsumfänge eine nicht zu große Versuchsfläche forderten.

Für die Einbeziehung unterschiedlicher Versuchstypen an O × J - Kombination sind VRS / VGL (s. 3.3.1, ad. Sorte) von herausragender Bedeutung, da sie oft den orthogonalen Kern zwischen diesen Einzelversuchen ohne jegliche weitere Sortenüberschneidungen darstellen. Auch in der Gesamtserie bilden VRS / VGL Brücken zwischen Sorten unterschiedlichen Prüfstatus' und ihrer zeitlich versetzten Zuordnung zu Versuchstypen.

Bundesland bzw. Land (L)

Anbaugebiete bzw. BKR und Bundesländer (L) sind nicht deckungsgleich. Länder spiegeln einen politischen Regionalisierungsansatz wieder, die im Zusammenhang mit der Sortenprüfung bezüglich des Standortbezuges eigentlich irrelevant ist und durch standortkundlich definierte BKR und Anbaugebiete ersetzt wurde (s. 5.1). Da die Organisation von LSV aber durch den Bundesländern zugeordnete Dienststellen erfolgt, ist die Zuordnung der Orte zu Ländern für die Problematik der Datengewinnung, der Sortimentsabstimmung und des Datenaustausches doch von logistischem Interesse. Insofern ist in Tab. 12 in der Zuordnung der Orte zu Ländern auch kein eindeutiger Zusammenhang zur Zuordnung zu BKR und Anbaugebieten zu ersehen.

In der Regel flossen Daten außer aus Mecklenburg-Vorpommern von den Ländern Schleswig-Holstein, Niedersachsen, Brandenburg und Sachsen-Anhalt ein (Tab. 6, 12). Zur Absicherung eines Mindestmaßes an Versuchsdaten (für ein Anbaugebiet!) wurden bei Lupinen und Körnermais zusätzlich Diluvial-Standorte aus dem Norden Sachsens einbezogen, also von insgesamt sechs Ländern (Tab. 6).

4.4 Dem Sortenprüfsystem immanente systematische Unbalanciertheit

In der traditionellen Auswertungspraxis vieler Länderdienststellen wurden bzw. werden vielfach noch ausschließlich orthogonale Kernsortimente der Landessortenversuche ausgewertet. Hierfür waren das Konzept der meist 2- und/oder 3-jährigen Verrechnung innerhalb orthogonaler Kerne und eine Regelprüfzeit erforderlich. Diese zeitliche Begrenzung und orthogonale Datenstrukturen entsprechen allerdings nicht der Natur des integrierten Sortenprüfsystems (LSV + WP + EUSV). Sortenversuche stellen ein längerfristig zeitlich nicht begrenztes Kontinuum dar, in dem jährlich Sorten aufgenommen und auch herausgenommen werden. Dabei ist auch die Prüfdauer der Sorten hochgradig unterschiedlich, beginnend mit der WP1 durchlaufen die Sorten einen ununterbrochenen Selektionsprozess. Insofern sind orthogonale Strukturen und eine Regelprüfzeit Artefakte.

Im Verlaufe des Prüfprozesses von WP1 oder EUSV1 bis zum LSV sowie über die LSV-Jahrgänge hinweg sinkt die Sortenanzahl kontinuierlich, während die Ortanzahl tendenziell steigt. Diese Gegenläufigkeit beider Prozesse dient der Optimierung des Aufwand-Nutzen-Verhältnisses, sie trägt aber zur Unbalanciertheit bei.

Orthogonale Strukturen im langjährig betrachteten integrierten Sortenprüfsystem wären folglich willkürlich eingegrenzte Datenauszüge, wobei zur Erzielung der Orthogonalität ein häufig großer und relevanter Anteil der Daten ausgeblendet werden müsste und in der Realität vieler Länderdienststellen auch wird.

Die hier diskutierten Ausprägungen der Unbalanciertheit können die Versuchsserienauswertung erheblich beeinflussen. Dies findet Berücksichtigung in der Modellwahl (s. 5.3). Darüber hinaus kommt der möglichst guten Einhaltung der Modellvoraussetzungen im Komplex Additivität – Varianzhomogenität – Normalverteilung der Residuen durch die Unbalanciertheit eine besonders große Rolle zu.

Von wesentlicher Bedeutung könnte die Frage sein, ob die durch Selektion hervorgerufene Unbalanciertheit zu Verzerrungen führt. Laut Untersuchungen von Piepho und Möhring (2006) kann für den vorliegenden Fall des Sortenprüfsystems die ‚missing at random'-Annahme (MAR) in der Definition nach Little und Rubin (2002) als erfüllt gelten, wenn alle im Prüfprozess involvierten Daten auch in die Auswertung

einfließen (s.a. Ausführung in Abschn. 3). Daher wird vom willkürlichen Ausschluss von Prüfgliedern abgeraten, unabhängig von Ihrer Besetzung und der Dauer ihres Verbleibs im Prüfprozess oder von ihrer aktuellen Bedeutung. Diese Aussage steht dem zunächst subjektiv nachvollziehbaren Ansatz vieler Versuchsansteller entgegen, wonach z.B. Stämme ohne erlangte Zulassung im Interesse kurzer Rechenzeiten und übersichtlicher Strukturen aus der Auswertung entnommen werden. Im Rahmen dieser Arbeit wurde aber auf eine derartige Datensatzreduktion verzichtet, um die MAR-Annahme hinreichend zu erfüllen.

Speziell der Faktor Sorte ist in zeitlicher und räumlicher Hinsicht hochgradig von Unbalanciertheit betroffen. Nachfolgend sollen besonders gravierende und fachlich relevante Balanciertheitsprobleme beispielhaft für den Datensatz *Kornertrag Winterweizen* gezeigt und diskutiert werden.

Unbalanciertheit hinsichtlich des Faktors ‚Sorte' (S)

Bereits Silvey (1978) hat beschrieben, dass die Offizialprüfung in Großbritannien fast ausnahmslos lückige Tabellen hervorbringt. Patterson (1997) beschreibt, dass daher eine einfache Mittelwertbildung über Umwelten zur Schätzung der Sortenmittelwerte nicht angebracht ist, da durch die generell großen Unterschiede zwischen den Versuchen die Mittelwertvergleiche dadurch verzerrt würden.

Auch die Sorten im hier diskutierten Datensatz Winterweizen weisen einen äußerst heterogenen Datenhintergrund auf, der bereits in der Anzahl Werte je Sorte zum Ausdruck kommt (Abb. 5). Von den 1118 im Ertrag einbezogenen Sorten des Gesamtsortimentes sind 11 Sorten nur in einem Versuch geprüft worden, während das Maximum bei der Sorte Tommi mit 243 Werten lag. Besonders Sorten die nicht zum Kernsortiment gehören (1074), die z.T. bereits im Prozess der WP ausgeschieden sind, weisen häufig sehr wenige Werte auf. Dadurch lag die häufigste Besetzung (bei 406 Sorten) bei nur zwei Werten, der Median bei drei und das arithmetische Mittel bei ca. neun Werten je Sorte. Obgleich der Beitrag der selten geprüften Sorten zur Schätzung der Varianzkomponenten etc. vergleichsweise gering ist, sollten auch die sehr schwach besetzten Sorten aus o.g. Gründen nicht von der Auswertung ausgeschlossen werden (um die MAR - Annahme nicht zu verletzen).

Im Kernsortiment (44 von 1118 Sorten) ist die Besetzung - in Abhängigkeit vom Prüfstatus (Definition s. Tab. 8) - deutlich überdurchschnittlich (Abb. 5). Als Kernsortiment sind in dieser Arbeit nicht diejenigen Sorten zu verstehen, die aus langjähriger Sicht am häufigsten geprüft wurden, sondern diejenigen Sorten, die im letzten Prüfjahr Beratungsrelevanz besaßen, die also in diesem Jahr im LSV stehen oder zwar aus der Prüfung genommen wurden, aber noch empfohlen werden. Somit zählen einige besonders häufig geprüfte Sorten (s.a. Abb. 5) im aktuellen (letzten) Jahr nicht mehr zum Kernsortimentes, sofern sie nicht mehr geprüft und auch nicht mehr empfohlen werden. Nach dem LSV1 liegen im Mittel bereits 20, nach LSV2 46 und nach LSV3 66 Werte je Sorte als Empfehlungsgrundlage vor (Tab. 13). Gerade unter den Sorten mit sehr hoher Anzahl Ergebnisse gibt es kaum ein Sortenpaar, das orthogonal genau in identischen Einzelversuchen geprüft worden ist - fast jede Sorte hat hier also einen individuellen Datenhintergrund (Tab. 10; Abb. 5). Die zweitstärkste Besetzung (Drifter, N=185) und die drittstärkste Besetzung (Batis, N=179) weisen zwei Sorten auf, die nicht mehr Teil des aktuellen Kernsortimentes sind (da sie nicht mehr im aktuellen LSV 2009 standen und 2009 auch nicht mehr in der Empfehlung waren und damit per Definition nicht in die Rubrik ‚alte Sorten' fallen). Beide Sorten waren ebenso wie die weiteren besonders häufig geprüften

Sorten Tommi (N=243), Türkis (N=172) und Hermann (N=164) langjährig in allen Versuchstypen als VRS vertreten. Dies zeigt nochmals, wie wichtig VRS / VGL und im Besonderen auch die Belassung alter Jahrgänge im Datensatz ist. Diese Sorten und Jahrgänge tragen erheblich zum Nutzen der Einbeziehung älterer Versuchsergebnisse bei, da sie die erforderlichen ‚Brücken' zwischen den Versuchen, Jahren und Versuchstypen herstellen.

Tab. 13: Anzahl Mittelwerte je Sorte im Kernsortiment, gruppiert nach Prüfstatus

	Prüfstatus (Definition s. Tab. 8)							
	alte Sorten	**LSV5-8**	**LSV4**	**LSV3**	**LSV2**	**LSV1**	**OV_WP**	**OV_EU**
	Anzahl Sorten je Prüfstatus-Gruppe							
	7	8	2	3	6	4	12	2
Maßzahl	Anzahl Mittelwerte aus Einzelversuchen je Sorte							
Min	22	85	66	65	24	15	10	9
Median	59	117	78	66	46	20	10	9,5
Max	73	243	84	74	68	28	17	10

Unbalanciertheit hinsichtlich der Kombination ‚Sorte × Jahr' (S × J)

Die systematischen Bindungen über die Jahre und den aufsteigenden Prüfstatus im Zuge des Prüfverlaufes jeder Sorte führt zu einer systematisch versetzten Datenlage. Dies kommt in Tab. 14 (s.a. Tab. 10) in den nach links-unten zunehmenden Leerzellen und im Vergleich von Tab. 10 zu Tab. 11 im zeitlichen Versatz der Versuchstypen zum Ausdruck. Auch die besetzten Felder in Tab. 10 und Tab. 14 unterscheiden sich erheblich in der Wertanzahl, was unmittelbar auf den Zusammenhang zwischen Prüfstatus (bzw. Jahrgang) und Versuchstyp (Tab. 11) zurückzuführen ist. Mit aufsteigendem Versuchstyp von WP1 oder EUSV1 bis zum LSV nimmt die Anzahl Werte je Jahr tendenziell zu (von links nach rechts). Durch diese Zusammenhänge verringert sich der Anteil des im Fokus der Auswertung stehenden Kernsortimentes an den jeweiligen Jahresdaten mit zunehmendem Abstand zum Auswertungsjahr (2009) von 42 % bzw. 46 % in den Prüfjahren 2009 bzw. 2008 auf 3 % im Prüfjahr 2001 (Tab. 14).

Tab. 14: Durchschnittliche Anzahl Mittelwerte je Sorte im Kernsortiment nach Jahren, gruppiert nach Prüfstatus

	Versuchsjahr 2001 - 09								
Prüfstatus im Kernsortiment (Def. s. Tab. 8)	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
alte Sorten	0.3	1	3	4	8	15	13	11	2
LSV5-8	3	4	11	15	23	23	21	19	17
LSV4	.	.	3	3	4	11	17	18	22
LSV3	.	.	.	3	3	5	17	22	19
LSV2	.	.	.	.	2	3	5	13	23
LSV1	.	.	.	.	.	2	2	4	13
OV_WP	.	.	.	.	.	2	2	4	4
OV_EU	.	.	.	.	.	.	3	3	4
Anteil Kern-sortiment an Jahresdaten	**3 %**	**4 %**	**10 %**	**14 %**	**21 %**	**31 %**	**35 %**	**46 %**	**42 %**

Unbalanciertheit hinsichtlich der Kombination ‚Sorte × Jahr × Ort' (S × J × O)

Von den (1118×9×39) = 392418 möglichen S × J × O - Kombinationen waren nur 9512, also lediglich ca. 2,4 %, besetzt.

Unbalanciertheit hinsichtlich der Kombination ‚Jahr × Ort' (J × O)

Aus Tab. 12 geht die sehr differenzierte J × O - Besetzung hervor. Gleichzeitig ist aber auch zu ersehen, dass an vielen J × O - Kombinationen bis zu vier verschiedene Einzelversuche unterschiedlichen Versuchstyps angelegt wurden. Hier stehen also jeweils Sorten, die im Prüfsystem (s. Abb. 4) / Prüfstatus (s. Tab. 8) zeitlich versetzt sind, zeitgleich nebeneinander. Liegen z.B. eine WP und ein LSV gemeinsam an einer J × O - Kombination, so stehen noch nicht zugelassene Sorten bereits im indirekten Vergleich zu zugelassenen LSV-Sorten - unter nahezu gleichen Anbaubedingungen. VRS / VGL stellen dabei auch innerhalb der J × O - Kombination ‚Brücken' für den indirekten Vergleich dar.

Unbalanciertheit in Einzelversuchen

Auch unterschiedliche Wiederholungszahlen in Einzelversuchen bzw. Fehlwerte bei einzelnen Sorten in einem Einzelversuch führen zu Unbalanciertheit. Da die Auswertung der Versuchsserie aber auf Mittelwerten basiert, tritt diese Unbalanciertheit hier nicht in Form lückiger Daten in Erscheinung, sondern schlägt sich in differenzierten Standardfehlern der Mittelwerte aus Einzelversuchen ($SE_{Versuch}$) und - bei Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen - indirekt auch in der Serienauswertung nieder (s. 4.5).

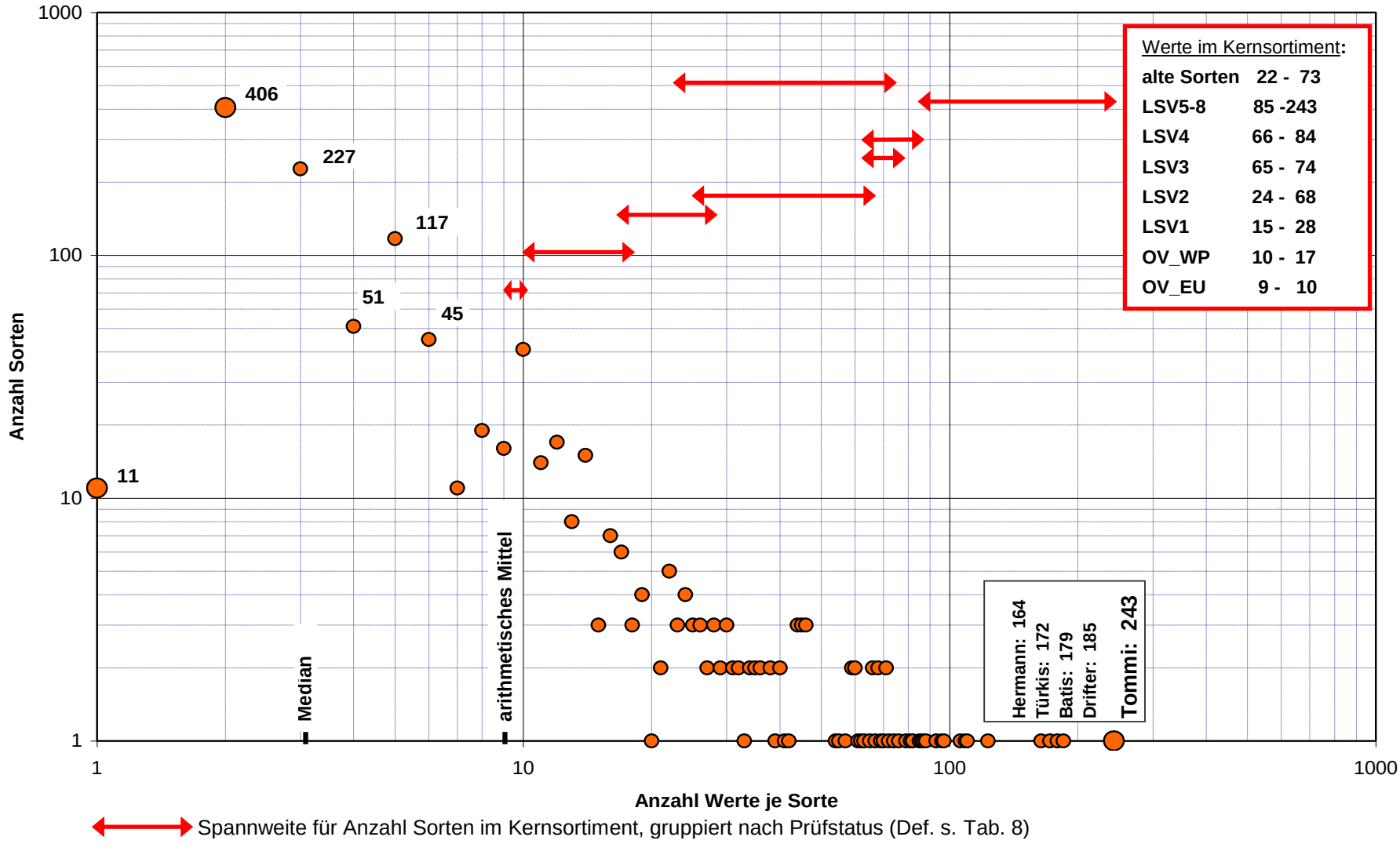


Abb. 5: Anzahl Mittelwerte je Sorte aus Einzelversuchen und deren Häufigkeiten im Sortiment Winterweizen (log-skaliert)

4.5 Differenzierte Präzision der Sortenmittelwerte aus Einzelversuchen

Die hier diskutierten Auswertungen von Versuchsserien basieren auf adjustierten Mittelwerten je Sorte aus Einzelversuchen (Zweischrittanalyse). Außer den ohnehin variierenden versuchsspezifischen Fehlervarianzen sind auch die Wiederholungszahl in den Einzelversuchen, die Anlagemethode, z.T. sogar die Faktorenanzahl nicht durchgehend identisch. Auch im jeweiligen Auswertungsmodell wurde versuchsspezifischen Besonderheiten Rechnung getragen - z.B. wurden z.T. räumliche Ansätze für die Modellierung der Fehlervarianz genutzt.

Die beide Komponenten Fehlervarianz und Wiederholungsanzahl ergeben z.T. sehr differenzierte Standardfehler der Mittelwerte bzw. der Mittelwertdifferenzen. Für die Gewichtung der Sortenmittelwerte aus Einzelversuchen in der Versuchsserienauswertung werden Standardfehler für die Gewichtung ermittelt. Zur Abgrenzung von den Standardfehlern der Mittelwerte aus der Versuchsserienauswertung (SE_{Serie}, s. 5.3.5) werden die Standardfehler der Mittelwerte aus Einzelversuchen im Weiteren mit $SE_{Versuch}$ bezeichnet. Die Bestimmung der $SE_{Versuch}$ wird in 5.3.3 dargelegt. Diese Bestimmung und damit die Gewichtung waren nur bei den Merkmalen möglich, für die in den Einzelversuchen Messungen an allen Wiederholungen erfolgten - dies trifft für *Ertrag* und überwiegend auch für *Minderertrag* zu. Bei den weiteren Merkmalskategorien lagen überwiegend keine Standardfehler vor, da in der Regel (außer beim Silo-und Körnermais) diese Qualitätsanalysen an Mischproben aus den Wiederholungen je Sorte und Versuch vorgenommen wurden, um Aufwand, Zeitbedarf und Kosten zu begrenzen.

Die Verteilungen der Variationskoeffizienten der Mittelwerte in Einzelversuchen sind in Abb. 2 deskriptiv dargestellt (bezogen auf den Mittelwert aller Urdaten der gesamten Serie).

Anmerkungen und Erläuterungen zur Kennziffer $SE_{Versuch}$% in Abb. 2:

Der Standardfehler eines Sortenmittelwertes in einem Einzelversuch wird in dieser Arbeit im Unterschied zu einem Sortenmittelwert über die Versuchsserie mit der Notation ‚$SE_{Versuch}$' versehen. Die Quadratwerte der Standardfehler, also die Fehlervarianzen ‚$SE^2_{Versuch}$', werden in 5.3.3 genauer definiert. Um Pflanzenart und Merkmal übergreifend eine Vergleichbarkeit dieser Standardfehler zu erzielen, erfolgt in Abb. 2 eine Relativierung auf das jeweilige arithmetische Mittel aller Urwerte, diese Relativzahlen werden als ‚$SE_{Versuch}$%' ausgewiesen. Es handelt sich hierbei zwar nicht um relative Standardabweichungen (von Einzelwerten), sondern um Standardfehler (von Mittelwerten) - trotzdem sollen sie hier zur Vereinfachung als *Variationskoeffizient von Mittelwerten* bezeichnet werden. Die Mediane dieser Variationskoeffizienten der Mittelwerte liegen etwa im Bereich 2% bis 4% des mittleren Ertragsniveaus. Auffällig ist die große Differenziertheit und die Schiefe der Variationskoeffizienten der Sorten-Mittelwerte bei Sommerbraugerste - ca. ¼ der Mittelwerte weist Variationskoeffizienten über 9,5% auf. Auf der absoluten Skala sind diese Standardfehler vergleichbar zu Winterweizen, aber das geringe Ertragsniveau führt zu sehr hohen Variationskoeffizienten. Diese dürften mit der bei Braugerste praktizierten geringen Stickstoffdüngung im Zusammenhang stehen. Auch bei Lupinen sind hohe Werte zu beobachten. Bei Erbsen ist ein sehr hohes Maximum bei sonst mittlerer Streuung zu erkennen. Sehr geringe $SE_{Versuch}$% in einem engen Werte-

bereich zeigt der Hafer, da diese Getreideart fast ausschließlich einfaktoriell und damit mit erhöhter Wiederholungszahl geprüft wurde.

In fast allen Datensätzen ist die Verteilung der Variationskoeffizienten deutlich linkssteil / rechtschief. Zum einen dürfte dies beim Ertrag mit der unteren Begrenzung des Parameterraums zusammenhängen. Hinzu kommt, dass Einzelwerte oder Wiederholungen in Einzelversuchen nicht selten (ungeplant) ausfallen, wohingegen zusätzliche geplante Wiederholungen (z.B. aufgrund einer größeren verfügbaren Versuchsfläche u.ä.) Ausnahmen sind. Diese Problematik hat u.a. die Frage der optimalen Gewichtung der Einzelversuchsergebnisse animiert.

Die Datenumfänge sind i.d.R. etwas kleiner als beim Originalmerkmal *Ertrag* (z.B. Winterweizen N=8874 ; 89% der Ertrags-Daten), da für einzelne Versuche keine Standardfehler vorlagen (z.B. von anderen Dienststellen importierte Versuche).

Die Variationskoeffizienten bei der Kategorie *Minderertrag* sind naturgemäß größer als die der Erträge, da sie auf Differenzen zweier Mittelwerte (unterschiedliche Intensität, gleiche Sorte) basieren.

Die Standardfehler sind i.d.R. versuchsspezifisch. Besonders bei Fehlwerten, aber auch bei Formen unvollständiger Blockung oder geostatistischer Ansätze, können die Mittelwert spezifischen Standardfehler innerhalb eines Versuches aber mehr oder weniger stark differieren. Da die automatische Speicherung von Ergebnissen im Zuge jeder Versuchsauswertung mit PIAFStat (s. 5.4) ohnehin Mittelwerte und deren assoziierte Standardfehler als $SE_{Versuch}$ speichert, werden für die Gewichtung nicht die versuchsspezifischen, sondern die Sortenmittelwert spezifischen $SE_{Versuch}$ genutzt. Dennoch wird im Weiteren der Terminus Einzelversuchsfehler synonym auch für Mittelwert spezifische Standardfehler innerhalb von Einzelversuchen verwendet.

5 Methoden

5.1 Regionalisierung

Das nachfolgend vorgestellte Gesamtkonzept für die Regionalisierung nach Michel und Zenk (2009) ist im unmittelbaren Zusammenhang und begleitend zur Entwicklung der *Hohenheim-Gülzower-Serienauswertung* unter maßgeblicher Mitwirkung des Autors dieser Arbeit entwickelt worden. Daher werden die methodischen Aspekte dieser Regionalisierung eingehender vorgestellt und diskutiert.

5.1.1 Boden-Klima-Räume für alle Pflanzenarten als Bausteine verschiedener Regionalisierungsziele

Die naturräumliche Gebietsgliederung der Bundesrepublik Deutschland nach Boden-Klima-Räumen (BKR) ist ein Ergebnis der Zusammenarbeit der Dienststellen für das Sortenversuchswesen der Bundesländer mit der (damaligen) Biologischen Bundesanstalt für Land- und Forstwirtschaft (dem jetzigen Julius-Kühn-Institut) (Roßberg et al., 2007).

Ein wesentliches Ziel war es, von Gebietsgliederungen basierend auf politischen Grenzen bzw. auch von sehr unterschiedlichen Regionalisierungsansätzen verschiedener Bundesländer oder Einrichtungen abzukommen. Stattdessen sollten Gebiete mit in sich relativ homogenen Standortbedingungen für die landwirtschaftliche Produktion auf einer rein standortkundlichen, methodisch deutschlandweit einheitlichen Grundlage abgegrenzt werden.

Darüber hinaus sollte mit den BKR die kleinste unteilbare regionale Untereinheit für verschiedene Regionalisierungsziele geschaffen werden. Dabei wird vorausgesetzt, dass für unterschiedlichste Zielstellungen und Pflanzenarten auch sehr unterschiedliche Regionalisierungen erforderlich sein können, zum einen in der Gebietsgröße und -anzahl, zum anderen aufgrund der differenzierten Bedeutung der einzelnen Standortparameter für die jeweilige Zielstellung bzw. Pflanzenart. Die Verwendung der BKR als Bausteine erleichtert und objektiviert die darauf aufbauenden Aggregationen und erhöht die Transparenz, Nachvollziehbarkeit und standortkundliche Fundiertheit.

Mit dieser einheitlichen naturräumlichen Gliederung werden die Belange des Pflanzenbaus sowie des Pflanzenschutzes berücksichtigt. Seitens der Biologischen Bundesanstalt kamen BKR u.a. bei der regionalisierten Erhebung zur Pflanzenschutzintensität im Ackerbau (NEPTUN-Projekt, s. Roßberg et al., 2007) zum Einsatz, wobei aus mehreren BKR größere Erhebungsregionen aggregiert wurden. Im Sortenwesen werden basierend auf BKR Pflanzenart spezifische Anbaugebiete für die Planung und Auswertung der Sortenversuche sowie für die Sortenberatung gebildet (s. 5.1.2).

Die BKR sind einheitliche Basis der Regionalisierung für alle Pflanzenarten. Erst in den unterschiedlichen Aggregationen zu übergeordneten Anbaugebieten erfolgt eine Differenzierung nach Pflanzenarten

BKR selbst stellen keinen Faktor im Auswertungsmodell dar. Da sie aber die Basis für den Faktor ‚Anbaugebiet' sind, werden sie hier so ausführlich erörtert.

Herleitung der Boden-Klima-Räume (BKR):

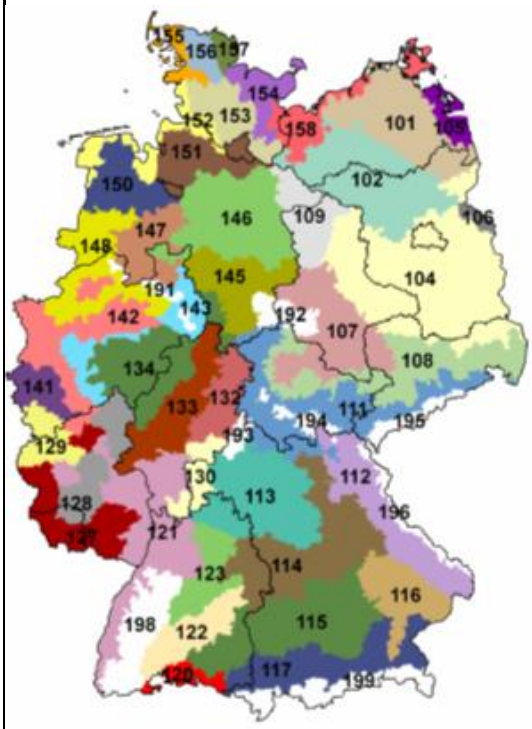
Als kleinste Flächenausgrenzung wurden Gemeindegrenzen gewählt, da dies den verfügbaren digitalisierten Standortdaten entsprach. Zur Quantifizierung des Einflussfaktors *Boden* wurden auf Basis der Leitbodenarten Deutschlands Bodenwertzahlen zugeordnet. Die Quantifizierung des Einflussfaktors *Witterung* erfolgte durch klimatologische Werte auf der Grundlage von 401 meteorologischen Stationen. Als Parameter wurden die mittleren Monatstemperaturen und die mittleren monatlichen Niederschlagssummen im Hauptvegetationszeitraum März bis August genutzt. Auf Basis einer Clusterung durch den Autor dieser Arbeit unter Berücksichtigung genannter Variablen (Bodenwertzahl, Temperatur und Niederschlag) wurden Gemeinden mit ähnlichen Eigenschaften bezüglich Bodengüte, Temperatur und Niederschlag zu größeren Gebieten zusammengefasst. Die Clusterung erfolgte mit der Software SAS® (PROC CLUSTER, METHOD=WARD). Die Bestimmung einer geeigneten Anzahl Gebiete erfolgte unter Berücksichtigung von lokalen Maximalwerten des *cubic clustering criterion* (CCC) in dem Teil des Definitionsbereiches, der die Bildung einer realistisch verwendbaren Anzahl an Clustern (Gebieten) erlaubt (Khattree und Dayanand, 2000). Es ergaben sich räumlich gut erkennbare, standortkundlich nachvollziehbare Strukturen. Durch Methoden der räumlichen Aggregation in geografischen Informationssystemen wurden sehr kleine Raumstrukturen in benachbarte Räume integriert. Räumlich nicht verbundene Gebiete, die zum gleichen Cluster zugeordnet waren, wurden als getrennte Gebiete betrachtet. In Ausnahmefällen mussten unter Einbeziehung von Expertenwissen in den Regionen Grenzkorrekturen erfolgen. Dies lag an der im Einzelfall ungeeigneten Übersetzung von ‚Leitbodenart' in ‚Bodenwertzahl', z.B. in Marschgebieten (BKR 152, 155 in Abb. 6).

Im Ergebnis wurden 44 Boden-Klima-Räume festgelegt (Abb. 6; in Mecklenburg-Vorpommern vorkommende BKR sind in der Legende grau unterlegt).

Struktur der Klassifizierungsgröße ‚BKR' in den Datensätzen (Tab. 5, 6, 7, 12):

Aus Gründen der fachlichen Einordnung und Bewertung wurden BKR als Klassifizierungsvariable in den Datensätzen mit ausgewiesen. Sie sind aber kein Faktor im Auswertungsmodell. In Tab. 6 zu beobachtende unterschiedliche BKR – Anzahlen von vier (Kartoffeln) bis elf (z.B. Winterweizen) resultieren nicht aus einer bewusst unterschiedlich gewählten Raumausdehnung, sondern aus unterschiedlicher Datenverfügbarkeit. So wurden z.B. Kartoffeln nur in sehr wenigen BKR geprüft, da sie auch nur in diesen BKR eine Anbaubedeutung besitzen.

Karte



Legende

Nr.	BKR-Bezeichnung
101	mittlere diluviale Böden MV und Uckermark
102	sandige diluviale Böden des nordostd. Binnentieflandes
104	trocken-warme diluviale Böden des ostdt. Tieflandes
105	vorpommersche Sandböden im Uecker-Randow-Gebiet
106	Oderbruch
107	Lößböden in der Ackerebene (Ost)
108	Lößböden in den Übergangslagen (Ost)
109	diluviale Böden Altmark u. Überlapp. Nördlich. Nieders.
111	Verwitterungsböden in den Übergangslagen (Ost)
112	Verwitterungsböden in Höhenlagen (östliches Bayern)
113	Nordwestbayern-Franken
114	Albflächen und Ostbayerisches Hügelland
115	Tertiär-Hügelland Donau-Süd
116	Gäu, Donau- und Inntal
117	Moränen-Hügelland und Voralpenland
120	Hochrhein-Bodensee
121	Rheinebene und Nebentäler
122	Schwäbische Alb, Baar
123	Oberes Gäu und körnermaisfähige Übergangslagen
127	Mittellagen Rheinland-Pfalz und Saarland
128	Hunsrück, Westerwald
129	sandiger Lehm / Eifel / Höhenlagen
130	Odenwald, Spessart
132	Osthessische Mittelgebirgslagen
133	Zentralhessische Ackerbaugebiete, Warburger Börde
134	Lehmböden / Sauerland, Briloner Höhen / Höhenlagen
141	Lößböden / Köln-Aachener Bucht / Niederungslagen
142	Lehmböden / oberer Mittelrhein, Niederrhein, südliches Münsterland / Niederungslagen
143	Lehmböden / Ost-Westfalen, Haarstrang, Bergisches Land, Voreifel / Übergangslagen
145	Lehmböden / Südhannover
146	sandige Böden / Lüneburger Heide, nördliches NRW
147	leichtere Lehmböden / mittleres Niedersachsen, nordöstliches NRW
148	Sandböden / südwestliches Weser-Ems-Gebiet, nördliches Münsterland / Niederungslagen
150	nordwestliches Weser-Ems-Gebiet / sandige Böden
151	Elbe-Weser-Dreieck / sandige Böden
152	Niedersächsische Küsten- und Elbmarsch
153	Geest - Süd
154	südliches schleswig-holsteinisches Hügelland
155	Marsch - Nord
156	Geest - Nord
157	nördliches schleswig-holsteinisches Hügelland
158	Nordwest-Mecklenburg / Rügen / Südost-Holstein
191	Teutoburger Wald
192	Harz
193	Rhön
194	Thüringer Wald
195	Erzgebirge
196	Bayrischer Wald
198	Schwarzwald
199	Alpen
103	Niederungsstandorte NO-Deutschland (überwiegend Moore)
160	Moore Nordwest-Deutschland

Legende: grau unterlegte BKR liegen anteilig im Bundesland Mecklenburg-Vorpommern

Abb. 6: Boden-Klima-Räume für alle Pflanzenarten in Deutschland (nach Roßberg et al., 2007)

5.1.2 Pflanzenartspezifische Anbaugebiete für die Sortenempfehlung

Im Sinne der Zielstellung einer *regionalen* Sortenempfehlung stellen die Anbaugebiete den regionalen Bezug her. Damit ist *Anbaugebiet* nach *Sorte* der zweite inhaltlich zentrale Faktor im Auswertungsmodell.

Auf Basis der Gliederung Deutschlands in Boden-Klima-Räume (5.1.1) wurden für 20 Pflanzenarten jeweils spezifische Anbaugebiete definiert (Graf et al., 2009). Ein Anbaugebiet setzt sich aus einem oder häufiger aus mehreren der insgesamt 44 Boden-Klima-Räume zusammen. Der unterschiedliche Zuschnitt der Anbaugebiete berücksichtigt sowohl die aktuelle Anbaubedeutung der einzelnen Pflanzenarten (Anbaugebiete sind also nicht für alle Zeit festgelegt) als auch deren unterschiedliche Präferenzen bezüglich der Standortparameter (Boden und Klima). Jedes Anbaugebiet stellt ein Beratungsgebiet im Sortenwesen dar.

Wichtigstes Ziel dieser Regionalisierung war es, die notwendige Genauigkeit sowie Repräsentativität der Versuchsergebnisse auch bei eingeschränkten Ressourcen im Versuchswesen zu gewährleisten. Die Anbaugebiete bilden gleichzeitig die Grundlage für eine optimierte überregionale länderübergreifende Versuchsplanung.

Stellvertretend für insgesamt 20 Karten sind hier die Anbaugebiete Winterweizen (Abb. 7) und Sommerweizen (Abb. 8) dargestellt. Aufgrund der erheblich geringeren Anbaubedeutung und der entsprechend geringeren Standortanzahl sind hierbei die Anbaugebiete beim Sommerweizen erheblich größer als beim Winterweizen. Beim Winterweizen sind bundesweit 22, beim Sommerweizen nur sieben Anbaugebiete ausgewiesen.

An dieser Stelle sei noch einmal darauf hingewiesen, dass die Abgrenzungen von BKR und Anbaugebieten nicht mit den politischen Grenzen der Bundesländer übereinstimmen. Die fachliche Zuständigkeit der regionalen Sortenberatung ist aber den Bundesländern zugeordnet. In dieser Arbeit wurden nur die Gebiete mit Relevanz für das Bundesland Mecklenburg-Vorpommern berücksichtigt. Dabei handelt es sich um einen Großraum, der sich aus Anbaugebieten bzw. BKR der diluvialen (eiszeitliche geprägten) Standorte im Nord-Osten Deutschlands ergibt (s. Abb. 9; 5.1.3).

Nur zwei der 22 Winterweizen-Anbaugebiete tangieren unmittelbar das Gebiet Mecklenburg-Vorpommerns (Anbaugebiete 4: Diluviale Standorte Nord, Kürzel: ‚D-Nord' und Anbaugebiet 45: Mecklenburg Süd/ Brandenburg Nordwest, Kürzel: ‚MV-Süd'). Dahingegen gehören alle Standorte in Mecklenburg-Vorpommern zum Sommerweizen-Anbaugebiet 5 (Diluviale Standorte Ostdeutschland).

Neben den Anbaugebieten, die unmittelbar Mecklenburg-Vorpommern tangieren wurden aber auch diluviale Anbaugebiete in einem südlich, südwestlich und westlich ausgeweiteten Überlappungsgürtel berücksichtigt. Die Größe des einbezogenen Raumes war über alle Pflanzenarten hinweg weitgehend identisch, soweit Versuchsergebnisse aus allen Bereichen vorlagen. Daher spiegeln die Unterschiede in der Anzahl Anbaugebiete (Tab. 6) die Größe der Anbaugebiete wieder – je mehr Anbaugebiete ausgewiesen sind, desto feiner aufgelöst erfolgte die Regionalisierung (s. 5.1.3). Im Maximum wurden fünf Anbaugebiete verwendet (Tab. 6, Winterweizen und Raps).

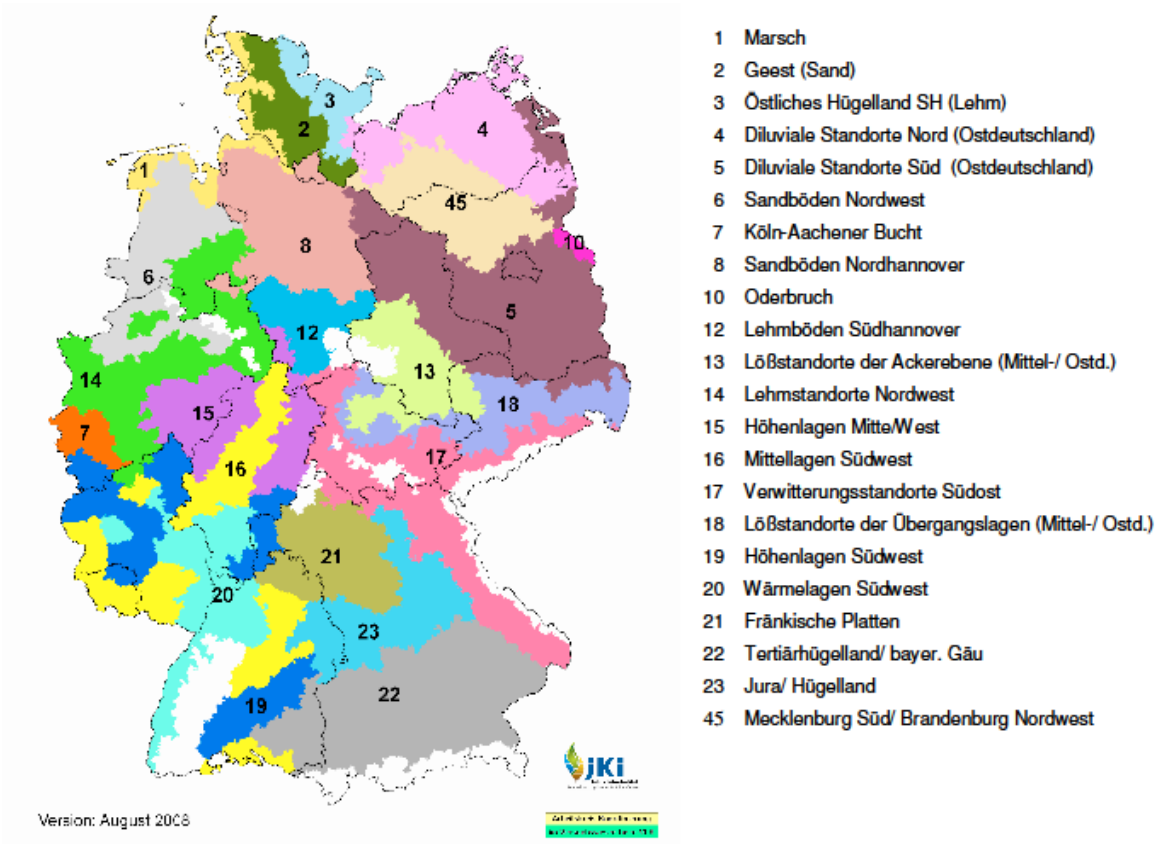


Abb. 7: Anbaugebiete Winterweizen in Deutschland (Graf et al., 2009)

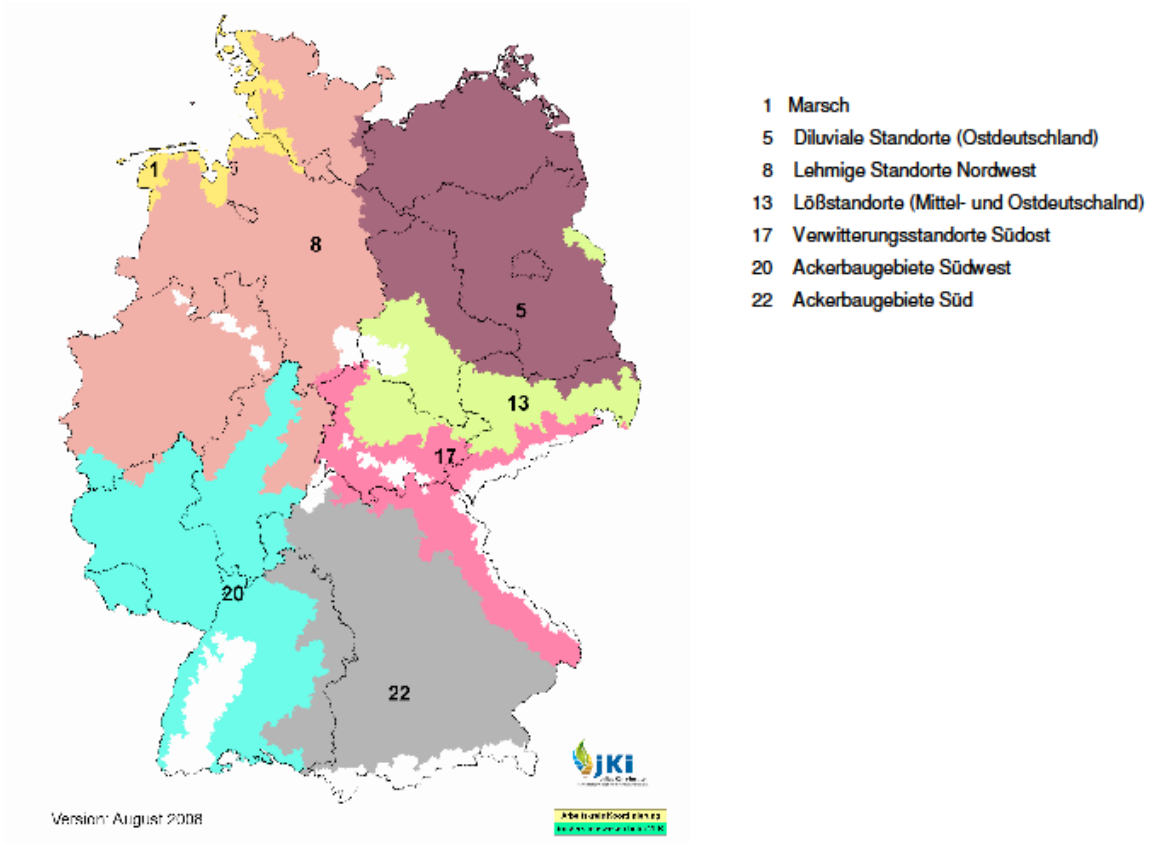


Abb. 8: Anbaugebiete Sommerweizen in Deutschland (Graf et al., 2009)

5.1.3 Konzept von Zielgebiet und Einbeziehung von Nachbargebieten

Standortkundliche Ähnlichkeiten zwischen benachbarten BKR begründen die Zusammenfassung zu einem Anbaugebiet (s. 5.1.2). Allerdings sollten auch die Grenzen zwischen diesen übergeordneten Anbaugebieten nicht als distinkte qualitative Sprünge verstanden werden. Auch zwischen benachbarten Anbaugebieten können Ähnlichkeiten und gleitende Übergänge beobachtet werden. Ähnlichkeiten in standortkundlichen Parametern lassen auch korrelierte Sorteneffekte erwarten. Dieser Aspekt bildet einen der grundlegenden Eckpfeiler der *Hohenheim-Gülzower-Serienauswertung* – die abgestuft gewichtete Einbeziehung von Versuchsergebnissen benachbarter Anbaugebiete.

Untersuchungen im Rahmen einer Studie im Auftrag von Bundessortenamt, Bund Deutscher Pflanzenzüchter und der Länderdienststellen für die Sortenprüfung ergaben, dass auswertungsseitig das Prinzip fixierter Anbaugebiete bzw. Großräume durch eine dynamischen Zusammenfassung von Anbaugebieten, die einen größeren Standortbereich abdecken, ersetzt werden sollte. Ein wesentlicher Aspekt des vorgeschlagenen Ansatzes ist nach Möhring et al. (2004a), dass die Anbaugebiete als solche bestehen bleiben (und nicht, wie z.T. vorgeschlagen, zu wenigen Großräumen ausgeweitet werden) und dass die integrierte Auswertung (trotz der eingetretenen Reduzierung der Versuchsanzahl) über ein gewichtetes Mittel aus z.T. mehreren Anbaugebieten erfolgt.

Das Anbaugebiet, für das im Rahmen einer Auswertung Schätzwerte ermittelt werden sollen, wird im Weiteren als *Zielgebiet* (ZG) bezeichnet. Benachbarte Anbaugebiete, deren Ergebnisse die Einschätzung für das Zielgebiet unterstützen sollen, werden als *Nachbargebiete* (NG) bezeichnet. Dabei stellen im Rahmen dieser Arbeit Anbaugebiete, die die Fläche des Bundeslandes Mecklenburg-Vorpommern überschneiden, jeweils ein Zielgebiet dar, während westlich, südwestlich und südlich gelegene Gebiete als Nachbargebiete mit abgestufter Gewichtung in die Auswertung einfließen (Abb. 9).

Zielgebiete entsprechen hier immer exakt einem bundesweit definierten Anbaugebiet. Nachbargebiete werden hier aus Gründen a) der Datenbesetzung oder b) aus dem standortkundlichen Blickwinkel des Zielgebietes zum Teil abweichend zugeschnitten. Sie setzten sich dann aber trotzdem immer exakt aus vollständigen BKR zusammen. Im Interesse der Verallgemeinerungsfähigkeit lassen sich alle einbezogenen Nachbargebiete grob den drei Räumen Süd (S̱), West (W̱) und Südwest (S̱W̱) oder deren verschiedenartigen Zusammenfassungen zuordnen. Dies stellt eine konzentrische Sicht dar, die von den Zielgebieten in MV ausgeht und sie in den Schwerpunkt legt.

Daraus ergibt sich die in Tab. 15 dargestellte Regionalisierung bei den einzelnen Pflanzenarten. Soweit nicht wie bei Raps und Winterweizen merkmalsspezifisch untersetzt gilt diese Einteilung für alle analysierten Merkmale. Diese Einteilung der Nachbargebiete nach Himmelsrichtungen wird im Weiteren durchgehend verwendet.

Tab. 15: Regionalisierungskonstellationen für Zielgebiete in Mecklenburg-Vorpommern und einbezogene Nachbargebiete

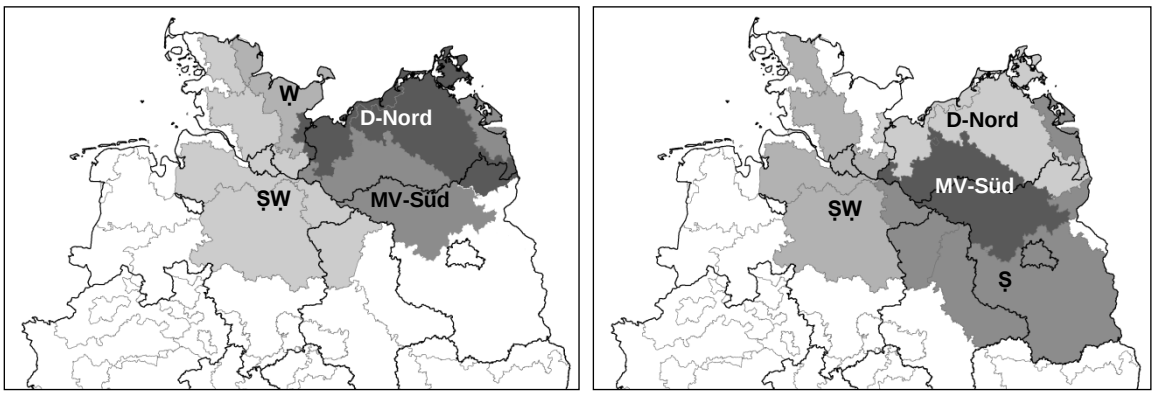
Pflanzenarten (Kürzel; s. Tab. 1)	**Anzahl (ZG+NG)**	**Zielgebiete (ZG)** und jeweilige Nachbargebiete (NG)				
		D-Nord	**MV-Süd**	**Ṣ**	**ṢẆ**	**Ẉ**
WW, RAW (nur *Ertrag*)	(5)					
Zielgebiet D-Nord	4	**ZG**	NG	-	NG	NG
Zielgebiet MV-Süd	4	NG	**ZG**	NG	NG	-
WW (nur *Minderertrag*)	3	**ZG**		NG	NG	
GW, MS	3	**ZG**		NG		NG
RW, TIW	2	**ZG**		NG		
GS_B, Ha, EF, MK	2	**ZG**			NG	
GS_F, LUB, KA, RAW (*Minderertrag, Qualitäten)*	1	**ZG**				

Nachbargebiete außerhalb des Bundeslandes Mecklenburg-Vorpommern in drei grob nach Himmelsrichtung ausgerichteten Räumen: Süd (Ṣ), West (Ẉ) und Südwest (ṢẆ)

Je nach Umfang des Datenbestandes, der von der Anbaubedeutung der Pflanzenarten in den einzelnen Anbaugebieten abhängt, wurden je Zielgebiet bis zu drei Nachbargebiete einbezogen. Beim Minderertrag Winterweizen wurden die beiden Ertrags-Zielgebiete zu einem zusammengefasst. Beim Minderertrag und den Qualitätsmerkmalen Raps wurde im Gegensatz zum Ertrag nur ein Großraum gebildet (Tab. 15). Die unterschiedliche Gebietszusammenfassung von Roggen und Triticale gegenüber anderen Pflanzenarten mit ebenfalls genau zwei Anbaugebieten ist in der regional unterschiedlichen Bedeutung und Besetzung mit Versuchsdaten begründet: bei Roggen und Triticale konnte das Zielgebiet für MV relevanter eingegrenzt werden, bei den anderen Arten musste das südlich gelegene Anbaugebiet ‚D-Süd' (entspricht hier i.d.R. ‚Ṣ') in das Zielgebiet integriert werden. Für die Pflanzenarten mit geringer Datenbasis (Sommerfuttergerste, Lupine, Kartoffeln sowie bei einzelnen Merkmalen für Raps) konnte daher nur ein sehr großes Gesamtgebiet - quasi als Großraum - gebildet werden, ohne dass in Zielgebiet und Nachbargebiete untersetzt wurde.

Aufgrund der besonders großen Anbaubedeutung und des entsprechend großen Datenbestandes bei Raps und Winterweizen konnte hier beim Ertrag für zwei separate Zielgebiete in Mecklenburg-Vorpommern ‚*D-Nord*' und ‚*MV-Süd*' ausgewertet werden. Die Überlappungsräume für diese beiden Zielgebiete decken sich nicht völlig (Abb. 9). Das Nachbargebiet Ẉ (in Abb. 7 Anbaugebiet 3; Östliches Hügelland SH) wurde als Nachbargebiet nur für das direkt benachbarte Zielgebiet D-Nord (AG 4) herangezogen (Abb. 9a). Das Nachbargebiet Ṣ (in Abb. 7 Anbaugebiet 5; diluviale Standorte Süd) wurde nur für das direkt benachbarte Zielgebiet MV-Süd (AG 45) herangezogen (Abb. 9b). Hierin kommt das Prinzip von sich *dynamisch überlappenden Großräumen* zum Ausdruck: jedes Zielgebiet liegt in einem nur auf dieses Zielgebiet zugeschnittenen Großraum. Die Großräume zweier benachbarter Zielgebiete haben eine mehr oder weniger große Schnittmenge. Von Zielgebiet zu Zielgebiet ist eine dynamisch gleitende Verschiebung des Großraumes zu verzeich-

nen. Diese Verschiebung ist allerdings kein Automatismus, sondern sollte standortkundlich begründet sein. Innerhalb der Anbaugebiete mit diluvialen Standorten können sich aus standortkundlicher Sicht die Überlappungsräume durchaus gleitend verschieben. Dagegen sollte z.B. die Grenze zwischen Diluvial- und Lößstandorten eine nicht zu überschreitende Grenze darstellen (u.a. BKR 104 zu BKR 108 in Abb. 6).



(a) D-Nord und Nachbargebiete (b) MV-Süd und Nachbargebiete
dunkelgraue Linien: Grenzen der Bundesländer; hellgraue Linien: Grenzen von BKR
dunkle Fläche: jeweiliges Zielgebiet; hellere Flächen: einbezogene Nachbargebiete

Abb. 9: Regionalisierung Winterweizen und Raps (*Ertrag*); Zielgebiete (a) D-Nord bzw. (b) MV-Süd und jeweils einbezogene Nachbargebiete

Die Auswertung setzt die Ergebnisse im Zielgebiet in den Schwerpunkt, Ergebnisse aus Nachbargebieten mit nachweislich sehr ähnlichen Sortenrangfolgen können mit geringerem Gewicht einfließen, wobei die Gewichtungsabstufung über die Schätzung der Ähnlichkeit von Sortenleistungen objektiviert und optimiert wird (5.3.5 ff).

5.2 Transformation

Gemischten Modellen liegen die Annahmen von Additivität der fixen und zufälligen Effekte (die Modellgleichung lässt sich als Linearkombination der Modellparameter schreiben), Varianzhomogenität und Normalverteilung (die Verteilung der Zufallseffekte folgt einer Gauß'schen Normalverteilung, zeigt insbesondere Symmetrie) zugrunde. Eine Verletzung dieser Annahmen kann zu einem Verlust an Power, sowie zu ungenaueren Mittelwertschätzungen führen (Stroup, 2014), u.U. in besonderem Maße bei einer ausgeprägten Unbalanciertheit wie im Sortenprüfsystem.

Varianzhomogenität, Normalverteilung und Additivität sollten aufgrund ihrer Wechselbeziehungen im Zusammenhang betrachtet werden. Gibt es z.B. eine Korrelation zwischen Erwartungswert und Varianz bzw. schiefe Verteilungen der Residuen, so sind in der Regel alle drei Voraussetzungen der Varianzanalyse verletzt (Thomas, 2006). Daher berührt die nachfolgende Diskussion implizit diesen Gesamtkomplex.

Unter Umständen kann bereits eine einfache Datentransformation die Varianzen so stabilisieren, dass eine Standardauswertung mit einfachen Varianz-Kovarianz-Strukturen erfolgen kann. Kommt Varianzheterogenität als maßgebliche Ursache für die Verletzung von Modellvoraussetzungen in Betracht, so wäre zu prüfen, ob die Zulassung von Varianzheterogenität im Auswertungsmodell geeigneter ist, als die Datentransformation (Piepho, 2009). Nach geeigneter Transformation soll die Auswertung immer unter Annahme eines linearen (additiven) Modells für die transfor-

mierten Daten erfolgen. Wenn Effekte z.B. multiplikativ wirkten, würde eine logarithmische Transformation Additivität auf der transformierten Skala gewährleisten. Eine logarithmische Transformation ist aber nicht zwangsläufig angemessen, es sollten auch alternative Transformationen in Betracht gezogen werden.

Definition des linearen Modells

Die nachfolgend dargelegten Transformationsansätze werden vor der Definition des zugrundliegenden linearen Modells vorgestellt. Diese Reihenfolge wurde gewählt, da im Praxiseinsatz die Schätzung eines Transformationsparameters und ggf. die damit durchgeführte Vorab-Transformation der Daten der erste Auswertungsschritt der Serienauswertung ist. An dieser Stelle muss dafür allerdings ein Vorab-Verweis auf die erst nachfolgende Darlegung des Auswertungsmodells in 5.3 erfolgen. Die Bestimmung des Transformationsparameters erfolgt mit dem Modell 7a in Tab. 17.

5.2.1 Box-Cox-Transformation

Im Rahmen dieser Arbeit wurde überwiegend die Klasse der Box-Cox-Transformationen (Box und Cox, 1964) genutzt. In 5.2.2 ist ein alternativer Transformationsansatz für die Auswertung von Prozentzahlen dargelegt.

Die Box-Cox-Transformation ist definiert durch:

$$y = f(x) = (x^{\varphi} - 1)/\varphi \qquad \text{für } \varphi \neq 0,$$
$$y = f(x) = \ln(x) \qquad \text{für } \varphi = 0.$$

wobei φ ein wählbarer Transformationsparameter und x ein Beobachtungswert ist.

Eine wesentliche Stärke der Box-Cox-Transformation liegt in einer flexiblen, stufenlos anpassbaren Transformation, die u.a. die logarithmische Transformation ($\varphi = 0$; impliziert ein multiplikatives Modell für die untransformierten Daten), die Wurzel-Transformation ($\varphi = 0{,}5$) und die Nicht-Transformation ($\varphi = 1$; impliziert ein additives Modell für die untransformierten Daten) als spezielle Fälle mit einschließt (s. Abb. 11). Der Definitionsbereich umfasst im Grundsatz auch die Wertebereiche $\varphi < 0$ und $\varphi > 1$. Z.B. ist die reziproke Transformation ($\varphi = -1$) zur Varianzstabilisierung angezeigt, wenn die Varianz proportional zur vierten Potenz des Erwartungswerts ansteigt (Thöni, 2007).

Der bestmögliche Wert für φ wird durch Maximierung der Likelihood-Funktion ermittelt (Piepho und Ogutu, 2003; Gurka et al., 2006). Es wird ein iteratives Verfahren verwendet, wobei für jeden Schritt das komplette Modell angepasst werden muss. In dieser Arbeit kommt dafür die Modellgleichung 7a (Tab. 17) zur Anwendung, wobei Varianzheterogenität bezüglich des Effektes S•AG in der Modellgleichung zugelassen wird und die Kovarianzstruktur durch ‚unstructered' Typ UN(1) definiert ist. Voraussetzung ist dabei eine Standardisierung, die die infolge der differenzierten Transformation veränderten Skalenabschnitte mittels der *Determinante der Jacobi-Matrix* berücksichtigt (Erläuterung: durch die Transformation ändert sich der Skalenabschnitt der Variablen, über den bei der Berechnung von Wahrscheinlichkeiten aus der Dichtefunktion zu integrieren ist; daher ändern sich die Integrationsgrenzen; die Determinante der Jacobi-Matrix berücksichtigt diese Änderung in der Variablen). Im Zuge der Schätzung des Transformationsparameters

wird entsprechend Möhring et al. (2004b) und mittels eines durch ihn erstellten Makros jeweils eine Grafik für das Likelihood-Profil erstellt (Abb. 10). Diese zeigt die Güte des Modells (die sog. Log-Likelihood bzw. ‚*loglik*') als Kurve in Abhängigkeit vom Transformationsparameter. Der approximativ ermittelte Wert von φ am Maximalpunkt der Kurve (maximaler Log-Likelihood) weist die beste Anpassung aus und ist somit der Maximum-Likelihood-Schätzer des Transformationsparameters.

Ein Vertrauensintervall für den Transformationsparameter wird grafisch ausgewiesen: Auf Basis des Likelihoodverhältnis-Tests (Likelihood ratio test) für die Chi-Quadrat verteilte Likelihood wird für eine vorzugebende Irrtumswahrscheinlichkeit eine horizontale Linie (Abb. 10, rote Linie) definiert, welche die Log-Likelihood-Kurve schneidet. Die Projektionen der beiden Schnittpunkte auf die x-Achse grenzen dementsprechend das Vertrauensintervall ab (Abb. 10, zwei Lote). Von Interesse ist insbesondere, ob $\varphi = 1$ (kein Transformationserfordernis) oder andere markante Eckwerte (wie $\varphi = 0$; logarithmische Transformation geeignet) sich vom optimalen Transformationsparameter signifikant unterscheiden.

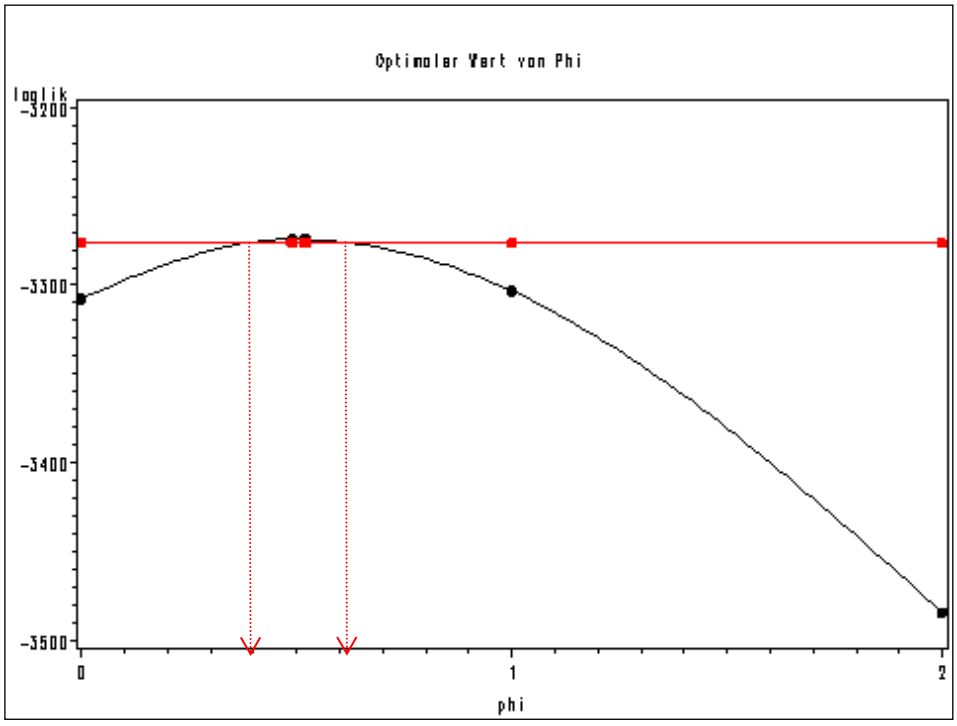


Abb. 10: Schätzung des Transformationsparameter phi (φ) (Beispiel)

Die Maximierung der Likelihood-Funktion basiert auf den üblichen Annahmen wie Normalverteilung, Varianzhomogenität und Additivität auf der transformierten Skala. Im Rahmen einer Studie (Möhring et al., 2004b) konnte gezeigt werden, dass in den meisten untersuchten Fällen die Modelle für die transformierten Daten signifikant besser als für die untransformierten Daten waren. Wenn auch die gleichzeitige Erfüllung der Modellvoraussetzungen hier nicht explizit getestet wird und nach dieser Methode nicht vorausgesetzt werden kann, so gewährleistet die Transformation zumindest, dass gegebene Daten bestmöglich an die Modelle angepasst werden, sodass nach dieser Transformation die Daten durch ein lineares, homoskedastisches Modell am besten beschrieben werden können (Thöni, 2007). Damit werden Verzerrungspotenziale auch bei der Mittelwertschätzung reduziert.

Im Zuge der Datentransformation müssen auch die Standardfehler der Mittelwerte (hier die $SE_{Versuch}$) aus Einzelversuchen transformiert werden, wenn diese als Gewichtungsfaktoren im Auswertungsmodell der Versuchsserie genutzt werden sollen. Diese Option der Gewichtung ist im derzeitigen Stadium der Implementierung in SAS-Makros und PIAFStat-Verfahren (s. 5.4) bei der Schätzung der Varianzkomponenten (Abb. 14) und der Mittelwerte (Abb. 15) umgesetzt, bei der Schätzung des Transformationsparameters aber noch nicht berücksichtigt (Abb. 13). Dies wäre eine Zielsetzung für die Weiterentwicklung der Verfahren, damit die genannten drei Schritte von identischen Ansätzen ausgehen. Die Transformation der Fehlervarianzen der Mittelwerte erfolgt nach der Delta-Methode (Johnson et al., 1993):

Transformation der Standardfehler der Sortenmittelwerte in Einzelversuchen:

$$SE^2_{Versuch}(y) = SE^2_{Versuch}(x) \times x^{(2\times(\varphi-1))}$$

wobei x den Original-Mittelwert und y den transformierten Mittelwert indiziert (dies gilt auch in nachfolgenden Transformationsanleitungen).

Wenn die Serien-Mittelwerte und deren Standardfehler (SE_{Serie}) im Ergebnis in der metrischen, fachlich interpretierbaren ursprünglichen Dimension dargestellt werden sollen, können diese nach dem Modell-Durchlauf wiederum rücktransformiert werden. Bei den rücktransformierten Mittelwerten handelt es sich um Schätzwerte für den Median, nicht für den Erwartungswert (Piepho, 2009). Wenn die Ursache für die Transformation z.B. mit einer schiefen Verteilung der Zufallseffekte im Zusammenhang steht, so ist auf der untransformierten Skala der Median ohnehin oft ein geeigneterer Parameter der Lage als das arithmetische Mittel.

Rücktransformation der Serien-Mittelwerte im Zuge der Box-Cox-Transformation:

$$x = (\varphi \times y + 1)^{(1/\varphi)} \qquad \text{für } \varphi \neq 0,$$

$$x = e^y \qquad \text{für } \varphi = 0.$$

Rücktransformation der Serien-Standardfehler:

$$SE_{Serie}(x) = SE_{Serie}(y) \times (x/x^{\varphi})$$

5.2.2 Relativzahlen und die ‚folded exponential transformation'

Für Prozentzahlen beschreibt Piepho (1998) verschiedene Möglichkeiten der Auswertung in Abhängigkeit von der Art der Erfassung (diskret / kontinuierlich) und von der Anzahl der Zufallskomponenten im Auswertungsmodell. Diskutiert wurden Generalisierte Lineare Modelle (GLM) für eine Zufallskomponente und Generalisierte Lineare Gemischte Modelle (GLMM) für mehr als eine Zufallskomponente. Beispieldaten wurden mit der SAS®-Prozedur GENMOD bzw. dem SAS®-Makro %GLIMMIX ausgewertet. Piepho zeigt, dass ein additives Modell für untransformierte Prozentzahlen häufig nicht sinnvoll ist. Verschiedene Transformationsansätze werden angesprochen. 2003 beschreibt Piepho die ‚*folded exponential transformation for proportions*'. Das dahinter stehende Grundprinzip wird im theoretischen Ansatz für die in dieser Arbeit verwendeten ausschließlich kontinuierlichen Prozentzahlen als besonders angemessen betrachtet. Sie vereinigt folgende Vorzüge in sich:

(1) Die in dieser Merkmalskategorie u.U. auftretenden Grenzwerte des Parameterraumes $x = 0$ und $x = 1$ (bzw. 0% und 100%) können enthalten sein, ohne dass ein Korrekturterm verwendet werden müsste.

(2) Sie weist die für Anteile angebrachte sigmoide Grundform auf (s-förmig; symmetrisch für x und $1\text{-}x$; genau ein Wendepunkt bei $x = 0{,}5$; s. Abb. 11). Mit Annäherung an die absoluten Grenzen ‚0' und ‚1' werden die Ausgangsdaten im Zuge der Transformation zunehmend gestreckt, im Bereich um $0{,}5$ gestaucht.

(3) Vergleichbar mit der Box-Cox-Transformation kann der Transformationsparameter auf einer stetigen Skala bestimmt werden, womit die Intensität der Stauchung bzw. Streckung der Situation angepasst wird.

Die ‚*folded exponential transformation for proportions*' wurde in dieser Arbeit für Prozentzahlen (hier also die Merkmalskategorie ‚*Qualität [%]*') einbezogen und in der Wirksamkeit mit der Box-Cox-Transformation verglichen.

Die ‚*folded exponential transformation for proportions*' ist definiert durch:

$$y = f(x) = \left(e^{(\varphi x)} - e^{(\varphi(1-x))}\right)/2\varphi \qquad \text{für } \varphi > 0\,,$$

$$y = f(x) = x - \tfrac{1}{2} \qquad \text{für } \varphi = 0\,.$$

Anmerkung: Im Gegensatz zur Box-Cox-Transformation, bei der $\varphi = 1$ Nicht-Transformation bedeutet, liegt dies hier quasi bei $\varphi = 0$, allerdings erfolgt eine Skalenverschiebung in der Form, dass eine relative Häufigkeit / Wahrscheinlichkeit von $x = 0{,}5$ (unabhängig von φ) in $y = 0$ transformiert wird, für $x > 0{,}5$ ist $y > 0$, für $x < 0{,}5$ ist $y < 0$.

Rücktransformation der Serien-Mittelwerte:

$$x = \tfrac{1}{2} + (1/\varphi)\,\sinh^{-1}\left(y\varphi/e^{(\varphi/2)}\right) \qquad \text{für } \varphi > 0\,,$$

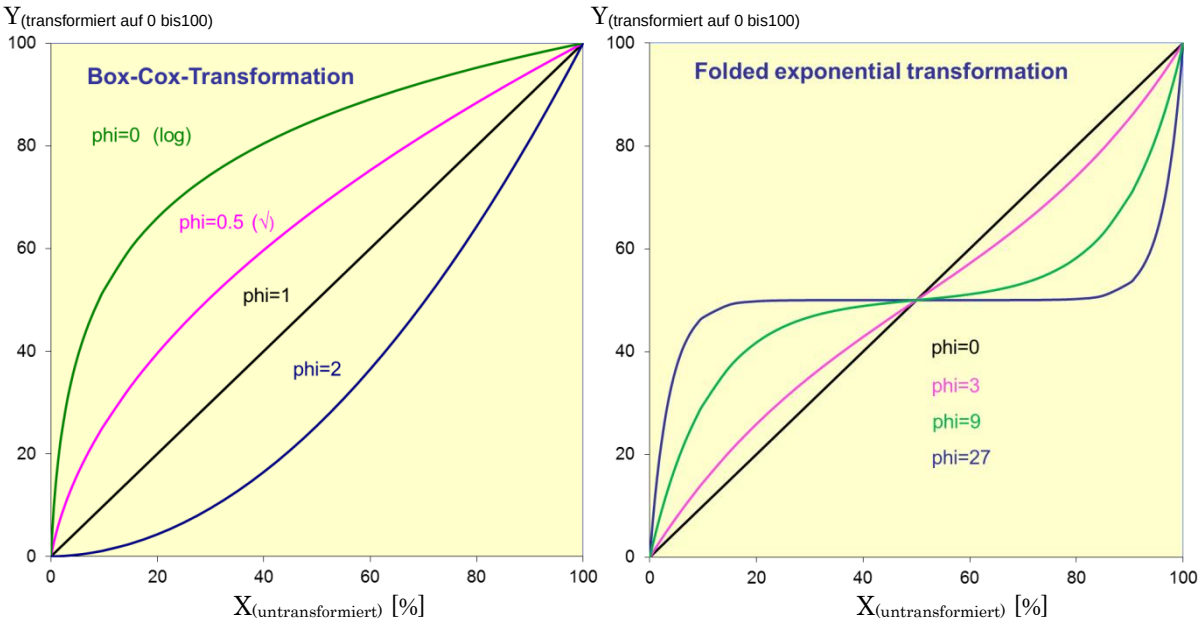
$$x = \tfrac{1}{2} + y \qquad \text{für } \varphi = 0\,.$$

Für extrem schiefe Verteilungen ist auch diese Transformation allerdings nur bedingt geeignet (Piepho, 2003). Aus Gründen der Umsetzbarkeit in der Praxis der Versuchsauswertung wurden hier dennoch keine weiteren Transformationsansätze einbezogen.

Wie bei der Box-Cox-Transformation erfolgt auch für diese Transformation die Bestimmung des bestmöglichen Transformationsparameters über die Maximierung des Log-Likelihood. Derzeit sind die Algorithmen der iterativen Schätzung des Transformationsparameters nach der ‚*folded exponential transformation*', der Hin- bzw. Rücktransformation der Ausgangs- bzw. Schätzwerte sowie deren Standardfehler noch nicht in die PIAFStat-Verfahren integriert (s. 5.4). Es konnte aber ein im Zusammenhang mit der Veröffentlichung von Piepho (2003) bereitgestelltes SAS®-Makro zur Bestimmung des optimalen Transformationsparameters genutzt werden. Ein Test gegenüber der Nicht-Transformation bzw. ein Vertrauensintervall ist in diesem Makro noch nicht enthalten, dies ist im Zuge weiterer Arbeiten vorgesehen. Der Güte-Vergleich zur Box-Cox-Transformation und zur Nicht-Transformation ist durch die Einbeziehung des *Jacobischen Terms* in die Bestimmung der Log-Likelihood aber möglich (Piepho 2003). Es sollte in Betracht gezogen werden, dass auch auf der Prozentskala (u.a. in Abhängigkeit von Lage und Verteilung der Werte eines Datensatzes) die Anwendung der Box-Cox-Transformation brauchbar sein

kann und nicht a priori der ‚folded exponential transformation' unterlegen sein muss (s. 6.1.1).

Die prinzipielle Wirkung der beiden hier genutzten Transformationsfamilien hinsichtlich Datenstreckung bzw. -stauchung in unterschiedlen Skalenabschnitten ist in Abb. 11 exemplarisch dargestellt.



ausgewählte Werte für den Transformationsparameter phi (φ)
y=f(x) auf den Skalenbereich von x transformiert (Prozent-Skala: 0% bis 100%)

Abb. 11: Stauchung und Streckung bei Transformationsfamilien ohne versus mit Wendepunkt

5.3 Das Auswertungsmodell

5.3.1 Das allgemeine gemischte Modell und die verwendete Syntax für die Formulierung der Modellgleichungen

Da in den anstehenden Auswertungen sowohl fixe als auch zufällige Faktoren zu berücksichtigen sind, kommen ausschließlich gemischte Modelle (mixed models) zum Einsatz.

Allgemeines gemischtes Modell:

In Matrixnotation lässt sich ein gemischtes Modell darstellen durch:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

wobei $\boldsymbol{y}$ der $n \times 1$ Vektor der Beobachtungen (hier Sortenmittelwerte aus Einzelversuchen), $\boldsymbol{\beta}$ der $p \times 1$ Vektor der p fixen Effekte, $\boldsymbol{u}$ der $q \times 1$ Vektor der q zufälligen Effekte, $\boldsymbol{X}$ und $\boldsymbol{Z}$ die $n \times p$ bzw. $n \times q$ Designmatrizen der fixen bzw. zufälligen Effekte und $\boldsymbol{e}$ der $n \times 1$ Vektor der Residuen (Zufallsterme) ist.

Bei den zufälligen Effekten und Residuen können unterschiedliche Varianzen und Kovarianzen existieren, sie sind durch die Matrizen $\boldsymbol{G}$ bzw. $\boldsymbol{R}$ beschrieben. Für die zufälligen Effekte und Residuen werden folgende Annahmen gemacht: $\boldsymbol{u} \sim N(\boldsymbol{O}, \boldsymbol{G})$, $\boldsymbol{e} \sim N(\boldsymbol{O}, \boldsymbol{R})$ und $Cov(\boldsymbol{u}, \boldsymbol{e}) = \boldsymbol{0}$. Danach ist $Var(\boldsymbol{y}) = \boldsymbol{ZGZ'} + \boldsymbol{R} = \boldsymbol{V}$.

Die Formulierung der Modellgleichungen (z.B. Tab. 16 und 17) erfolgt auf Basis einer von Patterson (1997) vorgeschlagenen und u.a. auch von Piepho und Michel (2001) und Piepho et al. (2003) genutzten Syntax für gemischte lineare Modelle. Diese Syntax korrespondiert direkt mit der Syntax von Computerroutinen für lineare Modelle und ist insofern für Leser im Bereich der angewandten Forschung einfacher als die Notation nach Rasch et al. (1998), in welcher Effekte mit Indizes versehen sind, und ebenso einfacher als eine Matrix-Notation.

Zur Erläuterung soll folgendes Beispiel für eine Modellgleichung (aus Tab. 16, Nr. 2):

S : O + J + O•J + T(O•J) + S•O + S•J + S•O•J dienen**.**

Der **Doppelpunkt** trennt fixe von zufälligen Effekten, wobei die fixen vor und die zufälligen Effekte nach dem Doppelpunkt stehen. Die Formulierung **T(O • J)** weist hier z.B. darauf hin, dass der Faktor Typ einen Stratifizierungsfaktor an Ort × Jahr - Kombinationen darstellt, wobei gilt: T(O • J) = T • O • J. Analog zur Syntax von Computerroutinen ist der **Resteffekt** (zufällig) in der Modellgleichung nicht explizit aufgeführt, ist aber unbedingter Bestandteil. Die Syntax (Kürzel) und Semantik der Faktoren ist in 4.3 dargelegt. Zur Darstellung des mathematischen Operators für die Multiplikation wird für Terme in der Modellgleichung der Malpunkt (•) verwendet. Zum Unterschied dazu wird für alle anderen Bedeutungen (Parameterschätzungen, Erläuterung von Faktoren-Kombinationen) das Malkreuz (×) verwendet.

Die im Weiteren dargelegte Strategie der gewichteten Auswertung über mehrere Anbaugebiete erfordert bezüglich der Definition des Faktors ‚Sorte' die Option einer differenzierten Betrachtung zwischen den Phasen Varianzkomponenten- und der

Mittelwertschätzung. Daher wird in 5.3.2 bis 5.3.4 zunächst das Basismodell zur Schätzung der Varianzkomponenten beschrieben, wobei in 5.3.2 ein kompakter Überblick über die Herleitung gegeben wird, während in 5.3.3 und 5.3.4. Besonderheiten dieses Modells vertieft werden. In 5.3.5 bis 5.3.8 werden unterschiedliche Ansätze zur Sorten-Mittelwertschätzung dargelegt.

5.3.2 Herleitung des Auswertungsmodells für die Varianzkomponentenschätzung

Als thematische Einführung wird die Frage „*Wozu Varianzkomponenten schätzen?*“ diskutiert. Das Leistungspotential einer Sorte kann nicht fehlerfrei ermittelt werden, sondern wird außer von den Einzelversuchsfehlern besonders von Wechselwirkungen mit der Umwelt beeinflusst: Die Sortenrangfolgen bzw. Sortenrelationen sind von Versuch zu Versuch nicht identisch, ebenso nicht von Jahr zu Jahr, von Ort zu Ort, von Anbaugebiet zu Anbaugebiet. Im Zuge der Schätzung der Varianzkomponenten werden die Variationsursachen der zufälligen Effekte im Auswertungsmodell analysiert und quantifiziert. Dies erlaubt den Vergleich der Wirkungsstärke zwischen den zufälligen Effekten der Modellgleichung und ermöglicht nachfolgend die geeignetste Verdichtung der Daten für die Mittelwertschätzung vor dem Hintergrund a) der Unbalanciertheit und b) der Sorte × Umwelt - Wechselwirkungen. Bezüglich der Berücksichtigung unterschiedlicher Besetzungen der Sorten in Umwelten (Jahren; Orten etc.) kann vereinfacht gesagt werden, dass ein Ergebnis aus einer zusätzlichen Umwelt einen größeren Informationsgewinn einbringt und entsprechend höher gewichtet wird, als ein Ergebnis aus einer bereits gut besetzten Umwelt, denn Ergebnisse aus dem gleichem Jahr bzw. Ort etc. sind korreliert (Möhring et al., 2004a).

Die erheblich unbalancierte, komplexe Struktur der Daten gestaltet die Auswertung der Versuchsserien kompliziert. Der Restricted Maximum Likelihood (REML) Algorithmus ist geeignet zur Verrechnung unbalancierter Datensätze (Patterson und Thompson, 1971) und wird in dieser Arbeit genutzt (s. Ausführungen zu REML in Abschn. 3).

Die Gesamtauswertung erfolgt als Zweischritt-Analyse. Wenn die Versuchsanlagen und die letztendlich jeweils genutzten Auswertungsmodelle (z.B. teilweise Einbeziehung von Bodentrends oder bei anderen Versuchen von räumlichen Korrelationen) zwischen den Einzelversuchen sehr variieren, so wäre die Modellierung der gesamten Versuchsserienauswertung in einem Schritt auf Basis der Einzelwerte u.U. zu komplex (Smith et al., 2005; Möhring und Piepho, 2009). Einzelwertbasierte Modelldurchläufe konnten als Einschritt-Analyse mit der SAS® Prozedur MIXED (SAS Institute, 1999) bei den gegebenen Bedingungen (Daten und faktorielle Strukturen) überwiegend nicht erfolgreich ablaufen. Zunächst wurde daher jeder Einzelversuch entsprechend seiner individuellen Spezifik mit einem gemischten Modell (PROC MIXED in SAS®) ausgewertet, um Sortenmittelwerte bei bestmöglich angepasstem Modell zu erhalten. Das AICC-Anpassungskriterium wurde hierbei als Optimalitätskriterium genutzt (Akaike’s Corrected Information Criterion; Burnham und Anderson, 1998). Eine Modellselektion in der Einzelversuchsauswertung erfolgt in Anlehnung an den bei Richter et al. (2007) vorgestellten Algorithmus. Der erste Schritt der Zweischrittanalyse - die Einzelversuchsauswertung - wird hier nicht weiter vertieft. Im zweiten Schritt werden die Sortenmittelwerte und deren Standardfehler (hier als SE_{Serie}; Definition s. 5.3.5) für die Gesamtserie geschätzt. Dies ist das primäre Ziel der Versuchsserienauswertung im Sortenversuchswesen und Gegenstand dieser

Arbeit. Die Versuchsserienauswertung erfolgt also ausschließlich basierend auf den geschätzten Mittelwerten aus den Einzelversuchen. Die Möglichkeit der Gewichtung der Sortenmittelwerte aus Einzelversuchen wird in 5.3.3 diskutiert.

Nachfolgend werden, ausgehend von einem einfachen Sorte × Umwelt - Modell, Schritte der Modellbildung sukzessiv hergeleitet (Tab. 16 und Tab. 17). Die einzelnen Schritte der Modellerweiterung müssen nicht zwingend in dieser Reihenfolge erfolgen. Im Ergebnis der Herleitung werden zwei Modelle als ‚*Endmodelle*' ausgewiesen, die in dieser Arbeit und in der Anwendung der Hohenheim-Gülzower-Serienauswertung für die Varianzkomponentenschätzung Verwendung finden.

Tab. 16: Auswertung für ein Anbaugebiet; schrittweise Herleitung des Auswertungsmodells

Nr.	Modellgleichung und Anmerkungen
	Vordefinitionen: – Serienauswertung basiert auf Mittelwerten aus Einzelversuchen – Umweltfaktoren werden als zufällig definiert (Aussagen für Anbaugebiet und dessen Klima)
1	Versuchsserie Sorte × Umwelt **S : Umwelt** Anmerkungen: – Umwelt synonym zu Versuch ohne Untersetzung nach Ort, Jahr und Typ; Versuch hier = T(O × J) in der Syntax lt. 4.3 – Restvarianz subsumiert S × Umwelt und $SE^2_{Versuch}$ (Fehler der Mittelwerte)
2	Umwelt wird in Ort und Jahr untersetzt **S : O + J + O•J + S•O + S•J** Anmerkung: Restvarianz subsumiert S • O • J und $SE^2_{Versuch}$
3	Aufnahme des Blockungsfaktors *Typ* (T) für Versuche an gleichen Ort × Jahr - Kombinationen **S : O + J + O•J + T(O•J) + S•O + S•J + S•O•J** Anmerkung: Restvarianz subsumiert S × T(O × J) und $SE^2_{Versuch}$
4	Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen **S : O + J + O•J + T(O•J) + S•O + S•J + S•O•J + S•T(O•J)** Anmerkungen – Die Varianzkomponente S × T(O × J) ist nun eigenständig schätzbar – gepoolte Restvarianz entspricht dem Mittelwert der $SE^2_{Versuch}$ – Vertiefung in Abschnitt 5.3.3
5	Modellvereinfachung durch Verzicht auf die Inter-Umwelt-Information **S + T(O•J) : S•O + S•J + S•O•J + S•T(O•J)** Anmerkung: – zur Schätzung von Sorten-Kontrasten können Effekte, die nicht den Faktor S enthalten, formal als fix betrachtet werden (Verzicht auf Inter-Umwelt-Information) – fixer Kombinationseffekt T(O × J) subsumiert nun alle (reinen Umwelt-) Effekte, die nicht den Faktor Sorte enthalten; also $T(O \times J)_{gesamt} = O + J + O \times J + T(O \times J)$ – S • T(O • J) nur in Modellgleichung, wenn Gewichtung entsprechend Modell 4 erfolgt

Tab. 17: Modellerweiterungen zur Auswertung für ein Zielgebiet bei gewichteter Einbeziehung von Nachbargebieten

Nr.	Modellgleichung und Anmerkungen
	Ausgangspunkt für nachfolgende Modellerweiterungen ist Modell 5 in Tab. 16
6	Modellanpassung zur Berücksichtigung der genetischen Korreliertheit der Sorteneffekte zwischen Anbaugebieten (AG) **T(O•J) + AG : S + S•O + S•J + S•J•AG + S•O•J + S•T(O•J) + S•AG** Anmerkungen: – Aufnahme des fixen Effektes AG und der zufälligen Effekte S•AG und S•J•AG – O vollständig in AG geschachtelt und eindeutig codiert, daher vertritt an allen Stellen im Modell O den dahinter stehenden Term O(AG) – Schätzung der Ähnlichkeiten zwischen Anbaugebieten (AG) erfordert: o Schätzung der Varianz des Haupteffektes Sorte (S) - daher ist S nun zufällig o sowie der S × AG - Interaktionsvarianz – S • T(O • J) nur in Modellgleichung, wenn Gewichtung entsprechend Modell 4 erfolgt – Vertiefung in Abschnitt 5.3.4
	Berücksichtigung einer Stratifizierung von Sorten innerhalb von Sortengruppen (Gr), es ergeben sich die beiden im Weiteren verwendeten **Endmodelle**:
7a	ohne Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen **T(O•J) + Gr•AG : S + S•O + S•J + S•J•AG + S•O•J + S•AG** Restvarianz subsumiert S × T(O × J) und $SE^2_{Versuch}$
7b	Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen **T(O•J) + Gr•AG: S + S•O + S•J + S•J•AG + S•O•J + S•T(O•J) + S•AG**
	Anmerkungen: – S vollständig in Gr geschachtelt und eindeutig codiert, daher vertritt an allen Stellen im Modell S den dahinter stehenden Term S(Gr) – Faktor Gr wird im fixen Kombinationseffekt AG•Gr aufgenommen; (LS Means AG*Gr als ‚Intercept' bzw. μ für die Mittelwertschätzung (s. 5.3.5) benötigt) – aus Gründen der Rechnerleistung (performance) und wegen marginaler Relevanz erfolgt derzeit noch keine Berücksichtigung eines Effektes Gr•O•J in der Modell-gleichung; Modellerweiterung vorgesehen – Aufnahme des Faktors Gr bewirkt Schätzung der Sorten-Varianz innerhalb von Gruppen, vermeidet damit eine Überschätzung der genetischen Korrelationen der Sorteneffekte zwischen Anbaugebieten – Vertiefung in Abschnitt 5.3.4 – Sofern nur eine Sortengruppe (Gr) und/oder nur ein Anbaugebiet (AG) in einem konkreten Datensatz vorkommt, könnte das Modell wieder reduziert werden. Für die Automatisierungen in der Auswertungssoftware PIAFStat (s. 5.4) wird aber für diese Faktoren dann nur eine Stufe klassifiziert und die diese Faktoren enthaltenden Varianzkomponenten werden ggf. auf ‚0' gesetzt. Dieses Vorgehen führt zu identischen Resultaten wie ein entsprechend reduziertes Modell.

Piepho und Michel (2001) haben gezeigt, dass das vierfaktorielle Modell 4 (Tab. 16) für die mehrortige mehrjährige Sortenprüfung die Realität der wichtigsten Einflussfaktoren und relevanten Varianzursachen gut widerspiegelt, wobei hier die abgestuft gewichtete Einbeziehung von benachbarten Anbaugebieten noch nicht angedacht war. Sie begründeten im Weiteren, dass alle Effekte, die nicht den Faktor Sorte enthalten, formal als fix betrachtet werden können, wenn die Schätzung der Sortenkontraste im Vordergrund steht und schlugen eine Reduktion zum Modell 5 vor. Der Umwelt-Effekt T(O • J) in der Modellgleichung subsumiert hier den Komplex der Effekte, die nicht den Faktor Sorte enthalten. Dies stellt einen Verzicht auf die Inter-Umwelt-Information dar - analog zur Inter-Block-Information bei Anlagen mit unvollständigen Blocks (s. John und Williams, 1995). Da die Umwelteffekte i.d.R. sehr dominant gegenüber den mit ‚Sorte' gekreuzten Effekten sind, führt dies aber zu keinem relevanten Informationsverlust. Diese Modellvereinfachung reduzierte die Rechenzeiten für große, komplex strukturierte Datensätze erheblich (z.T. um mehrere Stunden) bzw. ermöglichte erst einen erfolgreichen Durchlauf. Beide Modelle 4 und 5 liefern korrekte Standardfehler der Mittelwertdifferenzen. Sie implizieren eine relativ einfache Varianz-Kovarianz-Struktur der Daten. Modell 5 wurde seit 2000 bis zur Einführung der Erweiterungen lt. Tab. 17 vom Autor für die regionale Auswertung der Sortenversuche in Mecklenburg-Vorpommern eingesetzt.

Aufgrund ihrer fachlichen Bedeutung erfolgt für ausgewählte, in Tab. 16 und Tab. 17 kurz angerissene Modellschritte eine vertiefte Diskussion in eigenständigen Gliederungspunkten 5.3.3 (Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen) und 5.3.4 (Schätzung der genetischen Korrelationen zwischen Anbaugebieten).

Die Schätzung der Varianzkomponenten erfolgt mit der BOUND Option der REML Prozedur, welche nur nichtnegative Schätzwerte zulässt. Dies kann zu etwas überschätzten Varianzkomponenten führen (Laidig et al., 2008).

5.3.3 Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen

In Tab. 16. und Tab. 17 ist ab Modell 4 die Option vorgesehen, in der Serienauswertung eine Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen zu nutzen. Dies wird im Weiteren diskutiert und in der Umsetzung vorgestellt. Die grundsätzliche Problematik des Informationsverlustes in einer Zweischritt- gegenüber einer Einschritt-Analyse wurde in 3. diskutiert.

Für die sinnvolle Berücksichtigung der Sorte × Umwelt - Interaktionen in der Serienauswertung ist Voraussetzung, dass diese Interaktionen nur unwesentlich durch systematische oder zufällige Fehler in Einzelversuchen begründet sind (Bätz, 1984). Auf dieses Problem wies auch Roemer (1930) bereits hin. Vor diesem Hintergrund forderten Bätz et al. (1985), dass in Einzelversuchen beim Prüfmerkmal *Ertrag* der Variationskoeffizient höchstens bei 4% bis 5% liegen sollte. Der Autor dieser Arbeit plädiert allerdings nicht für eine starre Entscheidungs-Grenze, ab welcher ein Versuch komplett gestrichen wird, während ein ähnlich problematischer Versuch knapp unter der Grenze mit vollem Gewicht in die Serienauswertung einfließen würde. Gegen eine willkürliche starre Grenze spricht m.E. auch die relativ kontinuierliche, nicht abgestuft-diskrete Verteilung der Variationskoeffizienten bzw. der Standardfehler der Mittelwerte über die einbezogenen Versuche einer großen Serie

(Abb. 2 und Abb. 29 in 6.2.3). Insofern sollte eine kontinuierlich abgestufte, objektivierte Gewichtung förderlicher sein.

Für Merkmale, bei denen in den Einzelversuchen - basierend auf einem Modell mit fixen Sorteneffekten - Standardfehler der Mittelwerte ($SE_{Versuch}$) geschätzt werden konnten, ist in Abb. 2 die jeweilige Verteilung der Variationskoeffizienten der Mittelwerte dargestellt.

Im Weiteren wird der Terminus *Einzelversuchsfehler* auch für die Fälle synonym verwendet, bei denen innerhalb von Einzelversuchen differenzierte Mittelwert spezifische Standardfehler vorlagen, welche dann auch Mittelwert- und nicht Versuchspezifisch verwendet wurden.

Es ergibt sich die Frage, wie die Mittelwerte aus dem ersten Auswertungsschritt (Einzelversuchsauswertung) im zweiten Auswertungsschritt (Serienanalyse) in Abhängigkeit von ihrer jeweiligen Präzision gewichtet werden sollten. Wären die Versuche ausnahmslos als randomisierte vollständige Blockanlagen angelegt und basierend auf einem Modell mit festen Blockeffekten ausgewertet worden, so könnte eine exakte Gewichtung mittels der Kehrwerte ihrer Fehlervarianz ($1/SE^2$) erfolgen (Frensham et al., 1997; Möhring und Piepho, 2009). Da aber Versuche auch mit unvollständigen Blocks angelegt wurden, nicht selten ungeplante Fehlwerte auftraten oder auch Auswertungen basierend auf Modellen mit zufälligen Blocks und z.T. unter Berücksichtig geostatistischer Methoden erfolgten, wäre diese Gewichtung nicht exakt. Sie vernachlässigt die durch räumliche Ansätze oder zufällige Blocks verursachten Kovarianzen. Daher wären die Standardfehler der Mittelwerte im Sinne der Gewichtung überhöht. Erzielte Vorteilseffekte durch die geeignete Einzelversuchs-Anlage und -Auswertung würden sich unzureichend in der Gewichtung niederschlagen.

In dieser Arbeit erfolgte die Gewichtung auf Basis von durch Piepho und Möhring (2007) sowie Piepho und Möhring (2009) vorgestellten Methoden und dem in diesem Zuge durch Möhring bereitgestellten SAS-Makro. Drei dieser Methoden wurden für die Berücksichtigung von zufälligen Effekten im Auswertungsmodell und von Unbalanciertheit von den Autoren favorisiert. Bei diesen Methoden wird die Varianz-Kovarianz-Matrix $\boldsymbol{V}_{MW}$ der t adjustierten Sorten-Mittelwerte eines Einzelversuches durch eine Diagonal-Matrix $\boldsymbol{R}_{MW}$ in der Weise approximiert, dass die Varianzen der Differenzen die Summe der Abweichungsquadrate (S) zur assoziierten $\boldsymbol{V}_{MW}$ - Matrix minimieren (der Zusatz MW erfolgt zur Abgrenzung der Bedeutung, die $\boldsymbol{V}$ und $\boldsymbol{R}$ in der allgemeinen Schreibweise des gemischten Modells haben; $\boldsymbol{V}_{MW}$ und $\boldsymbol{R}_{MW}$ werden mittels der COVB Option in der PROC MIXED in SAS® erzeugt). Die Abwandlung in eine Diagonalmatrix bringt den Vorteil einer erheblichen Erhöhung der Rechnerleistung bei der REML-Analyse. Die Elemente der Diagonale von $\boldsymbol{R}_{MW}$ sind also so bestimmt, dass

$$S = \sum_{i=1}^{t} \sum_{i' \neq i} [(r_i + r_{i'}) - (v_{ii} + v_{i'i'} - 2v_{ii'})]^2$$

minimiert wird, wobei gilt:

$\boldsymbol{i}, \boldsymbol{i}'$	indizieren unterschiedliche adjustierte Sortenmittelwerte
t	Anzahl adjustierte Sortenmittelwerte; indiziert ‚letzten' Sortenmittelwert des Vektors der Sortenmittelwerte

$\boldsymbol{V_{MW}} = \begin{pmatrix} v_{1,1} & \cdots & v_{1,t} \\ \vdots & \ddots & \vdots \\ v_{t,1} & \cdots & v_{t,t} \end{pmatrix}$ geschätzte Varianz-Kovarianz-Matrix aller adjustierten Sortenmittelwerte des Einzelversuches

$v_{ii}\ bzw.\ v_{i'i'}$ geschätzte Fehlervarianzen der Sortenmittelwerte i und i'; Elemente der Hauptdiagonalen in $\boldsymbol{V_{MW}}$

$v_{ii'}$ geschätzte Kovarianz zweier unterschiedlicher Sortenmittelwerte i und i' in $\boldsymbol{V_{MW}}$

$v_{ii} + v_{i'i'} - 2v_{ii'}$ aus geschätzter $\boldsymbol{V_{MW}}$ abgeleitete Varianz der Differenz zweier Sortenmittelwerte i und i'

$\boldsymbol{R_{MW}} = \begin{pmatrix} r_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & r_t \end{pmatrix}$ approximierte, zu $\boldsymbol{V_{MW}}$ assoziierte Diagonalmatrix; Approximation unter der Bedingung der Minimierung von S (s.o.)

$r_i\ bzw.\ r_{i'}$ Elemente der Hauptdiagonalen in $\boldsymbol{R_{MW}}$; unter Nutzung der Informationen der Kovarianzen (in $\boldsymbol{V_{MW}}$) approximierte Fehlervarianzen der Sortenmittelwerte i und i';

$r_i + r_{i'}$ aus approximierter $\boldsymbol{R_{MW}}$ abgeleitete Varianz der Differenz zweier Sortenmittelwerte i und i'; unter Nutzung der Informationen der Kovarianzen (in $\boldsymbol{V_{MW}}$)

Die durch $(v_{ii} + v_{i'i'} - 2v_{ii'})$ in der $\boldsymbol{V_{MW}}$ - Matrix gegebene Varianz der Differenz zweier Sortenmittelwerte ist gut approximiert durch $(r_i + r_{i'})$ in der Diagonal-Matrix $\boldsymbol{R_{MW}}$. Zur Bestimmung der Gewichte der adjustierten Sorten-Mittelwerte werden die Elemente r_i der Diagonalen herangezogen und im Weiteren je Sorte als $SE^2_{Versuch}$ bezeichnet. Im Auswertungsmodell des Einzelversuches berücksichtigte zufällige Störgrößen gehen bei diesem Ansatz mit ein, ebenso wird Unbalanciertheit berücksichtigt. Die algorithmischen Unterschiede zwischen den drei vorgeschlagenen Methoden führten nach Piepho und Möhring (2009) in der Auswertung von vier Versuchsserien-Datensätzen und in Simulationen nur zu vernachlässigbar kleinen Unterschieden in den Serienauswertungen und die Unterschiede ließen keine generelle Tendenz erkennen, sondern hingen vom einzelnen Datensatz ab. Daher werden die theoretischen, eher marginalen Unterschiede zwischen diesen Methoden im Weiteren nicht vertieft, sondern es wird hierzu auf Piepho und Möhring (2009) verwiesen. Im Sinne eines robusten Risikoausgleichs wurden in den in dieser Arbeit einbezogenen Versuchen für alle drei Methoden die $SE^2_{Versuch}$ ermittelt und anschließend von diesen das arithmetische Mittel gebildet und in der Serienauswertung verwendet.

Jüngere Methoden, die die angemessene Einschritt-Analyse vollständig reproduzieren (Piepho et al., 2012), konnten noch keine Berücksichtigung in der Softwarelösung der Hohenheim-Gülzower-Serienauswertung finden. Sie basieren nicht auf je einem Standardfehler je Sortenmittelwert, sondern auf der Einbeziehung der Varianz-Kovarianz-Matrizen der adjustierten Mittelwerte aus Einzelversuchen.

Die $SE_{Versuch}$ werden je Sorte im Zuge der Einzelversuchsauswertung mit PIAFStat-Verfahren (s. 5.4) ermittelt und gemeinsam mit dem zugehörigen adjustierten Sortenmittelwert in der PIAF-Datenbank gespeichert.

In der Serienauswertung erhalten die einzelnen Sortenmittelwerte aus Einzelversuchen als Gewicht (w) den Kehrwert des quadrierten $SE_{Versuch}$:

$$w = 1/SE^2_{Versuch}$$

Bei Nutzung eines Auswertungsmodells *mit Gewichtung* (Tab. 16 Modell 4 ff; Tab. 17 End-Modell 7b) entspricht die gepoolte Restvarianz dem Mittelwert aller $SE^2_{Versuch}$. Die Varianzkomponente S × T(O × J) ist dann eigenständig schätzbar und wird als Effekt in das Modell aufgenommen.

Kann oder soll keine Gewichtung erfolgen, so ist die Varianzkomponente S × T(O × J) nicht eigenständig schätzbar und wird nicht im Modell berücksichtigt. In diesem Fall subsumiert die ausgegebene Restvarianz die Varianz S × T(O × J) und die Fehlervarianz der Mittelwerte (Tab. 16 Modell 3; Tab. 17 End-Modell 7a).

5.3.4 Modellanpassungen zur Berücksichtigung der genetischen Korreliertheit der Sorteneffekte zwischen Anbaugebieten

Im Jahre 2004 wurde das Modell 5 (Tab. 16) erweitert, um das Regionalisierungskonzept (s. 5.1.3) der Einbeziehung von Nachbargebieten mit abgestufter Gewichtung umzusetzen. Methodisch erfolgte dies im Wesentlichen in Kooperation der Universität Hohenheim und der LFA in Gülzow. Im Zusammenhang mit der Einarbeitung dieser Modelle in die Anwendersoftware PIAFStat (s. 5.4) erhielten diese Auswertungsverfahren (Tab. 17; Modelle 7a und 7b) die Sammelbezeichnung *Hohenheim-Gülzower-Serienauswertung*.

Methodische Grundlage dafür waren Studien zur regionalisierten Auswertung (Möhring et al., 2004 a, b und c). Im Wesentlichen erfolgten zwei grundlegende Modellerweiterungen gegenüber Modell 5 (Tab. 16):

a) Aufnahme des Effektes Anbaugebiet (AG) und Schätzung der *genetischen Kovarianzmatrix* $\sum_S$ (Erläuterung von $\sum_S$ s.u.)
b) Aufnahme des Faktors Sortengruppe (Gr)

ad a) Aufnahme des Effektes Anbaugebiet (AG) und Schätzung der *genetischen Kovarianzmatrix* → **Modell 6** in Tab. 17

Die Aufnahme des Effektes *Anbaugebiet* erfolgt in der Weise, dass *Orte* vollständig in übergeordneten *Anbaugebieten* geschachtelt sind (O(AG)). Auf die Modellformulierung bei den fixen Effekten braucht diese Stratifizierung keinen Einfluss haben, wenn Sorten- und alle mit ‚Sorte' gekreuzten Effekte im Fokus stehen. Auch auf der Seite der zufälligen Effekte kann ‚O' aufgrund der eindeutigen Schachtelung und Kodierung an allen Stellen ‚O(AG)' ersetzen. Entscheidend ist die Aufnahme der Effekte S • J • AG und S • AG in der Modellgleichung.

Die Schätzung einer genetischen Kovarianzmatrix wird als Grundlage genutzt, um die Gewichtungsabstufung der einbezogenen Anbaugebiete zu objektivieren. Komponenten in dieser Kovarianzmatrix sind die genetische Sortenvarianz (S) für Haupteffekte und die Sorte × Anbaugebiet – Varianzen (S × AG) für Interaktionseffekte. Die Schätzung dieser Varianzen erfordert zunächst die Zuordnung des Effektes *Sorte* als zufällig - unabhängig davon, ob letztlich die Sortenmittelwert-Bestimmung als ‚*best linear unbiased prediction'* (BLUP; Sorte zufällig) erfolgen soll oder ob der Faktor Sorte dann als fix betrachtet wird (s. 6.4).

Die Varianz-Kovarianz-Matrix kann in unterschiedlicher Struktur definiert werden.

Im Weiteren werden nur Ansätze verfolgt, die eine konstante (homogene) Kovarianz σ_S^2 zwischen zwei Anbaugebieten über alle r Anbaugebiete annehmen:

Es ergibt sich dann folgende genetische Varianz-Kovarianz-Matrix $\sum_S$:

$$\sum_S = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \sigma_S^2 + \begin{pmatrix} \sigma_{S\times AG_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{S\times AG_r}^2 \end{pmatrix},$$

wobei gilt:

$\sigma_S^2 = Varianz\ des\ Haupteffektes\ Sorte; Kovarianz\ zwischen\ zwei\ Anbaugebieten$

$\sigma_{S\times AG_j}^2 = Varianz\ des\ Interaktionseffektes\ S \times AG_j\ \ f\ddot{u}r\ ein\ Anbaugebiet\ j\ \ (j = 1 \ldots r)$ $mit\ r\ Anbaugebieten$

$F\ddot{u}r\ geschätzte\ Varianzen\ wird\ in\ dieser\ Arbeit\ das\ Malkreuz\ "\times"\ verwendet,$ $im\ Unterschied\ zum\ Malpunkt\ "\bullet"\ f\ddot{u}r\ Effekte\ in\ Modellgleichungen\ (Syntax\ s.\ 5.3.1).$

Der einfachste hier verwendete Ansatz legt darüber hinaus über alle Anbaugebiete eine homogene (S × AG) - Varianz- / Kovarianz-Struktur zugrunde (homogene Varianz des Haupteffektes Sorte und der Interaktionsvarianzen S × AG). Daraus ergibt sich in diesem Fall folgende zusätzliche Vereinfachung:

$\sigma_{S\times AG_1}^2 = \sigma_{S\times AG_2}^2 = \cdots = \sigma_{S\times AG_r}^2 = \sigma_{S\times AG}^2$.

Damit sind auch alle Korrelationen der Sorteneffekte zwischen jeweils zwei Anbaugebieten als konstant definiert. Dies ist eine sehr starke Annahme. Sie wird im Rahmen dieser Arbeit nur verwendet, wenn ein heterogenes Modell für S × AG - Varianzen nicht erfolgreich abläuft (Modellabbruch oder keine Konvergenz) bzw. wenn keine hinreichende Schätzgenauigkeit für die (S × AG) - Varianzen erzielbar war (z.B. wenn alle (S × AG) - Varianzen auf null gesetzt wurden). Tab. 18 zeigt, dass dieser compound symmetry - Ansatz regelmäßig nur bei den kleinen Datensätzen verwendet wurde - dies ist ein Indiz dafür, dass diese Interaktions-Varianzen ihrer Natur nach eher spezifisch je Anbaugebiet sind, dass aber bei Datensätzen, die für eine differenzierte Schätzung nicht hinreichend sind, die Vorgabe einer zunehmend einfachen Varianz-Kovarianz-Struktur pragmatisch sinnvoll bzw. erforderlich sein kann.

Die Umsetzung dieses Ansatzes erfolgt in SAS® mit der Option *CS* (CS für die c*ompound symmetry* - Struktur der Varianzheterogenität). In der Standard-Ausgabe der Varianzkomponenten durch das PIAFStat-Verfahren „VK“ (s. 5.4) wird die Varianz σ_S^2 und die Varianz $\sigma_{S\times AG}^2$ bereitgestellt.

Im Modellansatz mit heterogenen Interaktionsvarianzen wird für den Effekt S • AG in der Modellgleichung Varianzheterogenität zwischen den Anbaugebieten zugelassen, d.h. für jedes Anbaugebiet wird eine eigene (S × AG) - Varianz geschätzt. Es wird aber noch von konstanter Sorten-Varianz, also von homogener Kovarianz zwischen den Anbaugebieten ausgegangen. Die Modellierung erfolgt in SAS® mit der Option *UN(1)* (UN für unstructered). In der Standard-Ausgabe der Varianzkomponenten durch das PIAFStat-Verfahren „VK“ (s. 5.4 und Anlage 6) werden die Varianz σ_S^2 und alle Elemente der Matrix $\begin{pmatrix} \sigma_{S\times AG_1}^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{S\times AG_r}^2 \end{pmatrix}$ (‚rechter‘ Summand der oben definierten Varianz-Kovarianz-Matrix $\sum_S$) bereitgestellt.

Dieses heterogene Modell ist der am häufigsten verwendete Ansatz in den Versuchsserienauswertungen der LFA und wurde entsprechend auch im Rahmen dieser Arbeit überwiegend verwendet (Tab. 18).

Tab. 18: Modellwahl für die genetische Kovarianzmatrix in unterschiedlich besetzten Datensätzen

Planzenart (s. Tab. 2)	**Anzahl Merkmale** (s. Tab. 3)	**Anzahl Werte je Merkmal** (vgl. Abb. 1 und 3)		**Modell für die genetische Kovarianzmatrix**
		Min	**Max**	
Winterweizen	2	9104	9993	‚unstructered'; Typ UN(1) heterogene (S × AG) - Varianzen
Silomais	4	7147	7302	
Wintergerste	2	5656	5973	
Triticale	2	3110	3172	
Roggen	2	3051	3098	
Körnermais	2	1410	1410	
Hafer	1	1343		
Erbse	2	1124	1591	
Sommerbraugerste	6	764	2879	
Raps	1 (ERTR86DT)	6314		
	3	1972	4253	‚compound symmetry'; CS homogene (S × AG) - Varianzen
Kartoffeln	2	1373	1373	
Lupine	3	582	805	
Sommerfuttergerste	2	346	384	
Sommerweizen	2	338	346	

Die genetische Korreliertheit der Sorteneffekte zwischen Anbaugebieten ist durch die Varianz-Kovarianz-Matrix $\sum_S$ definiert. Als leichter bewert- und vergleichbare Zusatzinformation wird durch die bereitgestellte Anwendersoftware (s. 5.4) ein Ähnlichkeitsparameter in der Dimension eines Korrelationskoeffizienten zur Information bereitgestellt. Dieser wird nachfolgend als genetische Korrelation r der Sorteneffekte zwischen zwei Anbaugebieten ($\mathrm{AG_j}, \mathrm{AG_{j'}}$) bezeichnet und ergibt sich wie folgt:

$$r = \frac{\sigma_S^2}{\sqrt{\left(\sigma_S^2+\sigma_{S\times AG_j}^2\right)\times\left(\sigma_S^2+\sigma_{S\times AG_{j'}}^2\right)}} \quad ,$$

wobei gilt:

$\sigma_S^2 \quad = Varianz\ des\ Haupteffektes\ Sorte$

$\sigma_{S\times AG_j}^2 = Varianz\ des\ Interaktionseffektes\ S\ \times AG_j \quad (f\ddot{u}r\ j = 1, 2, \ldots r \quad und \quad j \neq j')$

Diese Betrachtung leitet bereits zu 5.3.5 über. Es muss betont werden, dass die genetischen Korrelationen häufig deutlich größer sind, als die phänotypisch beobachteten. Phänotypisch beobachtete Sorte × Anbaugebiet - Effekte führen erfahrungsgemäß oft zu einer subjektiven Überschätzung der Anbaugebietsspezifik.

ad b) Aufnahme des Faktors Sortengruppe (Gr) → **Modell 7a** und **7b** in Tab. 17

In vielen Datensätzen werden Sortengruppen gemeinsam ausgewertet, die z.B. bei einer Ertragsauswertung nicht zur gleichen Vergleichskategorie gehören (s. 4.3.1). Dies führt zu einer Überschätzung der Sortenvarianz mit der Konsequenz, dass die genetischen Korrelationen (s.u.) überschätzt und in Folge Nachbargebiete übergewichtet würden. Um dies zu vermeiden, werden die genetischen Varianzen (Sortenvarianzen) innerhalb von interpretationsrelevanten Sortengruppen entsprechend den Gruppierungen lt. 4.3.1 geschätzt. Es ergeben sich auf Grund der vollständigen Schachtelung und Kodierung der Sorten keine weiteren mit Gruppe gekreuzten zufälligen Effekte. Daher kann an allen Stellen im Modell ‚S(Gr)' durch ‚S' vertreten werden. Die Aufnahme des fixen Effektes ‚Gr' ist erforderlich.

Im Falle der Schätzung durch BLUP (best linear unbiased prediction) bewirkt die Einbeziehung des Faktors Gruppe, dass die Schrumpfungsschätzung nicht zum Mittel aller Sorten, sondern zum jeweiligen Gruppenmittel hin erfolgt.

Unter Berücksichtigung der zeitlichen und rechentechnischen Ressourcen der Prozedur Mixed unter SAS® (1999) wurden im Rahmen einer Studie (Möhring et al., 2004b) fachlich und statistisch interessante Erweiterungen des zugrunde liegenden Modells wie heterogene Sorte × Ort - Varianz und komplexe Sorte × Anbaugebiet - Varianz - Kovarianz - Strukturen untersucht. Es zeigte sich, dass eine weitere Untersetzung der Varianz - Kovarianz - Strukturen mit SAS® auf heutigen Standardrechnern bei den hier vorliegenden sehr umfangreichen Datensätzen und komplexen Modellen kaum realisierbar ist. Diese Ansätze wurden daher auch im Rahmen dieser Arbeit nicht weiter verfolgt.

5.3.5 Gewichtung von Zielgebiet und Nachbargebieten bei der Bestimmung der Sortenmittelwerte und ihrer Genauigkeit - Einführung

Der methodische Ansatz des Konzeptes von Zielgebiet und Einbeziehung von Nachbargebieten besteht darin, dass derjenige Bereich gemeinsam ausgewertet wird, der die Vorhersagegenauigkeit der Differenz von Sortenmittelwerten für das Zielgebiet maximiert (s. 5.1.3). Hierbei treten zwei Tendenzen auf. Wertet man mehrere Anbaugebiete gemeinsam aus, so erhöht sich die Anzahl der Versuche und die Genauigkeit für diesen Großraum steigt, Varianzen und zufällige Effekte können besser geschätzt und ggf. bereinigt werden. Andererseits führt die Einbeziehung von Versuchen aus Nachbargebieten zu einer gewissen Verzerrung der Schätzwerte für das Zielgebiet, da sich im Nachbargebiet möglicherweise eine andere Sortenrangfolge zeigt als im Zielgebiet. Die Verzerrung beruht dann auf regionalspezifischen Reaktionen von Sorten und steht damit in unmittelbarer Beziehung zur Varianzkomponente S×AG. Ist die Verzerrung größer als der Genauigkeitsgewinn durch die erhöhte Versuchsanzahl, so wäre die Einbeziehung der benachbarten Anbaugebiete kontraproduktiv. Ist die Verzerrung dagegen kleiner, so ist die Aussage aus den gemeinsam ausge-

werteten Anbaugebieten trotz einer gewissen Verzerrung besser als die alleinige Auswertung des Zielgebietes mit zu wenigen Versuchen.

Flexibler als eine Entweder-oder-Entscheidung über die Einbeziehung von Nachbargebieten ist eine gewichtete Auswertung. Versuche aus Anbaugebieten mit hoher Korrelation zum Zielgebiet erhalten ein höheres Gewicht als Versuche aus Anbaugebieten mit niedrigerer Korrelation zum Zielgebiet (Möhring et al., 2004a; c). Je nachdem wie hoch die genetische Korrelation zwischen den Anbaugebieten ist (Basis ist die genetische Varianz-Kovarianz-Matrix $\sum_S$, s. 5.3.4) werden die Daten aus den Nachbargebieten für die Vorhersage der Sortenleistung im Zielgebiet in unterschiedlichem Maße gewichtet. Wird im Sortenwesen von der Schätzung der Sortenleistungen gesprochen, so impliziert dies das eigentliche Ziel, Sorten miteinander zu vergleichen. Die isolierte Beurteilung der Leistung *einer* Sorte ist gegenstandslos, deren Leistungsschätzung erlangt erst durch den Vergleich zu anderen Sorten einen Nutzen. Insofern stehen Sortenkontraste im Fokus und das Optimierungskriterium sollte nicht die Minimierung der Fehlervarianzen von Mittelwerten, sondern von Mittelwertdifferenzen sein. Die Gewichte sollen daher im Folgenden so bestimmt werden, dass die mittlere Fehlervarianz der Vorhersage einer Differenz zweier Sorten, hier definiert als Funktion von Verzerrung der Differenz und Varianz der Differenz und im Weiteren als MSEP (mean squared error of prediction) bezeichnet, für ein Zielgebiet minimiert wird. Vereinfacht gilt:

$$MSEP = Verzerrung^2 + Varianz$$

Die Bestimmung des MSEP bei Annahme ‚fix' versus ‚zufällig' für den Faktor Sorte wird in den nachfolgenden Abschnitten hergeleitet.

Worin liegen die Konsequenzen der großräumigeren Betrachtung für die Nutzer der Ergebnisse? Wenn Informationen aus benachbarten Anbaugebieten herangezogen werden, so steigt im Mittel der Abstand (geografisch bzw. standortkundlich) zum Betrieb eines Landwirts. Das könnte zu dem Schluss führen, dass die Verlässlichkeit der Sortenempfehlung für den konkreten Betrieb deswegen geringer ist. Dies ist aber eine Fehleinschätzung, die aus der häufig zu beobachtenden intuitiven Überschätzung des oben beschriebenen Verzerrungseinflusses und Unterschätzung des Genauigkeitseinflusses herrührt. Die Aussagekraft eines einzelnen Versuches wäre selbst dann gering, wenn dieser auf Flächen des eigenen Betriebes liegen würde. Neben der selbst innerhalb kleiner Räume relativ hohen Differenziertheit der Böden sind insbesondere die unvorhersehbaren Jahresbedingungen (bezüglich der anstehenden Sortenwahl hier speziell die des Folgejahres) dafür ausschlaggebend, dass aussagekräftige Sorteneinschätzungen sich erst durch umfangreiche Versuchsserien über Orte und Jahre erarbeiten lassen.

Unterschiedliche Herangehensweisen zur Bestimmung der Gewichte von Anbaugebieten in Abhängigkeit vom Regionalisierungskonzept sind in Tab. 19 definiert.

Tab. 19: Bestimmung der Gewichte für Ziel- und Nachbargebiete in Abhängigkeit vom Regionalisierungskonzept

Regionalisierungskonzept	**Bestimmung der Gewichte / Anmerkungen**
1. Nur Daten des Zielgebietes fließen ein	$w_1 = 1,\ w_2 = w_3 = \cdots = w_r = 0$
2. Gleiche Gewichte für alle Gebiete (Großraumkonzept)	$w_1 = w_2 = w_3 = \cdots = w_r = 1/r$
3. Abgestufte Gewichte für alle Gebiete zur Minimierung des MSEP für das Zielgebiet	(a) $\sum_r w_r \leq 1$; Sorte zufällig, Schrumpfungsschätzung für S und S×AG-Effekte
	(b) $\sum_r w_r = 1$; Sorte fix, keine Schrumpfungsschätzung für S und S×AG-Effekte

w_1: Gewicht des Zielgebiets; w_2, w_3, ... w_r: Gewichte der Nachbargebiete; r = Anzahl der AG

Die im Ansatz suboptimalen Extremvarianten bestehen darin, dass entweder die Verzerrung ausgeschlossen wird, indem ausschließlich das Zielgebiet einfließt (Konzept 1) oder andererseits die Varianz minimiert wird, indem ein größerer Raum als das Zielgebiet gebildet wird (Konzept 2). Bereits Curnow (1988) und Atlin et al. (2000) zeigten, dass die Information aus Nachbargebieten einen Beitrag für das Zielgebiet liefern können.

Zwischen diesen Extremen liegt der Optimierungsansatz des Konzeptes 3, bei dem der MSEP als Funktion von Verzerrung und Varianz, minimiert wird. Die Gewichte hängen dann maßgeblich mit den genetischen Korrelationen zwischen Anbaugebieten zusammen (s. 6.3.1). Je geringer die Korrelation zwischen dem Ziel- und einem Nachbargebiet ist, desto geringer werden die Daten dieses Nachbargebietes gewichtet. Zusätzlich werden die Gewichte auch durch die Anzahl und Struktur der Daten in unterschiedlichen Gebieten beeinflusst, die z.T. zwischen den Sorten sehr unterschiedlich ausfällt (s. Tab. 36). Es erfolgt für jede Sorte in Abhängigkeit von ihren konkreten Prüfungsumfängen in Regionen/Jahren etc. eine adäquate Gewichtungsverteilung.

Für den grundsätzlichen Optimierungsansatz nach Konzept 3 gibt es wiederum zwei Ansätze in Abhängigkeit davon, ob der Faktor Sorte als zufällig oder als fix betrachtet wird. Diese Ansätze werden nachfolgend dargelegt. Die softwaremäßige Umsetzung erfolgt durch SAS®-Makros (Möhring et al., 2004b), welche in Auswertungsverfahren des Softwareproduktes PIAFStat integriert wurden (s. 5.4).

5.3.6 Vorhersage gewichteter Sortenmittelwerte bei Betrachtung des Faktors Sorte als zufällig - in einem Modellschritt

Können die geprüften Sorten als (quasi-zufällige) Stichprobe eines Sortenpools mit gemeinsamer Verteilung angesehen werden, z.B. zu Beginn der Selektion in einer Kreuzungsnachkommenschaft, so kommt man zu einer besten Vorhersage (minimaler Vorhersagefehler) durch BLUP (best linear unbiased prediction) für die zufälligen Effekte S und S × AG. Die Bestimmung der Lösungen für diese Effekte sowie die Schätzung fixer Effekte erfolgt mit den Modellgleichungen 7a bzw. 7b, mit denen auch die Varianzkomponenten geschätzt wurden (Tab. 17):

7a: T(O•J) + Gr•AG : S + S•O + S•J + S•J•AG + S•O•J + S•AG
ohne Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen

7b: T(O•J) + Gr•AG : S + S•O + S•J + S•J•AG + S•O•J + S•T(O•J) + S•AG
Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen

Ein gewichteter Mittelwert ($\bar{y}_{iz}$) für Sorte i im Zielgebiet (ag_z) wird im Nachgang zum Modelldurchlauf der SAS®-Prozedur MIXED wie folgt berechnet:

$$\bar{y}_{iz} = \hat{\mu}_z + \hat{s}_i + (\widehat{s \times ag})_{iz} \quad ,$$

wobei das allgemeine Mittel (μ_z) durch LS-Mittelwerte für diejenige Sortengruppe, zu der die Sorte i gehört, im Zielgebiet z definiert ist (LS Means Gr × AG). Die Lösungen für die zufälligen Effekte s_i und $(s \times ag)_{iz}$ werden als ‚solutions' ebenfalls durch die SAS®-Prozedur bereitgestellt.

Piepho und Möhring (2005) zeigen, dass dieser Ansatz die Ergebnisse aus Nachbargebieten quasi-automatisch in der Weise gewichtet, dass die MSEP für Sortendifferenzen im Zielgebiet minimiert werden, sofern die Varianzkomponenten bekannt sind. Real wird allerdings mit geschätzten Varianzkomponenten gearbeitet, so dass die geschätzten Sortenmittelwerte empirische gewichtete LS-Mittelwerte darstellen. Daher empfehlen Piepho und Möhring (2005) möglichst langjährige Datensätze sowie mehr als zwei Orte je Anbaugebiet. Um die zugrunde liegenden Varianzkomponenten so gut wie möglich zu schätzen, erfolgt dies in dieser Arbeit an langjährigen Datensätzen (mindestens neun Jahre). Aus diesem Grund und um bei der Mittelwertschätzung die Rechenzeit zu verringern, wird die Varianzkomponentenschätzung zeitlich losgelöst von der Mittelwertschätzung durchgeführt. Die Mittelwertschätzung erfolgt eigenständig auf Basis der in dieser Weise ermittelten Varianzkomponenten an einem i.d.R. nur sechsjährigen Datensatz.

Außer der Gewichtung von Ergebnissen aus Nachbargebieten beinhaltet die geschilderte Prozedur als markante statistische Charakteristik die Schrumpfung. Schrumpfung ist eine grundlegende Eigenschaft von BLUP. Gegenüber dem fixen Ansatz für Sorteneffekte werden zufällige Sorteneffekte in der hier dargestellten Datenstruktur zum Mittel aller Sorten der Sortengruppe ‚geschrumpft' - in Abhängigkeit von der Relation zwischen den Varianzkomponenten mehr oder weniger stark. Entsprechend erfolgt auch für den zufälligen Effekt S × AG eine Schrumpfungsschätzung, so dass Anbaugebietsspezifik von Sorten ‚gedämpft' wird. Im balancierten Fall ändern sich dadurch nicht die Rangfolgen, in den hier vorliegenden Fällen erheblicher Unbalanciertheit können sich durch Schrumpfung aber nicht nur die Beträge der geschätzten Sortenmittelwerte und deren Differenzen, sondern auch die Rangfolgen ändern.

5.3.7 Vorhersage gewichteter Sortenmittelwerte bei Betrachtung des Faktors Sorte als zufällig - in zwei Modellschritten

Piepho und Möhring (2005) beschreiben einen alternativen Algorithmus, welcher bei balancierten Daten zu einem identischen Ergebnis führt. Die Varianzkomponenten werden nach wie vor durch die Modellgleichungen 7a bzw. 7b (Tab. 17; Sorte als zufällig angenommen) bestimmt und im Weiteren eingesetzt. Nachfolgend benötigte LS-Mittelwerte für nun abweichend als fix definierte Sorteneffekte je Anbaugebiet

werden in einem ersten Schritt durch die Modellgleichungen 8a bzw. 8b geschätzt. Diese werden dann im zweiten Schritt zu einem gewichteten Sorteneffekt im Zielgebiet kombiniert.

1. Schritt: Schätzung fixer Sorteneffekte für jedes Anbaugebiet $(s{\times}ag)_{ij}$, wobei in der Modellgleichung der Haupteffekt S im Effekt S•AG enthalten ist.

8a: T(O•J) + Gr•AG + **S•AG** : S•O + S•J + S•J•AG + S•O•J
ohne Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen $SE^2_{Versuch}$

8b: T(O•J) + Gr•AG + **S•AG** : S•O + S•J + S•J•AG + S•O•J + S•T(O•J)
Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen $SE^2_{Versuch}$

2. Schritt: Kombination der S×AG-Effekte $(s{\times}ag)_{ij}$ zur gewichteten Linearkombination der Sortenmittelwerte im Zielgebiet (s. Möhring et al., 2004a)

Obgleich aufgrund der sehr hohen Anzahl Sorten Sortenmittelwerte anstelle aller paarweisen Differenzen ausgewiesen werden sollen, sind möglichst schätzgenaue Sortenkontraste das Ziel. Optimierungskriterium ist also ein minimaler quadrierter Vorhersagefehler von Sortendifferenzen (MSEP).

Vorabdefinitionen:

$\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots \delta_r)$ sei der Vektor der wahren Differenzen zweier Sorten in r Anbaugebieten.

$\widehat{\boldsymbol{\delta}}$ sei der Vektor der Kleinst-Quadratschätzungen dieser Differenzen δ.

$\boldsymbol{w}$ sei der Vektor der gesuchten Gewichte der r Anbaugebiete.

Gesucht wird der Vektor $\boldsymbol{w}$, welcher den MSEP (mittlerer quadrierter Vorhersagefehler einer Differenz zweier Sorten im Zielgebiet) minimiert, wenn man ihn wie folgt zur Schätzung der gewichteten über Anbaugebiete kombinierten Differenz $\tilde{\delta}_z$ für ein Zielgebiet z verwendet:

$\tilde{\delta}_z = \boldsymbol{w}'\boldsymbol{\delta}$

Durch die Einbeziehung der S×AG-Effekte von Nachbargebieten wird eine Verzerrung von Sorteneffekten im Zielgebiet verursacht. Daher muss der MSEP neben der Varianz der Differenz auch die Verzerrung der Differenz berücksichtigen:

$MSEP = Verzerrung^2 + Varianz$

Die Verzerrung der Differenz zweier Sorten im Zielgebiet ist nun definiert durch:

$Verzerrung = (\boldsymbol{w} - \boldsymbol{u_z})'\boldsymbol{\delta}$,

da das Zielgebiet nicht zur Verzerrung beiträgt; dies wird durch einen Hilfsvektor $\boldsymbol{u_z}$ mit r Elementen, einer Eins an der Position (Nummer) des Zielgebietes und Nullen für alle anderen Gebiete (Nachbargebiete) erreicht.

Der Erwartungswert der quadrierten Verzerrung ergibt sich in Abhängigkeit von den Varianzen der wahren Differenzen, aber unabhängig von den Differenzen selbst, durch:

$E(Verzerrung^2) = (\boldsymbol{w} - \boldsymbol{u_z})' \sum_{\boldsymbol{\delta}} (\boldsymbol{w} - \boldsymbol{u_z})$,

wobei $\sum_{\boldsymbol{\delta}}$ die r - dimensionale Varianz-Kovarianz-Matrix für die wahren genetischen Differenzen zweier Sorten in den r Anbaugebieten sei. Sie erfasst die Ähnlichkeit der Sorteneffekte zwischen Anbaugebieten.

Die Varianz der Differenz zweier Sorten im Zielgebiet ist definiert durch:

$Varianz = var(\boldsymbol{w}'\widehat{\boldsymbol{\delta}}|\boldsymbol{\delta}) = \boldsymbol{w}' \sum_{\boldsymbol{e}} \boldsymbol{w}$,

wobei $\sum_{\boldsymbol{e}}$ die (umweltbedingte) r - dimensionale Varianz-Kovarianz-Matrix für die Schätzfehler der aus adjustierten Mittelwerten geschätzten Differenzen zweier Sorten sei. In $\sum_{\boldsymbol{e}}$ fließen alle zufälligen Terme in den Modellgleichungen 8a bzw. 8b ein, also die unterschiedlichen Sorte × Umwelt - Interaktionsvarianzen und die Restvarianz.

Damit ergibt sich aus der Berücksichtigung von Varianz und Verzerrung folgende Schätzformel für den MSEP:

$$MSEP = (\boldsymbol{w} - \boldsymbol{u_z})' \sum_{\boldsymbol{\delta}} (\boldsymbol{w} - \boldsymbol{u_z}) + \boldsymbol{w}' \sum_{\boldsymbol{e}} \boldsymbol{w}$$

Die Bestimmung des Vektors $\boldsymbol{w}$, welcher den MSEP minimiert, erfolgt durch Differentiation nach dem Vektor $\boldsymbol{w}$, wobei die erste Ableitung gleich null gesetzt wird:

$$\frac{\partial MSEP}{\partial \boldsymbol{w}} = 2(\boldsymbol{w} - \boldsymbol{u_z})' \sum_{\boldsymbol{\delta}} + 2\boldsymbol{w}' \sum_{\boldsymbol{e}} = 0$$

$$\Leftrightarrow \boldsymbol{w}'(2\sum_{\boldsymbol{\delta}} + 2\sum_{\boldsymbol{e}}) = 2\,\boldsymbol{u_z}' \sum_{\boldsymbol{\delta}}$$

Die Umstellung der Gleichung nach $\boldsymbol{w}'$ ergibt die Schätzung des Gewichtungsvektors $\widehat{\boldsymbol{w}}$:

$$\widehat{\boldsymbol{w}}' = \boldsymbol{u_z}' \sum_{\boldsymbol{\delta}} \left(\sum_{\boldsymbol{\delta}} + \sum_{\boldsymbol{e}}\right)^{-1}$$

Die Berechnung des Mittelwertes ($\bar{y}_{iz}$) für Sorte i im Zielgebiet z erfolgt im Nachgang zum Modelldurchlauf der SAS®-Prozedur MIXED:

(1) für jede Sorte i in jedem Anbaugebiet j Berechnung eines Effektes $f_{ij} = (s{\times}ag)_{ij} - \mu_j$ und Bildung des Vektors $\boldsymbol{f}$ über die r Anbaugebiete (AG)

Die Anbaugebietsmittel μ_j für die Gruppe, der Sorte i zugehört, sind die durch die SAS®-Prozedur MIXED nach Modell 8a bzw. 8b bereitgestellten LS Means Gr×AG; die Effekte $(s{\times}ag)_{ij}$ sind die LS Means S×AG.

(2) für jede Sorte i Berechnung des gewichtet kombinierten Effektes im Zielgebiet $(s{\times}ag)_{iz} = \boldsymbol{w}'\boldsymbol{f}$

(3) $\overline{y}_{iz} = \hat{\mu}_z + (\widehat{s \times ag})_{iz}$

μ_z ist aus der Menge aller μ_j (s.o.) das Anbaugebietsmittel des Zielgebietes.

Piepho und Möhring (2005) konnten zeigen, dass bei balancierten Daten dieser Gewichtungsvektor $\widehat{\boldsymbol{w}}$ zu identischen Ergebnissen führt wie die in 5.3.6 dargelegte auf BLUP basierende Auswertung in einem Modellschritt. Allerdings ist die Auswertung in einem Schritt flexibler und auch einfacher in der Umsetzung. In den in dieser Arbeit behandelten durchweg unbalancierten Datensätzen ist die Auswertung mit BLUP in einem Schritt zu favorisieren.

Die Herleitung dieses Algorithmus‘ für die Schätzung gewichteter Sortenmittel ist allerdings die Grundlage für die nachfolgend in 5.3.8 aufgenommene zusätzliche Bedingung, dass die Summe der Gewichte ‚eins‘ ergibt ($\sum_r w_r = 1$), und wurde daher hier dargelegt.

5.3.8 Gewichtete Sortenmittelwerte ohne Schrumpfungsschätzung

Im Falle der Landessortenversuche betrachtet der Autor (wie auch allgemein im deutschen Sortenprüfsystem üblich) nicht die Gesamtheit der Sorten als Stichprobe mit einer gemeinsamen Verteilung, sondern jede Sorte als distinktes (fixes) Subjekt, das jeweils auf Grund ihrer ganz spezifischen Kombination wertbestimmender Eigenschaften zur Zulassung bzw. Aufnahme in den LSV kam. Dies entspricht der Sichtweise von Patterson (1997). Ein für die Gesamtheit aller Sorten gültiges Verteilungsmodell wird für ein einzelnes Analysemerkmal daher nicht angenommen.

Bei dieser Sichtweise ist eine Schrumpfungsschätzung für Sorteneffekte in Anbaugebieten abzulehnen. Andererseits soll das Potenzial der Einbeziehung von Nachbargebieten mit korrelierten Sorteneffekten weiterhin genutzt werden, um die Vorhersagegenauigkeit von Sortendifferenzen im Zielgebiet, den MSEP, zu minimieren. Letzteres entspricht nicht dem BLUE-Ansatz für als fix definierte Effekte (hier S×AG - Effekte). Vielmehr ist die Einbeziehung differenziert gewichteter Nachbargebiete dem BLUP-Ansatz entlehnt, die Vermeidung der Schrumpfungsschätzung aber dem BLUE-Ansatz.

Der Algorithmus baut auf dem in 5.3.7 dargelegten Ansatz und den genannten Modellgleichungen auf. Die Schätzung fixer Sorteneffekte für jedes Anbaugebiet (erster Schritt) und die Bestimmung der Verzerrung und der Varianz als Bestandteile des MSEP (im zweiten Schritt) erfolgen völlig analog zu 5.3.7.

Die Minimierung des MSEP erfolgt dann in Abweichung zu 5.3.7 unter der zusätzlichen Bedingung, dass die Summe der Gewichte ‚eins‘ ergibt ($\sum_r w_r = 1$),

Zur Realisierung dieser zusätzlichen Bedingung im Gleichungssystem wird das Verfahren des Lagrange-Multiplikators genutzt. Es handelt sich um eine Methode, Optimierungsprobleme mit einer Nebenbedingung (bzw. Nebenbedingungen) umzuformulieren. Dabei wird die Nebenbedingung durch das Setzen einer Funktion auf einen gegebenen Wert definiert, in diesem Fall durch:

$\boldsymbol{w}'\mathbf{1} = 1$ mit $\mathbf{1}$ als Hilfsvektor mit r Einsen

Diese Methode führt den Lagrange Multiplikator λ als neue skalare Variable für die Nebenbedingung ein. Der Lagrange Multiplikator wird u.a. auch in der Geostatistik

beim ‚*ordinary kriging*' als BLUP-Anwendung verwendet (Isaaks und Srivastava, 1989). Zusätzlich wird für das Minimierungsproblem eine Hilfsgröße C aufgenommen:
$C = MSEP + \lambda(\boldsymbol{w}'\mathbf{1} - 1)$,

Im Unterschied zum Ansatz in 5.3.7 muss nun außer nach $\boldsymbol{w}$ auch nach λ differenziert und jeweils die erste Ableitung gleich null gesetzt werden:

$$\frac{\partial C}{\partial \boldsymbol{w}} = 2(\boldsymbol{w} - \boldsymbol{u_z})' \sum_{\boldsymbol{\delta}} + 2\boldsymbol{w}' \sum_{\boldsymbol{e}} + \lambda\mathbf{1}' = 0$$

$$\Leftrightarrow \boldsymbol{w}'(2\sum_{\boldsymbol{\delta}} + 2\sum_{\boldsymbol{e}}) + \lambda\mathbf{1}' = 2\boldsymbol{u_z}\sum_{\boldsymbol{\delta}}$$

$$\Leftrightarrow \boldsymbol{v}' \begin{pmatrix} 2\sum_{\boldsymbol{\delta}} + 2\sum_{\boldsymbol{e}} \\ \mathbf{1}' \end{pmatrix} = 2\boldsymbol{u_z}\sum_{\boldsymbol{\delta}} ,$$

wobei $\boldsymbol{v} = (\boldsymbol{w}, \lambda)$ ein Vektor der Gewichte zuzüglich λ vom Typ $(r+1)$ sei.

$$\frac{\partial C}{\partial \lambda} = \boldsymbol{w}'\mathbf{1} - 1 = 0 \quad \Leftrightarrow \quad \boldsymbol{w}'\mathbf{1} = 1 \quad \Leftrightarrow \quad \boldsymbol{v}' \begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix} = 1$$

Die Zusammensetzung des Gleichungssystems führt zu:

$$v' \begin{pmatrix} 2(\sum_{\boldsymbol{\delta}} + \sum_{\boldsymbol{e}}) & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix} = (2\boldsymbol{u}_{\boldsymbol{z}}' \sum_{\boldsymbol{\delta}} \mid 1) ,$$

wonach nach $\boldsymbol{v}'$ umgestellt wird und sich mit $\hat{\boldsymbol{v}}$ Schätzungen für $\boldsymbol{w}$ und λ ergeben (Piepho in Michel et al., 2007a):

$$\hat{\boldsymbol{v}}' = (2\boldsymbol{u}_{\boldsymbol{z}}' \sum_{\boldsymbol{\delta}} \mid 1) \begin{pmatrix} 2(\sum_{\boldsymbol{\delta}} + \sum_{\boldsymbol{e}}) & \mathbf{1} \\ \mathbf{1}' & 0 \end{pmatrix}^{-1} = (\hat{\boldsymbol{w}}', \hat{\lambda})$$

.

Die Berechnung des Mittelwertes ($\bar{y}_{iz}$) für Sorte i im Zielgebiet z erfolgt im Nachgang zum Modelldurchlauf der SAS®-Prozedur MIXED:

(1) für jede Sorte i in jedem Anbaugebiet j Berechnung eines Effektes
$f_{ij} = (\widehat{s \times ag})_{ij} - \hat{\mu}_j$ und Bildung des Vektors $\boldsymbol{f}$ über die r Anbaugebiete (AG)
Die Anbaugebietsmittel $\hat{\mu}_j$ für die Gruppe, der Sorte i zugehört, sind die durch die SAS®-Prozedur MIXED nach Modell 8a bzw. 8b bereitgestellten LS Means Gr×AG, die Effekte $(\widehat{s \times ag})_{ij}$ sind die LS Means S×AG.

(2) für jede Sorte i Berechnung des gewichtet kombinierten Effektes im Zielgebiet
$(\widehat{s \times ag})_{iz} = \boldsymbol{w}'\boldsymbol{f}$; wobei $\boldsymbol{w}$ Teil von $\boldsymbol{v} \Leftrightarrow \boldsymbol{v} = (\boldsymbol{w}, \lambda)$

(3) $\bar{y}_{iz} = \hat{\mu}_z + (\widehat{s \times ag})_{iz}$
μ_z ist aus der Menge aller μ_j (s.o.) das Anbaugebietsmittel des Zielgebietes.

Nach der beschriebenen Außerkraftsetzung der Schrumpfung kann nicht weiter von BLUP gesprochen werden, da die Schrumpfung Eigenschaft von BLUP ist. Andererseits kann trotz fester S × AG - Effekte auch nicht von BLUE gesprochen werden, da dann konsequenterweise im Gegensatz zu diesem Ansatz die S × AG - Effekte aus Nachbargebieten nicht in die Schätzung dieser Effekte im Zielgebiet einbezogen würden.

Im Rahmen dieser Arbeit sowie in der praktischen Umsetzung im regionalen Sortenprüfwesen des Landes Mecklenburg-Vorpommern setzt der Autor standardmäßig den Ansatz ohne Schrumpfungsschätzung um. Er wird in 6.4 mit dem BLUP-Ansatz entsprechend 5.3.6 verglichen.

5.3.9 Schätzung spezifischer Anbaugebietsgewichte und Präzisionsmaße für jede Sorte

Die Herleitungen für den MSEP in 5.3.5 bis 5.3.8 sind für balancierte Daten exakt. Im balancierten Fall ist der Gewichtungsvektor für die Anbaugebiete $\boldsymbol{w}$ für alle Sortenpaare identisch und der mittlere quadrierte Vorhersagefehler der Differenz zweier Sorten gilt für alle Sortenpaare.

In Sortenprüfsystemen sind - wie dargelegt - mehrjährige Datensätze aber außerordentlich unbalanciert. Selbst Sorten gleichen Prüfstatus‘ weisen sehr häufig keine vollkommen identische Datenstruktur hinsichtlich der Besetzung in Versuchen unterschiedlicher Anbaugebiete, Jahre, Orte, Typen und deren Kombinationen auf. Die umweltbedingte Varianz-Kovarianz-Matrix $\sum_e$ bzw. die Fehlervarianzen der Differenz zweier Sorten sind spezifisch für jedes Sortenpaar und somit sollte auch der Gewichtungsvektor $\boldsymbol{w}$ für jedes Sortenpaar separat geschätzt werden.

Eine Auflistung aller paarweisen Vergleiche ist bei den vorliegenden großen Sortimentsumfängen im praktischen Einsatz allerdings kaum praktikabel. Insofern wurde nach einem geeigneten Ansatz gesucht, sortenspezifische Gewichte $\boldsymbol{w}$, Fehlervarianzen der Differenz zu einem definierten Bezug und letztlich als Präzisionsmaß einen sortenspezifischen Standardfehler ihres Mittelwertes zu bestimmen.

Eine mögliche Approximation für eine Sorte wäre die Berechnung der Gewichte für den Vergleich zu jeder anderen Sorte und die anschließende Mittelwertbildung über diese Gewichte. Diese Approximation würde allerdings zu einem großem Teil durch paarweise Vergleiche mitbestimmt sein, die aufgrund des sehr unterschiedlichen Prüfstatus‘ beider Sorten kaum relevant sind. Da Sortenvergleiche zwischen Sorten mit sehr ähnlichem Datenhintergrund / vergleichbarem Prüfstatus ungleich relevanter sind als bei sehr unterschiedlichem Prüfstatus wurde ein völlig gegensätzlicher Approximationsansatz gewählt, der zudem deutlich weniger rechenintensiv ist. Die Gewichte und die damit im Zusammenhang stehenden Vektoren und Maßzahlen sollen in der Weise sortenspezifisch sein, dass man einen Vergleich zu einer Sorte mit völlig identischem Datenhintergrund anstrebt.

Praktisch erfolgt dies nach Möhring et al. (2004a) in der Weise, dass für jede Sorte alle Daten mit ihren identischen Klassifikationen dupliziert werden und die Duplikate ausschließlich bezüglich der Stufen des Faktors Sorte neu indiziert werden ($i \rightarrow i'$). Dadurch ensteht je Sorte ein Duplikat mit identischer Daten- und Varianz-Kovarianz-Struktur. Daraufhin erfolgt eine erneute Auswertung mit identischem Modell bei fixierten Werten für die Varianzkomponenten (welche am nicht duplizierten Originaldatensatz geschätzt wurden und für den gedoppelten Datensatz eingesetzt werden).

Mit Hilfe eines ‚by'-Statements in der SAS®-Prozedur erfolgen paarweise Vergleiche für jede Sorte i zu ihrem Duplikat i'. Dabei werden für diese paarweisen Vergleiche je Anbaugebiet Standardfehler der Differenz sowie Standardfehler der Differenz der Differenz geschätzt.

Die Fehlervarianzen der Differenz bilden die r Elemente der Hauptdiagonale der sortenspezifischen Matrix $\sum_{\boldsymbol{e}}$ für r Anbaugebiete. Die Nichtdiagonalelemente $\sigma_{AG_j,AG_{j'}}$ in $\sum_{\boldsymbol{e}}$ werden nach Umstellung der Formel:

$$\sigma^2_{AG_i-AG_{j'}} = \sigma^2_{AG_i} - 2\sigma_{AG_j,AG_{j'}} + \sigma^2_{AG_{i'}} \qquad \text{nach } \left(\sigma_{AG_j,AG_{j'}}\right)$$

$$\sigma_{AG_j,AG_{j'}} = \left(\sigma^2_{AG_i-AG_{j'}} - \sigma^2_{AG_i} - \sigma^2_{AG_{i'}}\right)/(-2) \quad \text{bestimmt,}$$

mit $j, j' = 1 \cdots r$ für r Anbaugebiete (AG)
und $j \neq j'$

Während also $\sum_{\boldsymbol{\delta}}$ für alle Sorten konstant ist, wird nun für jede Sorte nach beschriebenem Approximationsansatz eine spezifische Matrix $\sum_{\boldsymbol{e}}$ bestimmt.

Die sortenspezifische Bestimmung von $\sum_{\boldsymbol{e}}$ setzt sich fort in der Schätzung des MSEP entsprechend den Algorithmen in 5.3.7 bzw. 5.3.8 und in der Schätzung eines Standardfehlers (SE_{Serie}) des Mittelwertes einer konkreten Sorte. Der MSEP ist nun ein sortenspezifischer Vorhersagefehler, was nachfolgend durch die Syntax MSEP* kenntlich gemacht werden soll. Der Parameter SE_{Serie} wird in der vorliegenden Arbeit durchgehend als sortenspezifischer Genauigkeitsparamter für den geschätzten Mittelwert einer Sorte verwendet. Er wird wie folgt aus dem MSEP* berechnet:

$$SE_{Serie(i)} = \sqrt{MSEP^*_{(i,i')}/2}$$

i, i'	i indiziert Sorte i, i' indiziert die Duplikation von Sorte i
$MSEP^*_{(i,i')}$	mittlerer quadrierter Vorhersagefehler der Sortenmittelwertdifferenz im Zielgebiet, hier konkret zwischen Sorte i und ihrem Double i'
$SE_{Serie(i)}$	Standardfehler des Mittelwertes von Sorte i im Zielgebiet

Die Mittelwert-Übereinstimmung einer Sorte zu ihrer Duplikation ist für die Bestimmung der genannten Parameter ohne Belang.

Einschränkend ist festzuhalten, dass die ermittelten $MSEP^*$ nur für Vergleiche zwischen Sorten exakt sind, welche die gleiche Datenstruktur und somit identische Kovarianz-Matrizen aufweisen.

Angesichts der sehr großen Unterschiede zwischen den Sorten in ihrer Datenstruktur ist es von besonderer Bedeutung, dass der Gewichtsvektor sortenspezifisch bestimmt wird.

5.4 Integration der Modelle in die SAS®-basierte Anwendersoftware PIAFStat

Die Länderdienststellen für das Sortenwesen der Bundesländer (LDS) haben in Zusammenarbeit mit dem Bundessortenamt das Softwaresystem PIAF (**P**lanungs-, **I**nformations- und **A**uswertungssystem für das **F**eldversuchswesen) als gemeinsame umfassende IT-Anwendung für das Versuchswesen geschaffen. PIAF ist eine Windows®-Anwendung, die auf relationale Datenbanksysteme (z.B. Oracle®) aufsetzt. Für die biometrische Auswertung findet das SAS®-System (1999) Verwendung. Eine entsprechende Schnittstelle zwischen den Systemen (PIAF und SAS) wird innerhalb des PIAF-Gesamtkonzeptes durch die eigenständige Softwarelösung PIAFStat bereitgestellt (Abb. 12). Mit PIAFStat können SAS-Verfahren auch durch nicht routinierte SAS-Anwender in maximaler Flexibilität und nutzerfreundlich eingesetzt werden. Eine Vielzahl biometrischer Verfahren wurde für PIAFStat bereitgestellt. Diese Verfahren werden in einer PIAFStat-Verfahrensbibliothek verwaltet, gepflegt und weiterentwickelt.



ADS = Auswertungs - Daten - Schnittstelle
RES = Rücklieferungs - Ergebnis - Schnittstelle
PIAF **P**lanungs-, **I**nformations- und **A**uswertungssystem für das **F**eldversuchswesen

Abb. 12: PIAFStat als Verbindung zwischen PIAF und SAS® nach Schmidtke in Michel et al. (2007b)

Aufgrund der oben beschriebenen Softwarekonstellation ergab sich die Notwendigkeit, die unter 5.2 bis 5.3 hergeleiteten Methoden auf SAS®-Basis umzusetzen. Im Auftrag der LDS und des Bundesverbandes Deutscher Pflanzenzüchter wurden hierfür an der Universität Hohenheim, Fachgebiet Biostatistik (Möhring et al., 2004b) sechs SAS-Makros entwickelt und bereitgestellt (Tab. 20). Zielstellung war die Entwicklung eines robusten und praxisrelevanten Verfahrens auf SAS®-Basis, welches mit möglichst geringem Aufwand in PIAFStat implementiert werden kann, sodass es auch von nicht-SAS-kundigen Versuchsauswertern der LDS eingesetzt werden kann.

Die Bindung an die Software SAS bringt allerdings auch gewisse Einschränkungen mit sich. Die Prozedur MIXED benötigt unter SAS bei großen unbalancierten Datensätzen mit mehreren zufälligen Effekten relativ lange Rechenzeiten. Der effizientere ‚Average Information Algorithmus' nach Gilmour et al. (1995) kann unter SAS noch nicht genutzt werden, die HPMIXED Prozedur unter SAS® wurde für die in dieser Arbeit vorgestellten erfahren noch nicht getestet. Des Weiteren kann für die Wechselwirkungen zwischen zufälligen und fixen Faktoren ausschließlich der Ansatz mit unabhängigen Wechselwirkungen verwendet werden, da SAS® nur diese verwendet (s. Ausführungen in 3.)

Tab. 20: Übersicht über die SAS® - Makros und deren Kurzbeschreibung

Makroname	**Kurzbeschreibung**
%zuordnung_typ	Ordnet den Daten die Stratifizierungsgröße *Typ* zu
%boxcox	Schätzt den Transformationsparameter φ
%varianzkomponenten	Schätzt die Varianzkomponenten;
%auswertung	Ermittelt Schätzwerte für die Sorten mit gewichteter Einbeziehung von Nachbargebieten durch BLUP (Schrumpfungsschätzung)
%lsm	Ermittelt Schätzwerte für die Sorten mit gewichteter Einbeziehung von Nachbargebieten ohne Schrumpfungsschätzung
%gewichte	Als Zusatzinformation optionale Ausgabe der Gewichte je Anbaugebiet für alle Sorten

Makro *%zuordnung_typ:*

Alle Versuche der Gesamtserie erhalten automatisiert eine Stufennummer für den Faktor *Typ*, wobei Versuche an einer Ort × Jahr - Kombination jeweils beginnend mit eins aufsteigend nummeriert werden (s. Anlage 1). Dies erfolgt ohne reproduzierbare semantische Bindung an den jeweiligen Versuchstyp (z.B. LSV, WP3, EUSV2 …). Damit kann *Typ* als Stratifizierungsfaktor an Ort × Jahr - Kombination fungieren und gleichzeitig wird die Anzahl Faktorstufen für *Typ* im Gesamtdatensatz minimiert (näheres s. 3.3.2; Anmerkungen zu *Versuchstyp* (T)). Dieses Makro muss allen folgenden Makros im Ablauf vorgeschaltet werden, um jeweils die dargelegte Klassifizierung zu gewährleisten.

Makro *%boxcox:*

Es erfolgt die iterative Bestimmung des Maximum-Likelihood-Schätzers für den Transformationsparameter φ auf Grundlage des Modells 7a. Im Standardfall beginnt die Iteration mit den Startwerten ‚0' und ‚1'. Unabhängig von den Startwerten konvergiert das Verfahren bei monomodaler Form der Log-Likelihood-Kurve. Allerdings benötigt dieses Makro in Abhängigkeit vom konkreten Datensatz häufig extrem lange Rechenzeiten (häufig viele Stunden), da für jeden Iterationsschritt das komplette Modell angepasst werden muss. Daher können optional von ‚0' und ‚1' abweichende Startwerte für φ vorgegeben werden (Abb. 13). Gibt es Vorkenntnisse zum erwarteten Schätzwert für φ, so lässt sich dadurch die Rechenzeit z.T. deutlich verkürzen. Ein Auszug aus dem Makro wird in Anlage 2 gezeigt. Ein entsprechende Makro für die Merkmalskategorie ‚*Qualität [%]*' zur Schätzung des Transformationsparameters mit der ‚*folded exponential transformation*' ist derzeit noch nicht in das PIAFStat-Verfahren integriert, sondern läuft nur direkt unter SAS®.

Makro *%varianzkomponenten*

In diesem Makro erfolgt die Schätzung der Varianzkomponenten entsprechend der Modelle 7a bzw. 7b. Die Zeit, die Rechner für die Abarbeitung der Prozeduren benötigen (performance), wurde in diesem Makro durch einige Anweisungen bzw. Optionen verbessert. Hierzu ist vorauszuschicken, dass die Zielstellung der Auswertung eine methodisch optimierte Schätzstatistik, keine Teststatistik erfordert. Die SAS-Optionen *notest* und *ddfm=residual* in der *model*-Anweisung, die Verwendung von *Subjects* (*subject=s*), die formale Betrachtung aller Effekte, die nicht den

Faktor S enthalten, als fix sowie die Subsumierung dieser Effekte im Kombinationseffekt T(O • J) (s. Tab. 16) verringern den Speicherbedarf und die Rechenzeit auf ein vertretbares Maß (derzeit < 24 Stunden bei Winterweizen). Ein Auszug aus dem Makro wird in den Anlage 4 und 5 gezeigt.

Makro *%auswertung:*

Es werden Schätzwerte für die Sorten mit gewichteter Einbeziehung von Nachbargebieten nach dem BLUP-Ansatz mit Schrumpfungsschätzung für Sorteneffekte bestimmt (entsprechend 5.3.6).

Makro *%lsm:*

Es werde Schätzwerte für die Sorten mit gewichteter Einbeziehung von Nachbargebieten bestimmt, wobei eine Schrumpfungsschätzung vermieden wird. Die Bestimmung der Gewichte der Anbaugebiete erfolgt entsprechend 5.3.8 spezifisch für jede Sorte entsprechend 5.3.9. Dieser Ansatz wird derzeit für die regionale Sorteneinschätzung in Mecklenburg-Vorpommern verwendet. Ein Auszug aus dem Makro wird in den Anlage 7 und 8 gezeigt.

Makro *%gewichte:*

Als Zusatzinformation für den Anwender können bei Bedarf darüber hinaus die verwendeten Gewichte je Anbaugebiet in einer Output-Datei abgerufen werden. Dies erfolgt nur für Sorten, die mindestens einmal im Zielgebiet geprüft worden sind. In der Ausgabedatei finden sich die näherungsweise bei BLUP verwendeten Gewichte sowie die relativen Gewichte pro Sorte. Letztere berechnen sich als Anteil des Gesamtgewichtes. Dieses Makro dient z.B. sekundären Auswertungen im Zuge der Evaluierung von unterschiedlichen Regionalisierungsansätzen (s. Tab. 36).

Die Makros wurden an der Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern (LFA) in PIAFStat-Verfahren eingebunden und dabei in eine nutzerfreundliche dialoggesteuerte Bedienungsoberfläche überführt. Innerhalb der PIAFStat-Verfahrensbibliothek wurde hierfür die Verfahrensgruppe *Hohenheim-Gülzower-Serienauswertung* mit drei Verfahren (Tab. 21) gebildet (Michel et al., 2007a).

Auszüge aus dem Programmcode derjenigen Makros, welche in PIAFStat-Verfahren integriert wurden (außer *%zuordnung_typ*, s. Tab. 21) werden in den Anlagen bereitgestellt. Dies soll allerdings lediglich einen Blick auf die Herangehensweise und auf ausgewählte Kernprozeduren gestatten. Der Auszug ist nicht durchgehend selbsterklärend aufbereitet und erlaubt es nicht, alle unter 5.3 gezeigten Algorithmen ablaufen zu lassen. Der vollständige Code kann u.a. auch aus eigentumsrechtlichen Gründen in dieser Arbeit nicht aufgezeigt werden. PIAFStat und die Verfahren können allerdings durch die LDS freigegen werden, der Vertrieb erfolgt im Auftrag der LDS durch die BioMath GmbH.

Tab. 21: PIAFStat-Verfahren der *Hohenheim-Gülzower-Serienauswertung*

Reihenfolge der Abarbeitung	**Kurz-bezeichnung in PIAFStat**	**Langbezeichnung (Label in PIAFStat)**	**zentrales Makro***
1	PHI	Datentransformation	*%boxcox*
	Auszug aus Programmcode des Makros und Output in Anlagen 2-3		
2	VK	Bestimmung der Varianzkomponenten	*%varianzkomponenten*
	Auszug aus Programmcode des Makros und Output in Anlagen 4-6		
3	MW	Bestimmung der Sortenmittelwerte ohne Schrumpfungsschätzung	*%lsm*
	Auszug aus Programmcode des Makros und Output in Anlagen 7-9		

* darüber hinaus ist zur Datensatzvorbereitung in jedem dieser PIAFStat-Verfahren das Makro *%zuordnung_typ* eingebunden (s.a. Anlage 1)

Das Verfahren *PHI* (Abb. 13) schätzt den Transformationsparameter φ. Dieser Transformationsparameter kann optional in den Verfahren *VK* (Abb. 14, Option O11) und *MW* (Abb. 15, Option O13) eingesetzt werden.

Das Verfahren *VK* (Abb. 14) schätzt die Varianzkomponenten, deren approximative Vertrauensgrenzen und die genetischen Korrelationen, stellt diese tabellarisch bereit und speichert eine definierte Datei mit den Varianzkomponenten, auf die das Verfahren *MW* zugreifen kann (Abb. 15, Option O10).

Die Verfahren *PHI* und *VK* werden durch den Autor dieser Arbeit in der Auswertung für Anbaugebiete in Mecklenburg-Vorpommern (in der Regel aufgrund neunjähriger Daten) außerhalb der Arbeitsspitzen nur alle 2-3 Jahre aktiviert. Die Parameter werden in diesem Turnus entsprechend aktualisiert. Im Rahmen dieser Arbeit sind die Datensätze für *PHI*, *VK* und *MW* aber ohne zeitlichen Versatz aufeinander abgestimmt. Derzeit sind die Algorithmen der iterativen Schätzung des Transformationsparameters nach der ‚*folded exponential transformation*' für Prozentwerte noch nicht in die PIAFStat-Verfahren integriert.

Mit dem Verfahren *MW* werden die gewichteten Sortenmittelwerte ohne Schrumpfungsschätzung bestimmt (entsprechend 5.3.8). BLUP-Schätzungen (Makro *%auswertung*) sowie auch eine Ausgabe der Gewichte (Makro *%gewichte*) können derzeit nur außerhalb von PIAFStat direkt auf SAS®-Ebene erzeugt werden. Da diese Makros im Routineeinsatz derzeit nicht genutzt werden, sind sie in den Anlagen auch nicht weiter ausgewiesen.

Aus der Notwendigkeit heraus, Ergebnisse zeitnah nach der Ernte zu veröffentlichen und Empfehlungen auszusprechen, wurden die in der Hohenheim-Gülzower-Serienauswertung enthaltenen Auswertungsschritte in separaten Makros und PIAFStat-Verfahren umgesetzt. Dadurch kann die mehrjährige Mittelwertschätzung mit dem Verfahren *MW* mit einem nur darauf ausgelegten Makro sehr zeitnah nach Eingang der jüngsten Erntedaten ausgelöst werden. Vorbereitende Arbeitsschritte wie die Schätzung von Varianzkomponenten und des eventuell notwendigen Transformationsparameters für die gesamte Versuchsserie können vorab in einer arbeitsärmeren Periode an einem langjährigen, aber zeitversetzten Datensatz erfolgen.

Gerade die letztgenannten Arbeitsschritte beanspruchen sehr viel Rechnerlaufzeit. Aufgrund des Zugriffs auf feste Werte für den Transformationsparameter und die Varianzkomponenten kann durch das Verfahren *MW* (Abb. 15) die Mittelwertschätzung zügig erfolgen, insbesondere wenn die Zeitspanne zwischen Ernte, Sortenempfehlung und Neuaussaat zu gering ist, um φ und Varianzkomponenten neu zu bestimmen.

In der Landesforschungsanstalt für Landwirtschaft und Fischerei in Mecklenburg-Vorpommern (LFA) werden die Varianzkomponenten und Transformationsparameter alle drei Jahre an einem Datensatz der letzten neun Jahre geschätzt - die Parameter werden dann drei Jahre bis zur nächsten Auswertung verwendet. Die Mittelwertschätzung dagegen erfolgt jährlich an den letzten sechs Jahren. Varianzkomponenten und Transformationsparameter sind bei diesen langjährigen Auswertungen sehr robust und verändern sich nur geringfügig graduell (s.a. Validierung in 6.5), so dass in diesem zeitlichen Versatz kein relevantes Problem gesehen wird.

Die in 5.3.3 beschriebene Möglichkeit der Gewichtung der Ausgangsdaten mittels der Kehrwerte der Fehlervarianzen ($SE^2_{Versuch}$) ist in den Verfahren VK und *MW* optional gestaltet (Abb. 14, Option O14 und Abb. 15, Option O15), die SAS-Anweisungen *weight*, *parms* und *repeated* sind dafür als sogenannte Platzhalter verfügbar und werden gegebenenfalls dynamisch im SAS-Code eingebunden.

Die wesentlichen Optionen in den drei Verfahren sind in Abb. 13 - 15 als *Screenshot* zu ersehen. Damit soll eine Vorstellung von Nutzeroberfläche und Nutzeroptionen vermittelt werden, ohne dass an dieser Stelle tiefergehende, vollverständliche Erläuterungen gegeben werden können oder sollen.

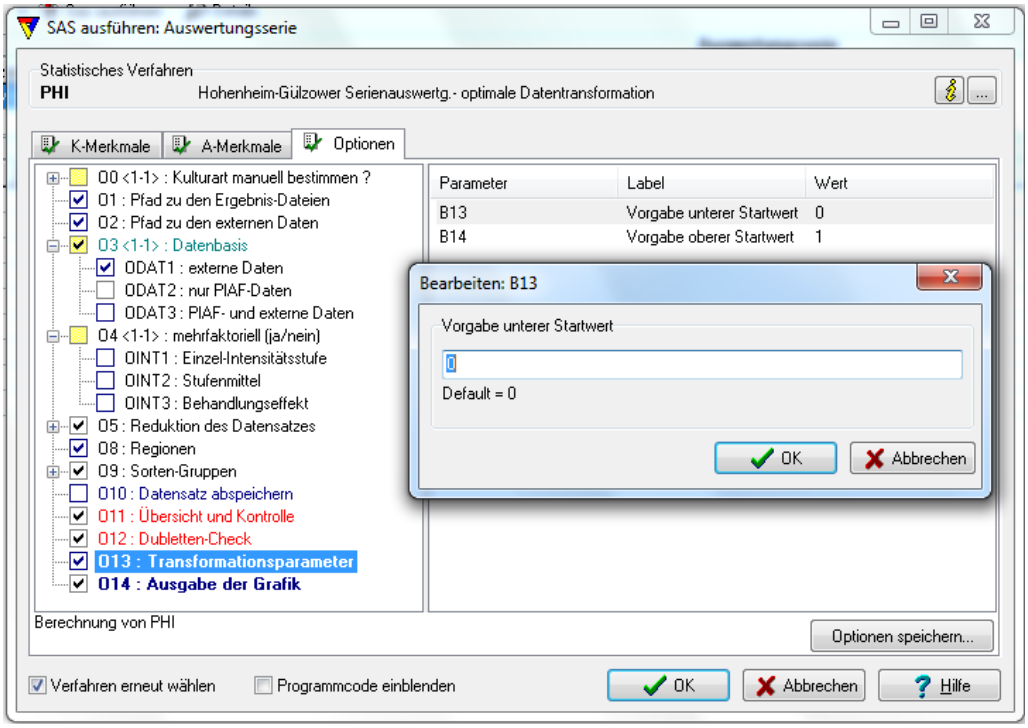


Abb. 13: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *PHI*

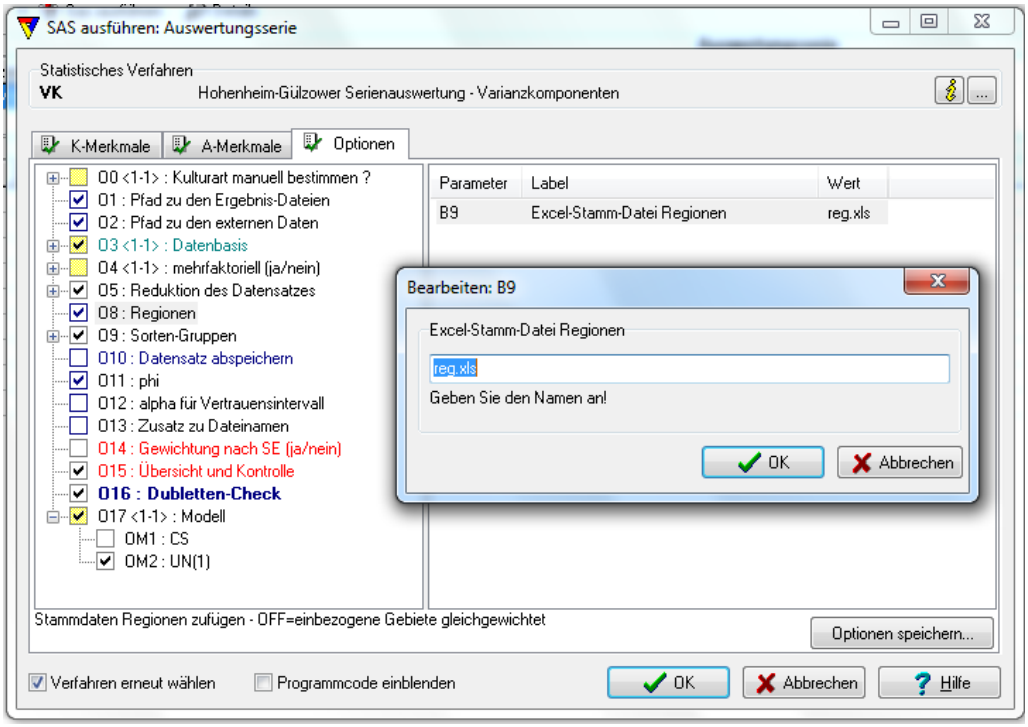


Abb. 14: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *VK*

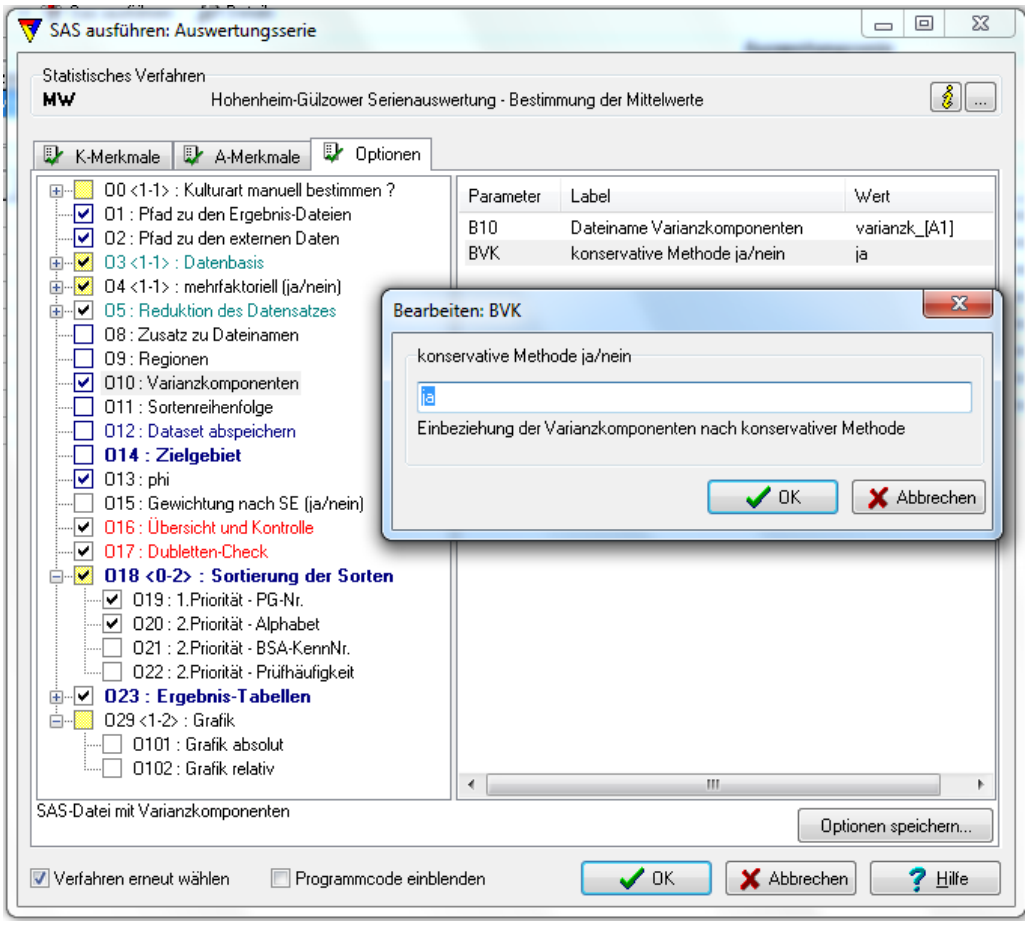


Abb. 15: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *MW*

5.5 Definition von Szenarien

Schwerpunkte der Untersuchungen sind drei Fragenkomplexe: (1) Datenverhalten hinsichtlich der Erfüllung der Modellvoraussetzungen und des Transformationserfordernisses, (2) Struktur der Varianzkomponenten und (3) Struktur der Korrelationen der Sorteneffekte zwischen Anbaugebieten. Hierbei erfolgen zunächst für alle verfügbaren Pflanzenarten und Merkmale Parameterschätzungen, vorrangig um biometrisch / methodisch verwertbare Verallgemeinerungen für Sortenprüfsysteme zu gewinnen.

Für Versuchsansteller im angewandten Versuchswesen interessiert neben diesen biometrischen Kennziffern insbesondere die Relevanz unterschiedlicher methodischer Ansätze für die Mittelwertschätzungen (womit hier im Weiteren implizit die Effekte bzw. Mittelwertdifferenzen gemeint sind). Hierzu werden zu den o.g. Fragestellungen Szenarien definiert und beispielhaft an ausgewählten Merkmalen (schwerpunktmäßig am best-besetzten Merkmal Kornertrag von Winterweizen) in ihrer Auswirkung auf die Mittelwerte bewertet.

Als **Grundszenarium** zieht sich eine im Weiteren als **Szenarium 1** (Sz 1) bezeichnete methodische Kombination durch alle Abschnitte. Diese entspricht der Standardauswertung, wie sie in Mecklenburg-Vorpommern seit Einführung der *Hohenheim-Gülzower-Serienauswertung* durch den Autor praktiziert wird.

Definition Sz 1

a) Einbeziehung vorgelagerter Versuche wie WP und EUSV,
b) Datentransformation aufgrund der Schätzung des optimalen Transformationsparameters φ mit Modell 7a an einem neunjährigen Datensatz,
c) Varianzkomponentenschätzung mit Modell 7b an einem neunjährigen Datensatz,
d) Schätzung der Sortenmittelwerte mit Modell 8b an einem sechsjährigen Datensatz, das bedeutet im Besonderen:
   - Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen,
   - Abgestuft gewichtete Einbeziehung von Nachbargebieten,
   - fixe Sorteneffekte je Anbaugebiet (BLUE).

Diese Eckpfeiler des Sz 1 werden zunächst folgenden Alternativ-Szenarien (Sz 2 bis Sz 8) gegenübergestellt, welche traditionell verbreitete Auswertungsansätze beinhalten:

Definition Sz 2

Sz 2 stimmt in a), c) und d) mit Sz 1 überein, aber eine Datentransformation entfällt (b). Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.1.2.

Definition Sz 3

Sz 3 stimmt in a), b) und c) mit Sz 1 überein, aber die Mittelwertschätzung (d) erfolgt an einem nur dreijährigen Datensatz. Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.2.2.

Definition Sz 4

Sz 4 stimmt in b), c) und d) mit Sz 1 überein, aber die Mittelwertschätzung (d) erfolgt ohne Einbeziehung vorgelagerter Versuche wie WP und EUSV (a). Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.2.2.

Definition Sz 5

Sz 5 stimmt in b) und c) mit Sz 1 überein, aber die Mittelwertschätzung (d) erfolgt an einem nur dreijährigen Datensatz ohne Einbeziehung vorgelagerter Versuche wie WP und EUSV (a). Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.2.2.

Definition Sz 6

Sz 6 stimmt in a), b) und c) mit Sz 1 überein, aber die Mittelwertschätzung (d) erfolgt anstelle des Modells 8b mit Modell 8a. Es erfolgt somit keine Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen. Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.2.3.

Definition Sz 7

Sz 7 stimmt in a), b) und c) mit Sz 1 überein, aber die Mittelwertschätzung (d) erfolgt ohne Einbeziehung von Nachbargebieten, also nur mit Daten aus dem Zielgebiet. Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.3.2.

Definition Sz 8

Sz 8 stimmt in a) und b) mit Sz 1 überein, die Mittelwertschätzung (d) erfolgt ebenfalls unter Einbeziehung von Nachbargebieten, wobei aber alle einbezogenen Anbaugebiete (Ziel- und Nachbargebiete) zu einem einzigen Gebiet zusammengefasst werden, Somit entfällt eine abgestufte Gewichtung der ursprünglichen Anbaugebiete. Die Varianzkomponentenschätzung (c) erfolgt für Sz 8 separat mit Modell 7b, wobei die Varianzkomponenten S•J•AG und S•AG auf null vordefiniert werden. Der Vergleich zu Sz 1 erfolgt im Abschnitt 6.3.3.

Auf die Untersuchung von Kombinationswirkungen der Sz 2 bis 8 wird mit einer Ausnahme verzichtet. Lediglich das Sz 5 stellt eine Kombination der Sz 3 und 4 dar.

Im Zuge der Methodenvalidierung wird im Abschnitt 6.5.3 untersucht, in welchem Maße die aus einem sechsjährigen Datensatz geschätzten Sorteneffekte (Sz1; 2004-2009) mit den geschätzten Sorteneffekten eines einzelnen unabhängig ausgewerteten Folgejahres (2010) korrelieren. Für die Bestimmung der Varianzkomponenten in diesem Folgejahr werden zwei Varianten (Sz 9 und Sz 10) definiert:

Definition Sz 9

Für die einjährige Auswertung des Folgejahres (2010) werden die Varianz–komponenten als gegeben angesehen und übereinstimmend zu Sz 1 aus der neunjährigen Auswertung übernommen (unabhängig davon, dass der Faktor ‚Jahr' nur noch mit einer Faktorstufe besetzt ist).

Definition Sz 10

Für die einjährige Auswertung des Folgejahres (2010) werden die Varianz–komponenten nicht als gegeben angesehen, sondern für die *einjährige* Auswertung

als unabhängig betrachtet und eigenständig geschätzt. Die den Faktor ‚Jahr' enthaltenden Effekte der Modellgleichung 7b werden auf null gesetzt, womit sich die Modellgleichungen für die Varianzkomponentenschätzung (7b) bzw. die Schätzung der Sorteneffekte (8b) de facto auf die Modelle $7b_{einjährig}$ bzw. $8b_{einjährig}$ reduzieren (s.u.). Zusätzlich wird Regionalspezifik der Sorteneffekte nicht weiter berücksichtigt, weil die Varianzkomponenten für S × AG aus einem einjährigen Datensatz nicht hinreichend reproduzierbar zu schätzen sind (s. 6.5.1).

Reduktion der Modellgleichungen für die Varianzkomponentenschätzung (7b) und für die Schätzung der Sorteneffekte (8b) für einjährige Auswertungen:

7b: T(O•J) + Gr•AG : S + S•O + S•J + S•J•AG + S•O•J + S•T(O•J) + S•AG

$7b_{einjährig}$: T(O) + Gr : S + S•O + S•T(O)

8b: T(O•J) + Gr•AG + S•AG : S•O + S•J + S•J•AG + S•T(O•J) + S•O•J

$8b_{einjährig}$: T(O) + Gr + S : S•O + S•T(O)

BLUE versus BLUP

Unter Sz 1 erfolgen Schätzungen von Sorteneffekten mittels BLUE (Modell 8b). Als Alternativ-Szenarium wird eine Schätzung mittels BLUP durchgeführt. Dies erfolgt mit Modell 7b. Der Vergleich von BLUP zu BLUE (Sz 1) erfolgt im Abschnitt 6.4.

Bewertung der Relevanz beim Vergleich von Szenarien

Die Bewertung der Relevanz für die Mittelwertschätzungen erfordert eine Vorabeinschätzung, welche Mindestdifferenzen als bedeutsam bewertet werden können. Erfahrungsgemäß werden Ertragseffekte über einem Prozent von Ergebnisnutzern subjektiv bereits als relevant wahrgenommen und in der Entscheidungsfindung über Sortenwahl oder anbautechnische Maßnahmen berücksichtigt, wenn es sich um mehrortige mehrjähre Ergebnisse handelt. Dies erfolgt i.d.R. ohne Berücksichtigung eines Signifikanz- oder Äquivalenztests (letzterer setzt wiederum die Vorgabe der Relevanzschwelle voraus), die ohnehin aufgrund der Unbalanciertheit für nahezu jedes Sortenpaar eine individuelle Grenzdifferenz ergäben. Im Bereich der realistischen Prüfumfange und Datenstrukturen werden nach mehrjähriger Prüfung beim Winterweizen Standardfehler der Sortenmittelwerte im Bereich geringfügig über einer dt/ha (entspricht grob gerundet wiederum etwa einem Prozent) erzielt (Abb. 27). Aus diesen Erwägungen werden Szenarien bedingte Mittelwertunterschiede, die deutlich über etwa einem Prozent bzw. konkret beim Kornertrag Winterweizen deutlich über etwa einer dt/ha liegen, im Weiteren als relevant betrachtet.

6 Ergebnisse und Diskussion

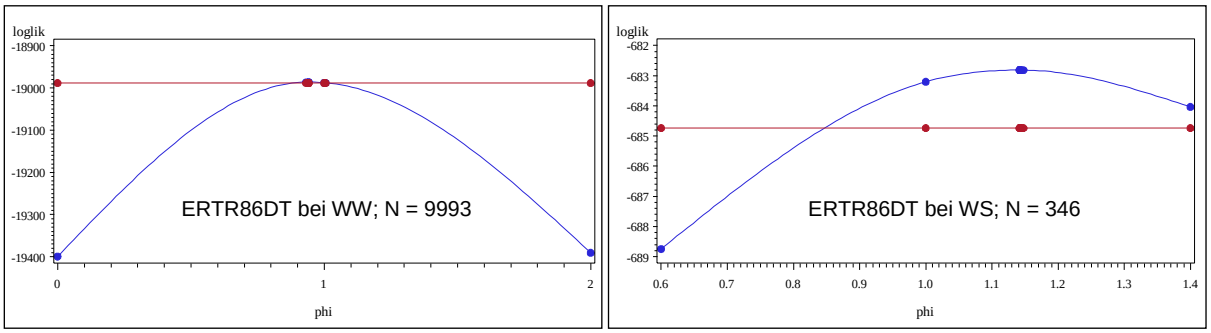
6.1 Modellvoraussetzungen, Datenverhalten und Transformation

6.1.1 Typische Muster beim Datenverhalten

Die Muster beim Datenverhalten wurden im Kontext zur Anwendung der Box-Cox-Transformation für alle Pflanzenarten und Merkmale untersucht. Anschließend erfolgte für Prozentmerkmale zusätzlich eine Untersuchung zur Anwendung der ‚folded exponential transformation'.

Zunächst wurde für jede Pflanzenart × Merkmal - Kombination der Transformationsparameter φ für die Box-Cox-Transformation geschätzt, welcher den log-likelihood - Wert maximiert (s. 5.2). In dieser Arbeit kam dafür die in Abschnitt 5.3.2 dargelegte Modellgleichung 7a zur Anwendung.

Die Güte der φ - Schätzung hängt nicht nur vom Datenverhalten ab, sondern steht u.a. in engem Zusammenhang zur Größe des jeweiligen Datensatztes. So ist z.B. bei dem größten Datensatz (Ertragsauswertung Winterweizen) das Vertrauensintervall für φ ausgesprochen klein, während es bei dem kleinen Datensatz zur Ertragsauswertung Sommerweizen relativ groß ist (Abb. 16). Dieser Zusammenhang ist trivial und soll hier nicht weiter analysiert und nachgewiesen werden.



φ = phi (Transformationsparameter)
Vertrauensintervall begrenzt durch Projektionen auf x-Achse an den Schnittpunkten der roten Linie mit Log-Likelihood-Kurve (blau); (s.a. 5.2.1, Abb. 10)

Abb. 16: Vertrauensintervall für den Transformationsparameter bei großem und bei kleinem Beispiel-Datensatz

Box-Cox-Transformation:

Die geschätzten Transformationsparameter sowie die Ergebnisse der Signifikanztests gegen die Eckwerte $\varphi = 1$ und $\varphi = 0$ sind in Tab. 22 dargestellt.

Tab. 22: Geschätzte Transformationsparameter φ; *Box-Cox-Transformation*

Pflanzenart (Kürzel)	**Merkmaleskategorie**					
	Ertrag *	**Minderertrag** **	**Qualität** [%]		**sonstige Qualität**	
	φ	φ	**Merkmal** (Label kurz)	φ	**Merkmal** (Label kurz)	φ
WW	0,94 {1}	0,69 {-}				
WS	1,14 {1}	0,80 {1}				
GW	1,07 {1}	0,77 {-}				
GS_F	0,54 {-}	0,99 {1}				
GS_B	1,03 {1}	1,07 {1}	SORTGR25	17,4 {-}	TKG_LUTR	1,26 {1}
			RP__I_TM	0,57 {-}	HLG	1,03 {1}
RW	0,46 {-}	0,74 {-}				
TIW	0,85 {-}	0,41 {-}				
HA	0,58 {-}					
RAW	1,23 {-}	0,88 {1}	GLUC91	0,53 {-}		
			FETT91	1,85 {-}		
EF	0,67 {-}		RP__86	1,62 {-}		
LUB	0,57 {-}		RP__86	0,76 {1}	TKG_86	0,22 {0}
MK	1,09 {1}		TS_KORN	1,94 {-}		
MS	0,72 {-}		TSNIGEPF	0,10 {0}	NELKGX	2,59 {-}
			STAEGEH	1,37 {-}		
KA	0,61 {-}		STAEGEKN	0,74 {-}		
Median	**0,79**	**0,79**		**1,07**		

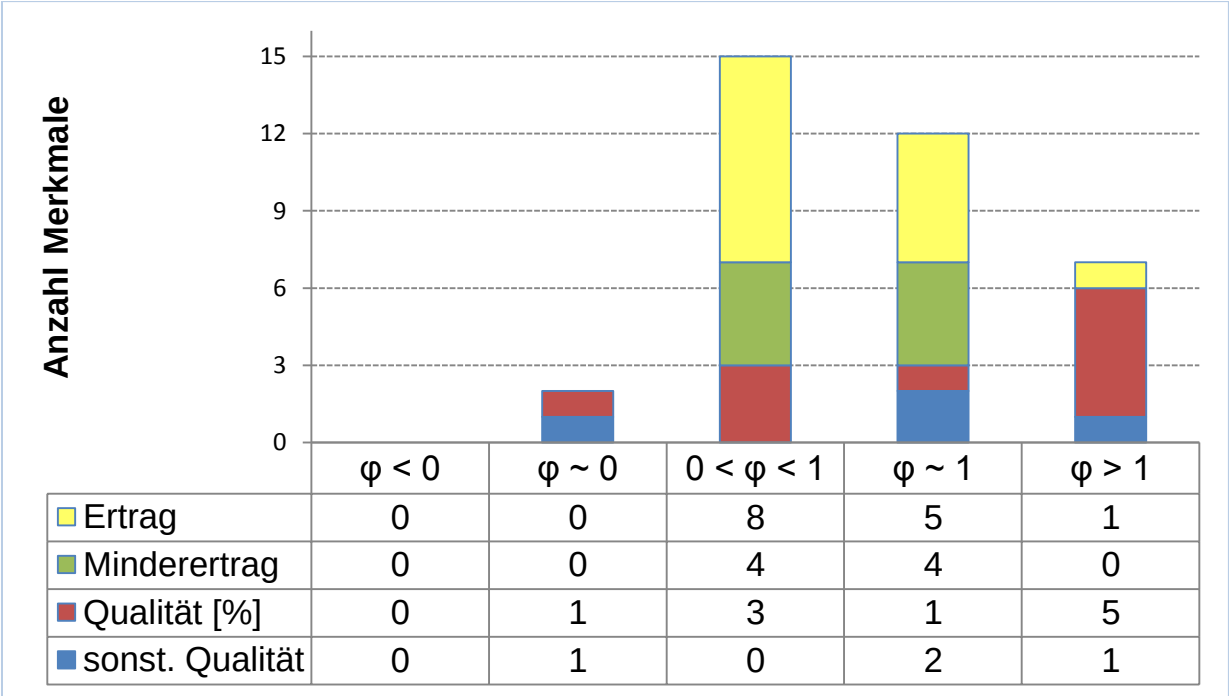
{1} $\varphi = 1$ im 95%-Vertrauensintervall enthalten, $\varphi = 0$ nicht enthalten
{0} $\varphi = 0$ im 95%-Vertrauensintervall enthalten, $\varphi = 1$ nicht enthalten
{-} $\varphi = 0$ und $\varphi = 1$ nicht im 95%-Vertrauensintervall enthalten
* Ertragsmerkmale für die einzelnen Pflanzenarten (*Label kurz* s. Tab. 3)
** Minderertrag: *Label kurz* ist ‚MinderE' (s. Tab. 3)

Existiert eine signifikante Abweichung des Transformationsparameters φ von 1, so ergibt sich die Schwierigkeit, Art und Ursache der Verletzung näher zu interpretieren. In Tab. 23 werden charakteristische Größenklassen für φ definiert und mögliche Erklärungen für ein entsprechendes Datenverhalten gegeben.

Tab. 23: Klassifizierung und Interpretation des Datenverhaltens für den Transformationsparameter φ (*Box-Cox-Transformation*)

Größenklassen für φ	**Definition der Größenklassen; Interpretation des Datenverhaltens und der Transformation**
$\varphi < 0$	$\{0;\ 1\}$ liegen nicht im Vertrauensintervall, wobei gilt: $\varphi < 0$; reziproke Transformation (*z.B. ist* $\varphi = -1$ zur Varianzstabilisierung angezeigt, wenn die Varianz proportional zur vierten Potenz des Erwartungswerts ansteigt (Thöni, 2007))
$\varphi \sim 0$	$\{0\}$ liegt im Vertrauensintervall; mögliche ursächliche Erklärung: annähernd multiplikatives Datenverhalten, Varianzen steigen mit steigenden Erwartungswerten an; adäquat zur logarithmischen Transformation
$0 < \varphi < 1$	$\{0;\ 1\}$ liegen nicht im Vertrauensintervall, wobei gilt: $0 < \varphi < 1$; ‚*intermediäres*' Datenverhalten; mögliche ursächliche Erklärung: Varianzen steigen (unterproportional mit Erwartungswerten an, Zusammenwirken additiver und multiplikativer Effekte?
$\varphi \sim 1$	$\{1\}$ liegt im Vertrauensintervall; mögliche ursächliche Erklärung: annähernd additives Datenverhalten, Varianzen unabhängig vom Erwartungswert; ist identisch mit der Auswertung untransformierter Daten
$\varphi > 1$	$\{0;\ 1\}$ liegen nicht im Vertrauensintervall, wobei gilt: $\varphi > 1$; mögliche ursächliche Erklärung: Abnahme der Varianz mit steigendem Erwartungswert

Die Häufigkeit des Auftretens dieser Größenklassen wird in Abb. 17 differenziert nach den Merkmals-Kategorien dargestellt.



	φ < 0	φ ~ 0	0 < φ < 1	φ ~ 1	φ > 1
Ertrag	0	0	8	5	1
Minderertrag	0	0	4	4	0
Qualität [%]	0	1	3	1	5
sonst. Qualität	0	1	0	2	1

Abb. 17: Häufigkeit von Größenklassen für φ nach Box-Cox-Transformation; gruppiert nach Merkmalskategorien

In der Mehrzahl der Fälle führte eine Transformation ($\varphi \neq 1$) zu einer signifikant besseren Modellanpassung als die Auswertung untransformierter Daten. Besonders häufig tritt die Größenklasse $0 < \varphi < 1$ mit signifikanter Abweichung zu diesen beiden Eckwerten auf, was hier als *intermediäres* Datenverhalten bezeichnet werden soll. In den Kategorien Ertrag und Minderertrag liegt der Median der geschätzten Transformationsparameter über die Merkmale bei $\varphi \sim 0{,}8$ (Tab. 22). Transformationsparameter bei $\varphi \sim 0$ traten in zwei Einzelfällen bei Qualitäts- bzw. Prozentmerkmalen auf. Werte unter null wären denkbar und lagen z.T. auch innerhalb der Vertauensintervalle für φ. Bei dem Prozentmerkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste (der Median der Urwerte betrug 96,4%), einem weiter unten tiefgehender diskutierten Merkmal mit extrem schiefer Verteilung der Urdaten und der Residuen, wäre das geschätzte Optimum für φ bei der Box-Cox-Transformation signifikant kleiner als null, wenn die Auswertung statt mit den Originalwerten (x) mit $(100\% - x)$ erfolgte. Diese Merkmalsdefinition wäre fachlich ebenso sinnvoll, wenn man das Merkmal nicht als Gewichtsanteil der Körner definiert, die ein Sieb mit 2,5 mm breiten Schlitzlöchern passieren, sondern als Gewichtsanteil der durch dieses Sieb fallenden Körner. Werte mit $\varphi > 1$ wurden mehrfach gefunden - überwiegend bei Prozent–Merkmalen, welche weiter unten gesondert diskutiert werden. Die beiden Prozentmerkmale mit einem oberhalb des Skalenwertes ‚50%' liegenden Wertebereich hatten erwartungsgemäß hohe, deutlich über $\varphi = 1$ liegende Schätzwerte. Alle anderen Prozentmerkmale lagen über den gesamten jeweiligen Wertebereich unter dem Skalenwert ‚50%' - sie verhielten sich bezüglich der φ - Schätzung aber ohne klares System, auch Schätzungen $\varphi > 1$ sind bei diesen Merkmalen zu beobachten (Abb. 24).

Erläuterungen zu den nachfolgend ausgewerteten Residuen

Ein Residuum stellt die geschätzte Fehlerkomponente / den Zufallsterm jeder Beobachtung im Vektor $\boldsymbol{y}$ dar (im allgemeinen gemischten Modell enthalten im Vektor $\boldsymbol{e}$; s. 5.3.1), wobei in dieser Arbeit außer den festen auch die zufälligen Effekte vom durch das Modell vorhergesagten Wert (predicted value) abgezogen wurden (in SAS® realisiert durch die Option OUTPRED; also unter Berücksichtigung der eBLUPS aller zufälligen Effekte).

Für den statistischen Test gegen die Normalverteilung wurden die Roh-Residuen verwendet (keine Studentisierung). Auswertungsvarianten mit differenzierter Transformation lassen sich bezüglich ihrer jeweiligen Güte der Modellanpassung nur anhand studentisierter Residuen miteinander vergleichen. Daher wurden für die Darstellungen in Abb. 19 studentisierte Residuen verwendet.

Normalverteilung der Residuen

Zunächst wurde die Übereinstimmung der Verteilung der Residuen mit der Normalverteilung getestet (SAS®, PROC Univariate; Shapiro-Wilk-Test für N<2000 mit Testgröße W; Kolmogorov-Smirnov-Test für N>2000 mit Testgröße D). Mit Ausnahme des Minderertrages bei Sommerfuttergerste (GS_F; MinderE) und des Rohproteingehaltes bei Lupinen (LUB; RP__86) wurde durchweg eine signifikante Abweichung von der Normalverteilung festgestellt und zwar sowohl bei Auswertung untransformierter Daten als auch nach der Box-Cox-Transformation mit dem jeweils geschätzten φ (lt. Tab. 24). Daher wurden in Schiefe bzw. Exzess der Residuen besonders auffällige Merkmale - vor oder nach Transformation - anhand von Q-Q-

Plots (Quantile-Quantile-Plots) der studentisierten Residuen explorativ-visuell auf Anpassung an die Normalverteilung überprüft (Abb. 19, weiter unten diskutiert).

Die Parameter Schiefe und Exzess sind in Tab. 24 dargestellt - jeweils bei unterlassener und bei erfolgter Transformation. Während sich in der Kennziffer Schiefe keine klare Tendenz abzeichnet, ist der ausnahmslos positive, z.T. sehr große Betrag im Exzess augenfällig - sowohl bei der Auswertung ohne Transformation als auch nach Auswertung mit Transformation. Es handelt sich also durchweg um eine *hochgipflige* bzw. *leptokurtische* Verteilung der Residuen mit einer starken Konzentration um den Mittelwert einerseits und großen Wahrscheinlichkeit auffälliger Werte außerhalb der Grenzen $\mu \pm 3\sigma$ andererseits (Autorenkollektiv, 1987b).

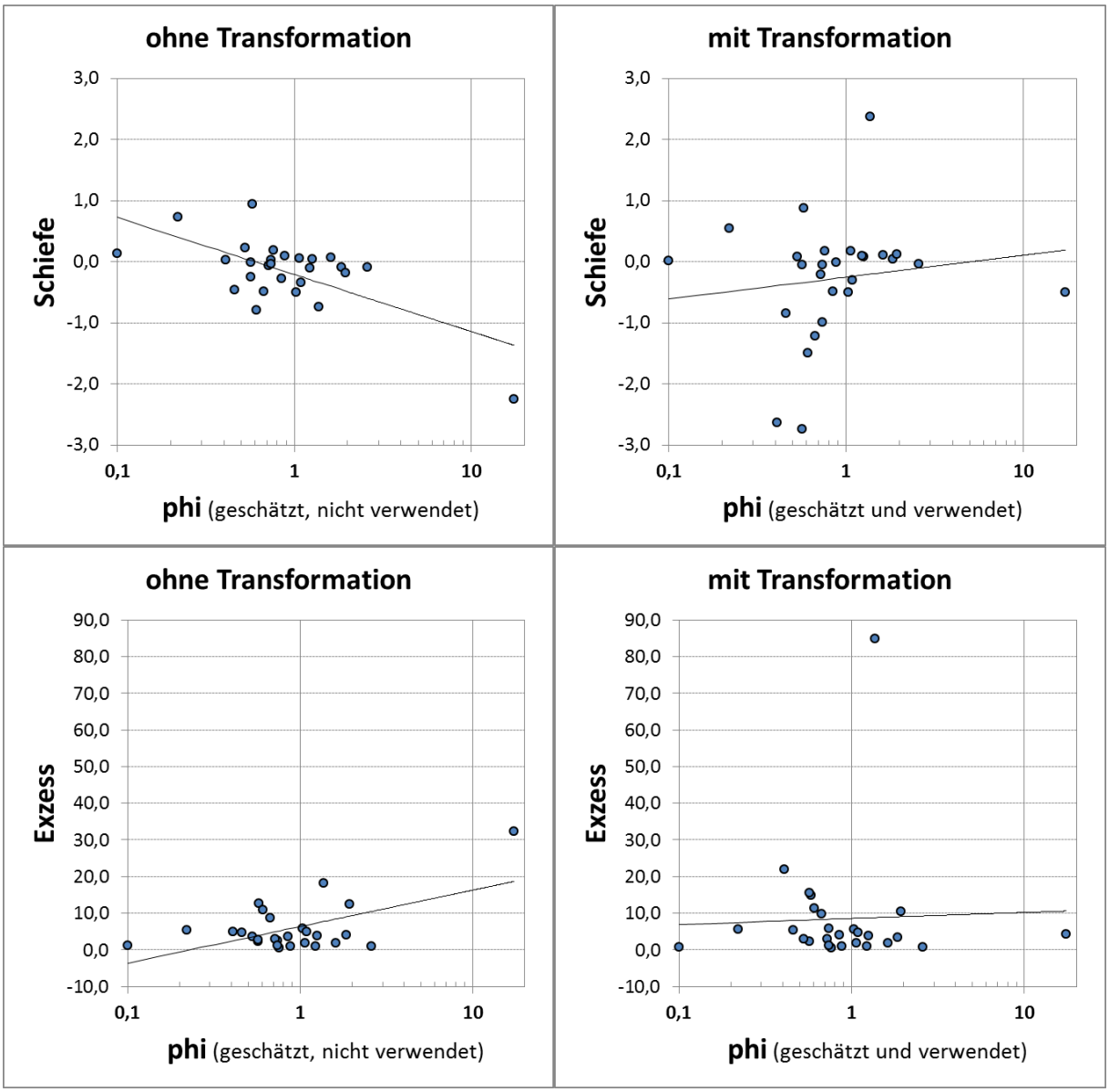
Tab. 24: Schiefe und Exzess der Verteilung der Residuen; ohne versus mit Transformation

Pflanzenart	**Merkmal**	**ohne Transformation**		**mit Transformation** (Transformationsparameter *φ)*		
		Schiefe	**Exzess**	***φ***	**Schiefe**	**Exzess**
WW	Ertrag	-0,07	1,1	0,9	-0,12	1,6
SW	Ertrag	0,25	1,4	1,1	0,35	1,5
GW	Ertrag	0,09	8,9	1,1	0,19	9,4
GS_F	Ertrag	0,05	5,0	0,5	-0,95	4,2
GS_B	Ertrag	0,07	2,5	1,0	0,08	2,4
RW	Ertrag	-0,47	4,7	0,5	-0,85	5,3
TIW	Ertrag	-0,28	3,6	0,9	-0,48	4,0
HA	Ertrag	0,94	12,6	0,6	0,88	14,9
RAW	Ertrag	-0,10	1,0	1,2	0,09	1,0
EF	Ertrag	-0,49	8,8	0,7	-1,22	9,7
LUB	Ertrag	-0,25	2,7	0,6	-2,74	15,5
MS	Ertrag	-0,06	2,9	0,7	-0,21	2,9
MK	Ertrag	-0,34	4,9	1,1	-0,30	4,8
KA	Ertrag	-0,79	10,9	0,6	-1,50	11,3
WW	MinderE	0,24	3,3	0,7	-1,19	8,5
SW	MinderE	0,48	6,5	0,8	0,47	4,4
GW	MinderE	0,09	3,1	0,8	-1,24	14,3
GS_F	MinderE *****	0,00	0,2	1,0	0,00	0,2
GS_B	MinderE	0,05	1,9	1,1	0,17	1,9
RW	MinderE	0,03	2,6	0,7	-0,99	5,9
TIW	MinderE	0,03	5,0	0,4	-2,64	21,9
RAW	MinderE	0,09	0,9	0,9	-0,01	0,9
MS	TSNIGEPF	0,14	1,3	0,1	0,02	0,8
MK	TS_KORN	-0,19	12,4	1,9	0,12	10,5
RAW	GLUC91	0,22	3,7	0,5	0,08	2,9
RAW	FETT91	-0,09	4,1	1,9	0,08	2,9
EF	RP__86	0,07	1,7	1,6	0,11	1,9
LUB	RP__86 *****	0,18	0,5	0,8	0,17	0,5
MS	STAEGEH	-0,74	18,2	1,4	2,38	84,8
KA	STAEGEKN	-0,04	1,1	0,7	-0,06	1,1
GS_B	SORTGR25	-2,25	32,2	17,4	-0,51	4,2
GS_B	RP__I_TM	-0,01	2,3	0,6	-0,05	2,3
GS_B	TKG_LUTR	0,04	3,8	1,3	0,08	3,9
GS_B	HLG	-0,51	5,7	1,0	-0,51	5,6
LUB	TKG_86	0,73	5,3	0,0	0,54	5,5
MS	NELKGX	-0,10	0,8	2,6	-0,04	0,8

(SAS®, PROC Univariate; Schiefe=Skewness, Exzess=Kurtosis; Residuen nicht studentisiert)

* Merkmale ohne signifikante Abweichung von Normalverteilung ($\alpha \leq 5\%$), sowohl ohne wie auch mit Transformation, ansonsten ausnahmslos signifikante Abweichungen von Normalverteilung

Nachfolgend werden die Kennziffern Schiefe und Exzess gegen die geschätzten Transformationsparameter geplottet (Abb. 18).



Abszisse logarithmisch skaliert; alle Merkmale entsprechend Tab. 24
Plot der einfachen linearen Regressionsfunktion ohne weitere Parameterausweisung

Abb. 18: Plot der Kennziffern Schiefe und Exzess über alle Merkmale gegen die geschätzten Transformationsparameter; ohne und mit Transformation

Schiefe und Exzess zeigen nach Transformation einen abgeschwächten Zusammenhang zum Transformationsparameter. Die Streuung um den mittleren linearen Funktionsverlauf ist aber sehr groß. Eine etwa erwartete negative Beziehung zwischen Schiefe und φ bei der Auswertung untransformierter Daten zeigt sich nur graduell. Ein Merkmal mit extremer negativer Schiefe (Merkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste →in Tab. 24: GS_B SORTGR25) zeigt aber deutlich eine derartige Kombination. Dieses Merkmal wird weiter unten tiefgehender analysiert. Die Transformation verringert die Schiefe dieses Merkmals zwar sehr erheblich, bei anderen Merkmalen (LUB Ertrag; TIW MinderE; MS STAEGEH) führt die Transformation teilweise aber zu erheblichen - positiven wie negativen - Ausschlägen in

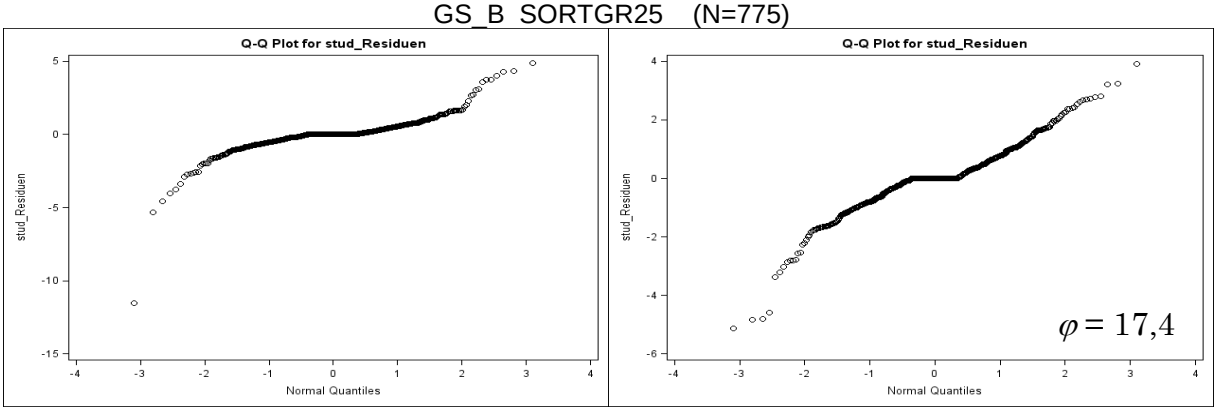
diesem Parameter. Die Verbesserung der Modellanpassung durch Transformation steht also offensichtlich nicht vordergründig im Zusammenhang mit der Schiefe der Verteilung der Residuen.

Das auffällige Merkmal ‚GS_B SORTGR25' reagiert auch in der Kennziffer Exzess positiv auf Transformation. Darüber hinaus ist aber auch hier keine unverkennbar eindeutige Tendenz im Vergleich mit versus ohne Transformation erkennbar. Vielmehr fällt einheitlich der regelmäßig positive Exzess auf - ohne wie auch mit Transformation. Im Merkmal ‚MS STAEGEH' führt die Transformation zu einem besonders extremen Exzess.

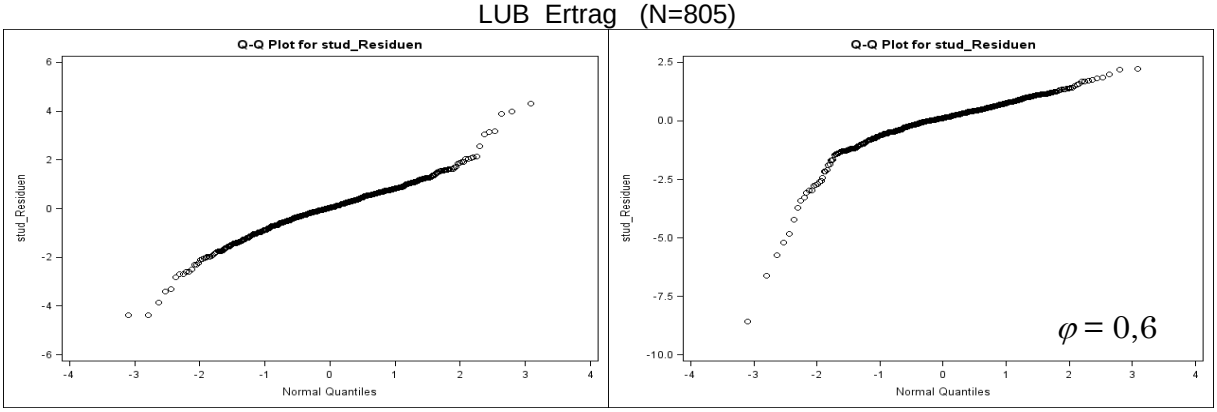
Die Residuen der vier o.g. Merkmale mit besonders auffälligen Anpassungskriterien an die Normalverteilung werden nachfolgend grafisch-explorativ anhand von Q-Q-Plots (Quantile-Quantile-Plots) der studentisierten Residuen mit der Normalverteilung verglichen (Abb. 19). Die Interpretationen der Anpassungskriterien (Tab. 24) bestätigen sich auch in dieser Grafik: die Verbesserung oder Verschlechterung der Anpassung ist offensichtlich vorrangig mit einem sehr geringen Anteil jener Daten in Verbindung zu bringen, deren Residuen im äußersten Randbereich der Verteilung liegen. Immerhin bei drei Merkmalen verschlechtert sich die Anpassung an die Normalverteilung infolge der Transformation nicht unerheblich (LUB Ertrag; TIW MinderE; MS STAEGEH). Während bei ‚GS_B SORTGR25' eine sichtbar verbesserte Anpassung durch Transformation erzielt wird, ist bei den Merkmalen ‚LUB Ertrag' und ‚TIW MinderE' zu erkennen, dass im untersten Abschnitt des Wertebereiches der Residuen die Datenstreckung durch Transformation zu einer zusätzlichen deutlichen Übersteigerung der Abweichung gegenüber der Normalverteilung führt. Beim Stärkegehalt von Silomais (MS STAEGEH) scheint die Problematik der Verschlechterung der Anpassung an die Normalverteilung nur an einem einzigen von insgesamt 7302 Werten zu liegen.

ohne Transformation	mit Transformation

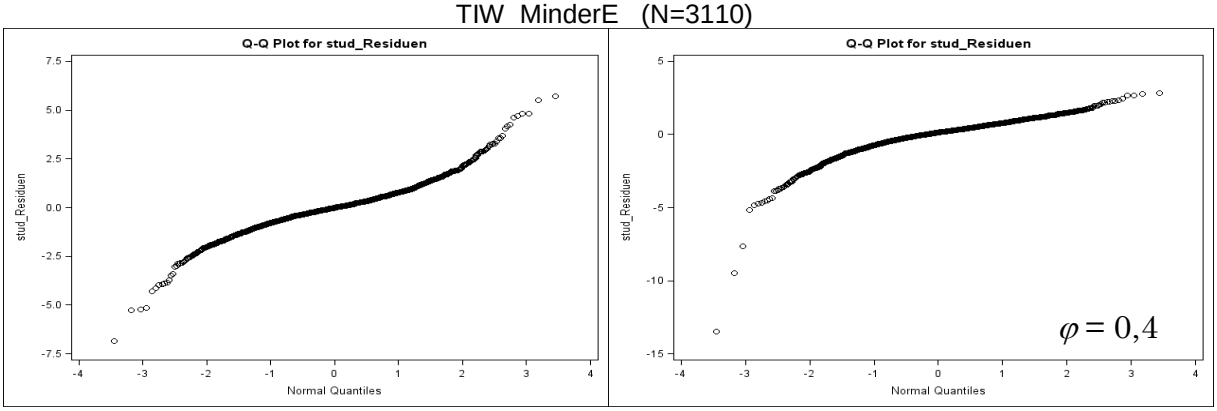
GS_B SORTGR25 (N=775)



LUB Ertrag (N=805)



TIW MinderE (N=3110)



MS STAEGEH (N=7302)

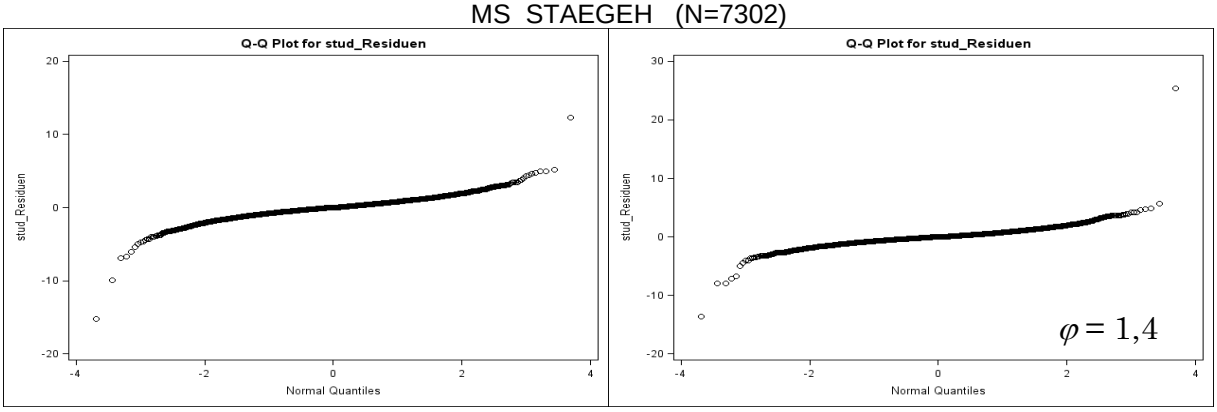


Abb. 19: Anpassung studentisierter Residuen an die Normalverteilung bei besonders auffälligen Merkmalen (Q-Q-Plots)

Unabhängigkeit der Varianzen vom Erwartungs- bzw. Vorhersagewert

Ein weiterer Erklärungsansatz für die häufige Verbesserung der Modellanpassung durch Transformation wäre, dass die Varianzen nicht unabhängig vom Betrag des Erwartungswertes sind.

Zur Untersuchung dieser Erklärungsansätze wurde geprüft, ob die absoluten Beträge der Residuen eine lineare Abhängigkeit von den Beträgen der Vorhersagewerte zeigen (Tab. 25; Definition der Residuen s.o. in 6.1.1). Außer dem Ansatz für die lineare Beziehung wurden keine weiteren Ansätze untersucht. Da die in funktionalen Zusammenhang zu bringenden Variablen ‚Vorhersagewert' und ‚Residuum' in ihrer Schätzung aus dem Modell der Versuchsserienauswertung gebunden sind, erfolgt für die Residuen in diesem Zusammenhang keine nachträgliche Studentisierung.

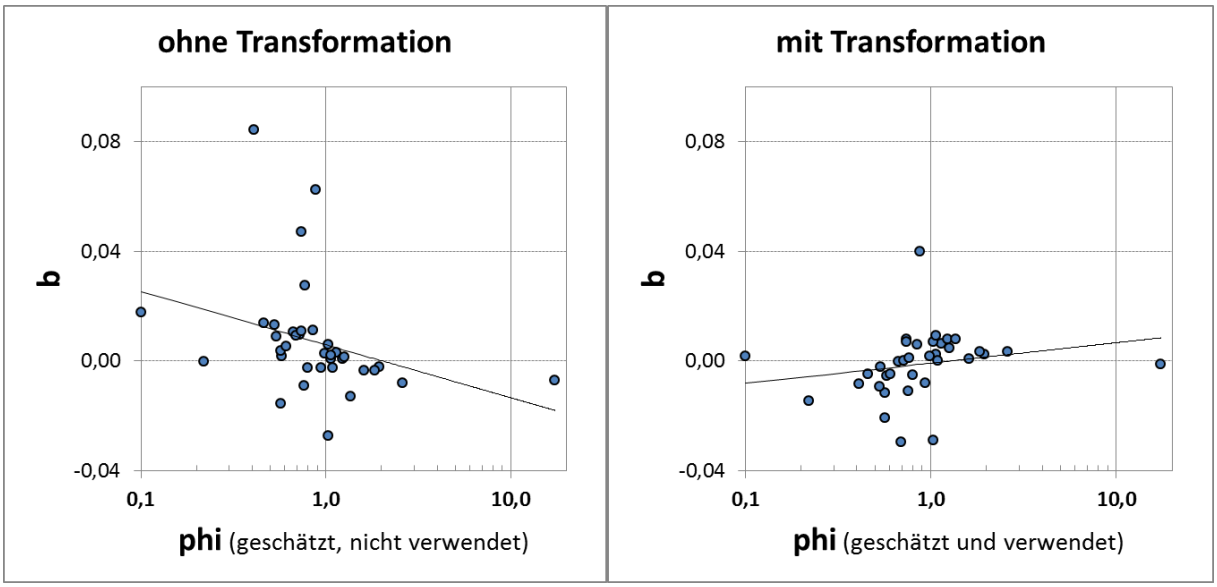
Die Bestimmtheitsmaße der einfachen linearen Regressionsfunktionen sind - erwartungsgemäß - überwiegend sehr gering (Tab. 25). Aufgrund der außerordentlich hohen Stichprobenumfänge und der daraus resultierenden großen Test-Power kann trotzdem für die Mehrzahl der Merkmale ein signifikanter linearer Zusammenhang zwischen Vorhersagewerten und absoluten Residuen nachgewiesen werden. Die konkreten numerischen Beträge der Regressionskoeffizienten sind von untergeordnetem Interesse, sie sind daher der besseren Übersicht halber mit wenigen Dezimalstellen ausgewiesen.

Tab. 25: Lineare Abhängigkeit der absoluten Residuen von den Vorhersagewerten

Pflanzenart	**Merkmal**	**ohne Transformation**		**mit Transformation** (Transformationsparameter *φ)*		
		b	**B%**	***φ***	**b**	**B%**
WW	Ertrag	-0,00 *	0,4	0,9	-0,01 *	1,4
SW	Ertrag	0,00	0,4	1,1	0,01 *	2,7
GW	Ertrag	0,00	0,0	1,1	0,00 *	0,1
GS_F	Ertrag	0,01 *	7,6	0,5	-0,00	0,0
GS_B	Ertrag	0,01 *	0,5	1,0	0,01 *	0,6
RW	Ertrag	0,01 *	2,8	0,5	-0,00 *	0,3
TIW	Ertrag	0,01 *	2,0	0,9	0,01 *	0,6
HA	Ertrag	0,00	0,1	0,6	-0,01 *	2,2
RAW	Ertrag	0,00	0,0	1,2	0,01 *	1,5
EF	Ertrag	0,01 *	2,9	0,7	-0,00	0,0
LUB	Ertrag	0,00 *	2,2	0,6	-0,01 *	9,2
MS	Ertrag	0,01 *	1,3	0,7	0,00	0,0
MK	Ertrag	-0,00	0,1	1,1	0,00	0,0
KA	Ertrag	0,01 *	2,3	0,6	-0,00 *	1,8
WW	MinderE	0,01 *	0,3	0,7	-0,03 *	2,2
SW	MinderE	-0,00	0,0	0,8	-0,01	0,8
GW	MinderE	0,03 *	1,6	0,8	0,00	0,0
GS_F	MinderE	0,00	0,0	1,0	0,00	0,0
GS_B	MinderE	0,00	0,0	1,1	0,01	0,0
RW	MinderE	0,05 *	2,8	0,7	0,01	0,0
TIW	MinderE	0,08 *	8,9	0,4	-0,01	0,0
RAW	MinderE	0,06 *	0,8	0,9	0,04 *	0,3
MS	TSNIGEPF	0,02 *	2,3	0,1	0,00	0,0
MK	TS_KORN	-0,00	0,2	1,9	0,00 *	0,3
RAW	GLUC91	0,01 *	0,3	0,5	-0,01 *	0,2
RAW	FETT91	-0,00	0,0	1,9	0,00	0,0
EF	RP__86	-0,00	0,0	1,6	0,00	0,0
LUB	RP__86	-0,01 *	2,0	0,8	-0,01 *	2,9
MS	STAEGEH	-0,01 *	1,1	1,4	0,01 *	0,8
KA	STAEGEKN	0,01 *	1,5	0,7	0,01 *	0,6
GS_B	SORTGR25	-0,01 *	5,2	17,4	-0,00	0,0
GS_B	RP__I_TM	-0,02 *	2,5	0,6	-0,02 *	4,3
GS_B	TKG_LUTR	0,00	0,0	1,3	0,00	0,0
GS_B	HLG	-0,03 *	6,3	1,0	-0,03 *	6,1
LUB	TKG_86	-0,00	0,0	0,0	-0,01 *	1,0
MS	NELKGX	-0,01 *	0,3	2,6	0,00	0,0

b linearer Regressionskoeffizient x= Vorhersagewerte, y=|Residuen|; Residuen nicht studentisiert
B% zugehöriges relatives adjustiertes Bestimmtheitsmaß,
t-Test des Regressionskoeffizienten gegen null nach PROC REG in SAS®:
* signifikant mit $\alpha \leq 5\%$

Von besonderem Interesse ist die Frage, inwieweit negative versus positive Anstiege im Zusammenhang zum geschätzten Transformationsparameter stehen. Dazu werden in Abb. 20 (basierend auf den Daten aus Tab. 25) die Transformations-parameter gegen die Regressionskoeffizienten geplottet.



alle Merkmale entsprechend Tab. 25; Abszisse logarithmisch skaliert
b linearer Regressionskoeffizient bei x=Vorhersagewert, y=|Residuum|
Plot der einfachen linearen Regressionsfunktion ohne weitere Parameterausweisung

Abb. 20: Plot der linearen Anstiege (absolute Residuen als Funktion der Vorhersagewerte) über alle Merkmale gegen die Transformationsparameter φ

Bei Merkmalen mit $\varphi < 1$ treten bei unterlassener Transformation gehäuft positive Anstiege auf, bei Merkmalen mit geschätztem $\varphi > 1$ überwiegend negative. Mit steigenden Vorhersagewerten steigt also die Varianz der Residuen. Nach erfolgter Transformation ist diese Beziehung abgeschwächt bzw. sogar leicht umgekehrt. Allerdings ist die Streuung um die mittlere lineare Funktion erheblich. Bei der Auswertung untransformierter Daten treten in einigen Fällen deutlich größere Abweichungen des Anstiegs von null auf, als bei einer Auswertung nach einer Transformation auf Basis des geschätzten Transformationsparameters. Nachfolgend werden die drei augenfälligsten Merkmale nochmal aufgeführt (analog zu Tab. 25). In allen drei Fällen handelt es sich um *Mindererträge*, die Transformationsparameter sind mit $\varphi < 1$ geschätzt und der Zusammenhang zwischen Vorhersagewerten und Residuen ist bei allen deutlich positiv.

Minderertrag Triticale: untransformiert $b = 0{,}08$; transformiert $b = -0{,}01$; $\varphi = 0{,}4$

Minderertrag Raps: untransformiert $b = 0{,}06$; transformiert $b = -0{,}04$; $\varphi = 0{,}9$

Minderertrag Roggen: untransformiert $b = 0{,}05$; transformiert $b = 0{,}01$; $\varphi = 0{,}7$

Das aufgrund des extremen Transformationsparameter von $\varphi = 17{,}4$ auffällige Merkmal ‚Sortierung > 2,5 mm (%)' bei Sommerbraugerste hatte ohne Transformation einen gesicherten, leicht negativen Anstieg der absoluten Residuen bei steigenden Vorhersagewerten, während nach der Transformation kein Zusammenhang mehr nachweisbar war.

‚Folded exponential transformation' bei Prozentmerkmalen:

Besonders bei Merkmalen der Kategorie Qualität [%] wurde häufig eine signifikante Verbesserung der Modellanpassung durch Transformation gefunden (Abb. 17). Für diese Kategorie wird nachfolgend eine u.U. adäquatere Transformationsfamilie untersucht.

Unter 5.2.2 wurde die ‚*folded exponential transformation*' als eine für Prozentmerkmale ggf. besonders angemessene Transformation beschrieben. In Tab. 26 erfolgt der Vergleich der ‚*folded exponential transformation*' zur *Box-Cox-Transformation* sowie zur Nicht-Transformation. Dabei ist zu beachten, dass die numerischen Werte für φ zwischen beiden Transformationsfamilien nicht direkt verglichen werden dürfen (Nicht-Transformation entspricht bei der Box-Cox-Transformation einem $\varphi = 1$, bei der ‚folded exponential transformation' einem $\varphi = 0$).

Tab. 26: Vergleich der Transformationsansätze bei Prozentmerkmalen (Qualität [%])

Pflanzenart	**Merkmal**	**untransformiert**	**Box-Cox**		**folded exponential**	
Label kurz		**loglike**	φ	**loglike**	φ	**loglike**
GS_B	SORTGR25	965	17,4 {-}	1235	18,5	1244
	RP__I_TM	2999	0,57 {-}	3002	4,32	3002
RAW	GLUC91	6700	0,53 {-}	6735	4,06	6739
	FETT91	8763	1,85 {-}	8774	0,00	8763
EF	RP__86	4255	1,62 {-}	4258	0,00	4255
LUB	RP__86	1432	0,76 {1}	1433	2,10	1433
MK	TS_KORN	3702	1,94 {-}	3707	2,80	3710
MS	TSNIGEPF	10956	0,10 {0}	11008	4,48	11003
	STAEGEH	9546	1,37 {-}	9583	0,00	9546
KA	STAEGEKN	3235	0,74 {-}	3237	2,25	3237
Median (N=)	**alle Merkmale**		**1,07**	(N=10)	**2,53**	(N=10)
	bei höchstem loglike		**1,07**	(N= 6)	**4,19**	(N= 4)

Wert unter ‚loglike,: maximaler *log-likelihood* - Wert je Transformationsansatz
Wert unter ‚loglike,: maximaler *log-likelihood* - Wert der drei Transformationsansätze
Box-Cox-Transformation $\varphi = 1$ entspricht untransformiert
Folded-exponential-transformation $\varphi = 0$ entspricht untransformiert
Testergebnis nur für Box-Cox-Transformation:
{1} $\varphi = 1$ im 95%-Vertrauensintervall enthalten, $\varphi = 0$ nicht enthalten
{0} $\varphi = 0$ im 95%-Vertrauensintervall enthalten, $\varphi = 1$ nicht enthalten
{-} $\varphi = 0$ und $\varphi = 1$ nicht im 95%-Vertrauensintervall enthalten

Bei neun von zehn dieser Prozentmerkmale war bereits die *Box-Cox-Transformation* der Nicht-Transformation überlegen. Zwischen den beiden Transformationsfamilien zeigten sich überwiegend eher graduelle als prinzipielle Unterschiede. Nur in vier von zehn Fällen erbrachte die ‚*folded exponential transformation*' höhere log-likelihood-Werte als die *Box-Cox-Transformation*, wobei in allen diesen Fällen bereits die letztere der Nicht-Transformation signifikant überlegen war. In drei Fällen wäre bei

ausschließlicher Berücksichtigung der ‚*folded exponential transformation*' eine Transformation unterlassen worden, während mit der *Box-Cox-Transformation* eine Verbesserung zur Nichttransformation möglich war (in allen diesen Fällen $\varphi > 1$).

Für das sehr linksschief verteilte, nahe an der Obergrenze des Parameterraumes (100%) liegende Merkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste (Median aller Urwerte betrug 96,4%) wurde bei beiden Transformationsfamilien ein auffällig hohes φ geschätzt - hier bewirkt also die Transformation eine gegen die Obergrenze zunehmend extreme Daten-Streckung, was einer Linksschiefe entgegenwirkt. In Abb. 21 wird für dieses Merkmal explorativ dargestellt, wie sich die Transformationen auf die Verteilung der Urdaten auswirken (Histogramm und Boxplot). Die extreme Schiefe wird durch beide Transformationsansätze in nahezu gleicher Weise recht gut behoben, zwischen beiden Transformationsansätzen sind hier visuell keine relevanten Unterschiede auszumachen. Aufgrund der extremen Schiefe der Urdaten (Abb. 21) und der Residuen (Abb. 22) sind aber beide Transformationen hier nicht völlig unkritisch zu beurteilen.

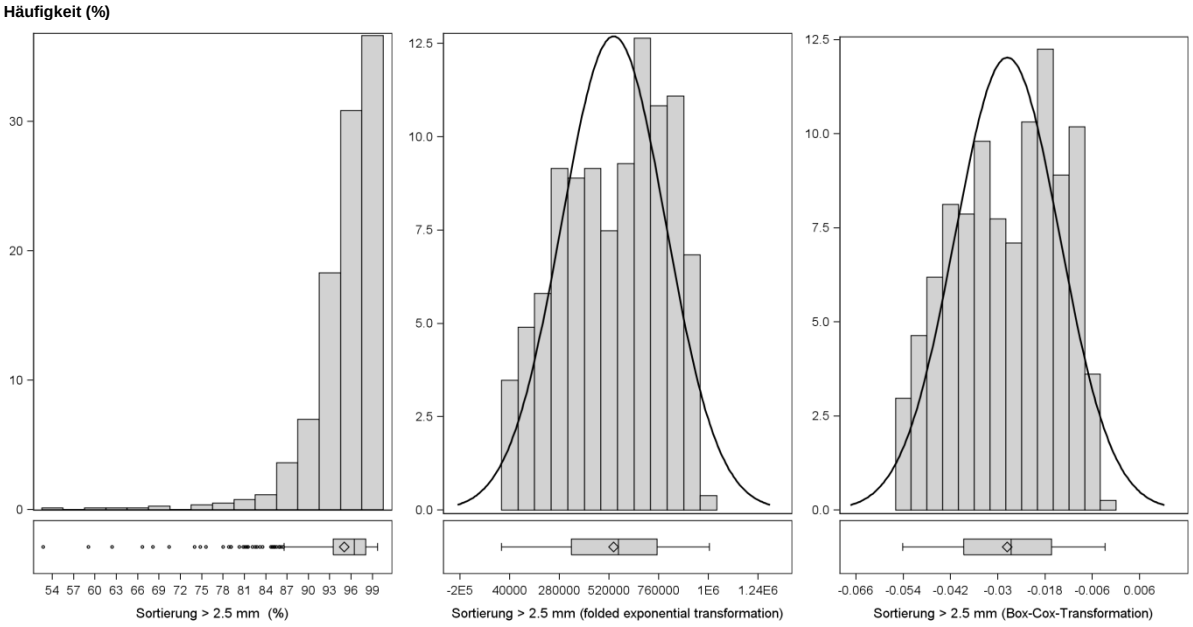


Abb. 21: Häufigkeitsverteilungen; Merkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste; Original-Daten und zwei Transformationsansätze

Wenn die Bestimmung des Transformationsparameters für die Box-Cox-Trans–formation statt an den Originalwerten (x) an deren Symmetrie-Werten $(1\text{-}x)$ bzw. $(100\%\text{-}x)$ erfolgte, so ergäbe sich bei diesem extrem schief verteilten Merkmal ein Transformationsparameter φ = -0,24. Dies wäre dann der einzige in den Untersuchungen gefundene Fall mit einem signifikant negativen Transformationsparameter.

Während Abb. 21 die Verteilung der untransformierten und transformierter *Urdaten* zeigt, wird nachfolgend die Auswirkung der ‚*folded exponential transformation*' auf die Einhaltung der Modellvoraussetzungen hinterfragt. Abb. 22 zeigt, dass bei Auswertung der entsprechend transformierten Daten eine deutlich bessere Annäherung der *Residuen* an die Normalverteilung zu verzeichnen ist, als bei unterlassener Transformation.

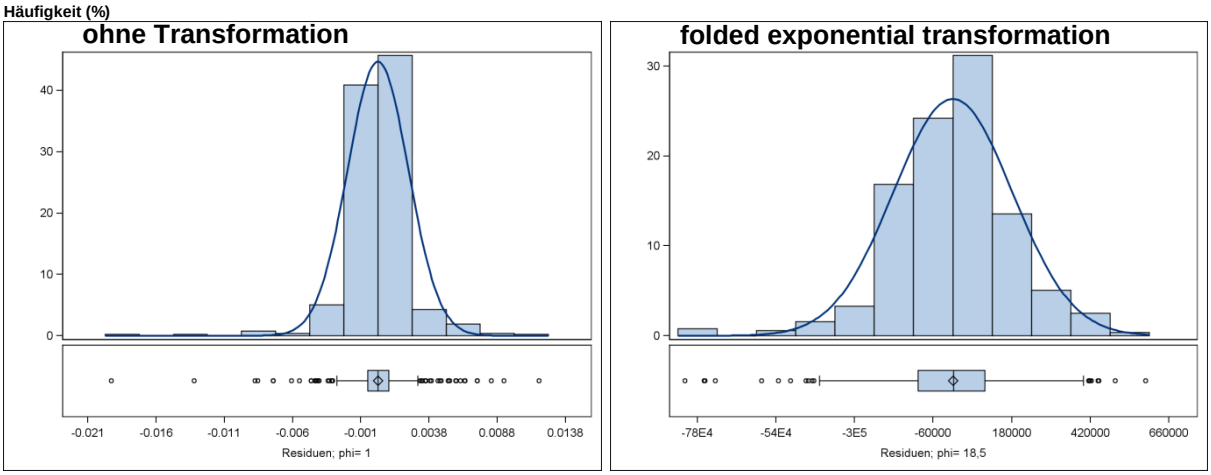
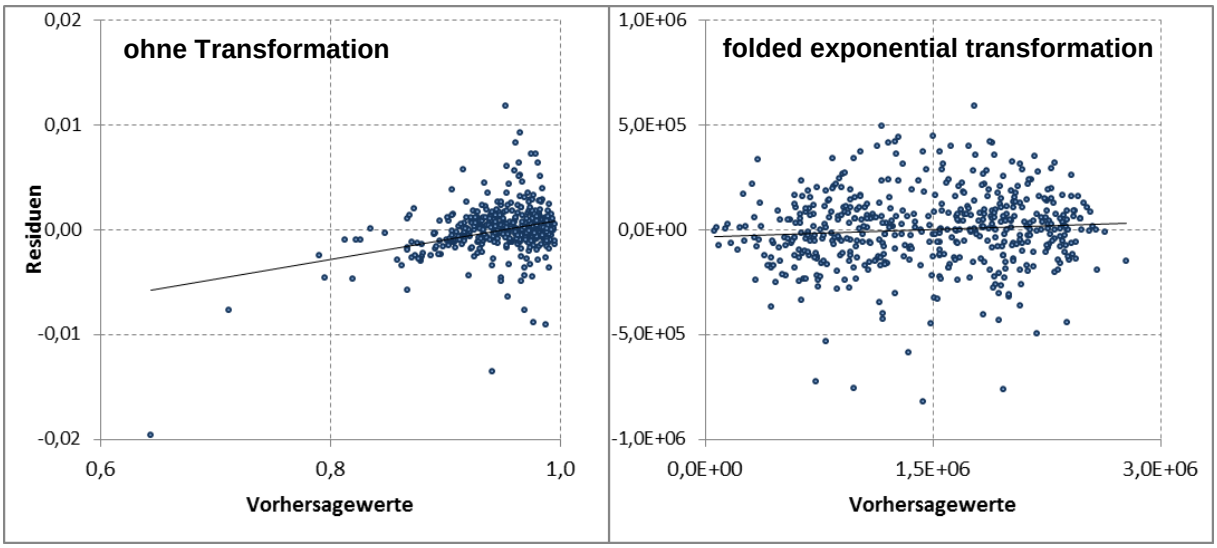


Abb. 22: Häufigkeitsverteilung der Residuen, ohne und mit der *‚folded exponential transformation‘* ; ‚*Sortierung > 2,5 mm (%)*‘ bei Sommerbraugerste;

Die Plots der Residuen gegen die geschätzten Vorhersagewerte (Abb. 23) zeigen, dass die Unabhängigkeit der Residuen vom Betrag der Vorhersagewerte nach der ‚*folded exponential transformation‘* als deutlich besser gegeben erscheint, als bei unterlassener Transformation. Bei der Auswertung untransformierter Daten ist die Einschätzung aufgrund der extrem rechtslastigen Datenbesetzung (x-Achse) beeinträchtigt. Es zeigt sich, dass ohne Transformation Ergebnisse mit auffällig geringen Vorhersagewerten negative Residuen aufweisen (im Gegensatz zu Tab. 25 und Abb. 20 sind hier nicht die absoluten Residuen, sondern die originalen Residuen (mit Vorzeichen) verwendet worden).

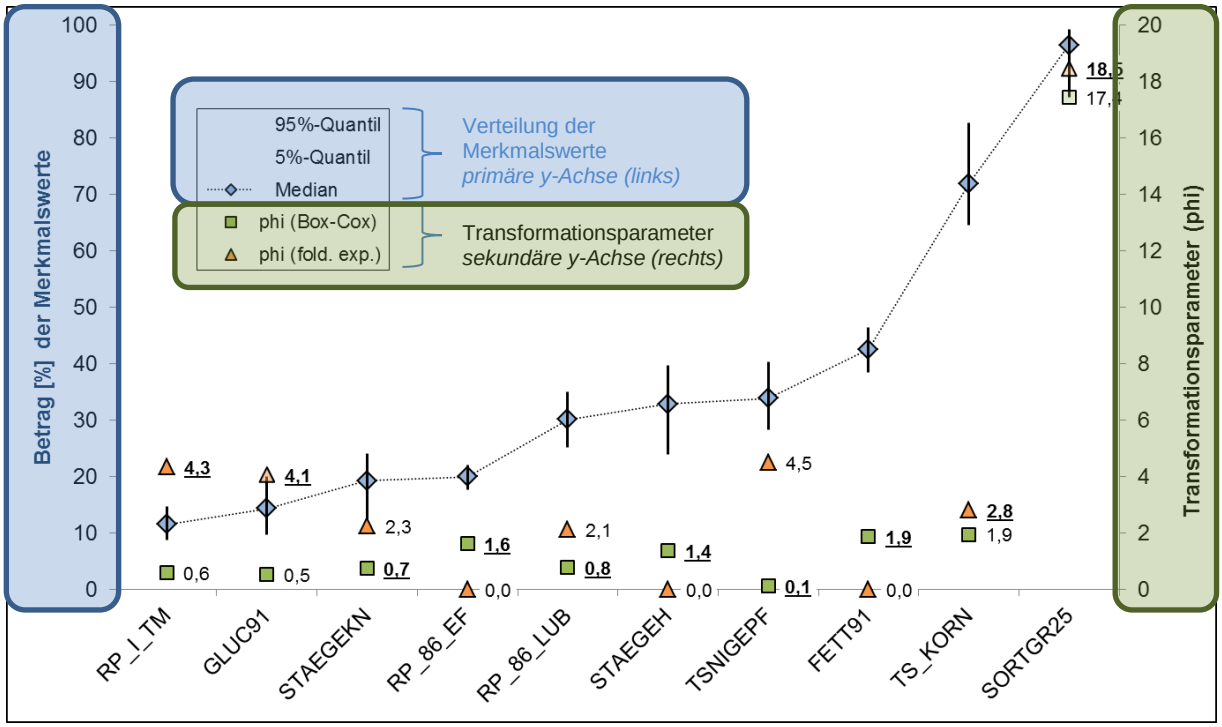


Plot der einfachen linearen Regressionsfunktion ohne weitere Parameterausweisung

Abb. 23: Plot der Residuen gegen die geschätzten Vorhersagewerte, ohne und mit der *‚folded exponential transformation‘*; ‚*Sortierung > 2,5 mm (%)*‘/100 bei Sommerbraugerste

Weiterhin wird speziell für Prozentmerkmale in Abb. 24 dargestellt, welchen Wertebereich der Prozentskala (0% bis 100%) die Urdaten der einzelnen Merkmale einerseits überdecken und welche Transformationsparameter sie - evtl. im Zusammenhang mit diesem - andererseits aufweisen. Auch hier ist zu beachten, dass die numerischen Werte für φ zwischen beiden Transformationsfamilien nicht direkt ver-

glichen werden dürfen. Offensichtliche Hypothesen ergeben sich aus der Abbildung nicht zwingend. Erwartungsgemäß zeigen Merkmale, deren Wertevorrat im unteren Bereich des Parameterraumes liegt, eher niedrige Schätzwerte für die Transformationsparameter, Merkmale im oberen Bereich des Parameterraumes höhere. Besonders auffällig ist hierbei wiederum das oben bereits diskutierte Merkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste. Der mittlere Bereich um 50% wird von keinem Merkmal überdeckt - dort wären im Analogieschluss eher keine oder marginale Transformationen zu erwarten.



Betrag von φ (phi) in der Transformationsfamilie mit höherem log-likelihood
Box-Cox-Transformation $\varphi = 1$ entspricht untransformiert
Folded-exponential-transformation $\varphi = 0$ entspricht untransformiert

Abb. 24: Transformationsparameter in Beziehung zur Lage des Wertebereiches bei Merkmalen der Kategorie ‚*Qualität [%]*'

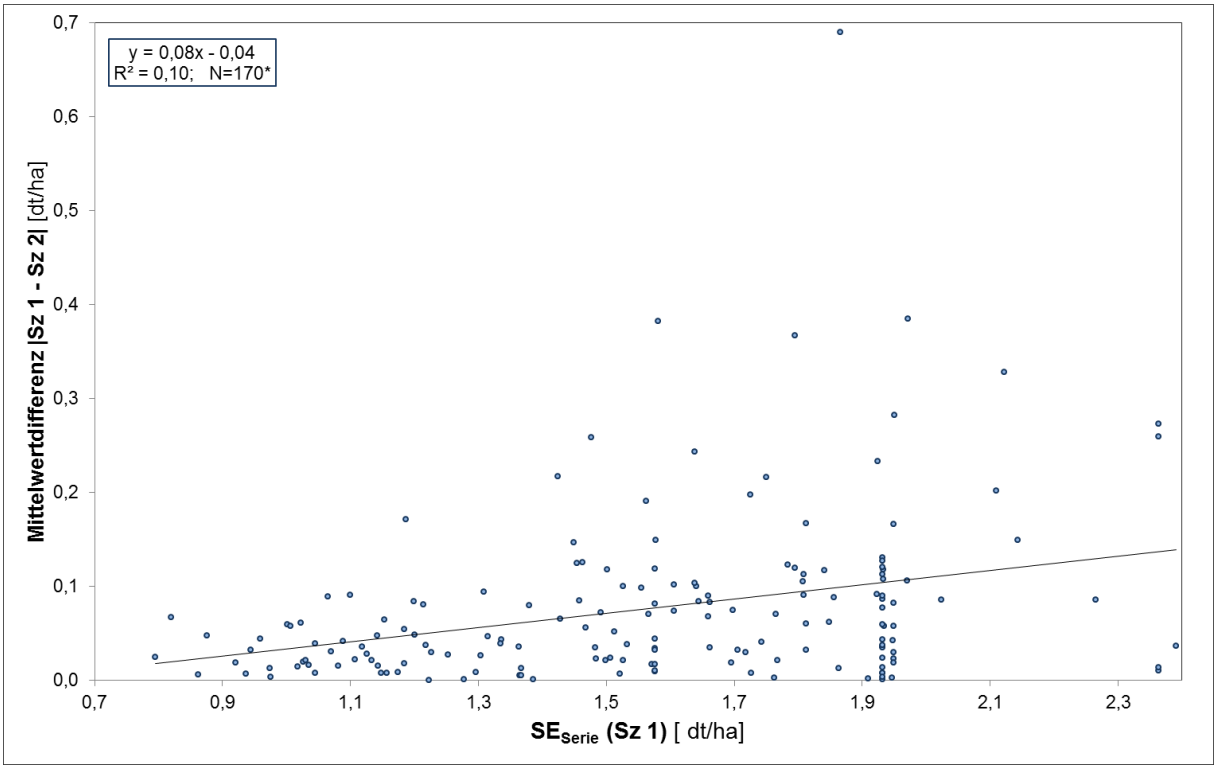
6.1.2 Effekte der Transformation auf die Mittelwerte in der Serienauswertung

Die Analyse der Effekte der Transformation lassen beim zunächst vorgesehenen Merkmal Kornertrag beim Winterweizen kaum Informationen erwarten, da hier keine signifikante Ablehnung von $\varphi = 1$ gefunden wurde (Tab. 22). Es wird daher auf das Merkmal *Minderertrag geringe Intensität* bei Winterweizen mit ebenfalls sehr großer Datenbesetzung (N = 9104), aber signifikanter Abweichung von $\varphi = 1$ ausgewichen ($\varphi = 0{,}69$ bei Box-Cox-Transformation).

Dazu werden in Abb. 25 für alle Sorten die absoluten Differenzen zwischen dem jeweiligen Mittelwert der Auswertung untransformierten Daten ($\varphi = 1$; Sz 2; Sz-Definition s. 5.5) und dem rücktransformierten Mittelwert der Auswertung mit Transformation ($\varphi = 0{,}69$; Sz 1) abgebildet (Ordinate). Rücktransformierte Mittelwerte nach Transformation sind nicht bedingungslos mit den Mittelwerten ohne Transformation zu vergleichen, da es sich bei ersteren um Schätzungen für den jeweiligen Median handelt, bei letzteren aber um Schätzungen für das arithmetischem Mittel (Erwar-

tungswert). Bei Erfüllung der Modellvoraussetzungen, im Speziellen bei symmetrischer Verteilung, ist eine übereinstimmende Lage von Median und Erwartungswert zu erwarten. Bei nichtsymmetrischer Verteilung ist der Median ein vorteilhafter Lageparameter, allerdings mit abweichenden statistischen Eigenschaften. Da in der Praxis der Sortenberatung bei Durchführung einer Transformation mit den rücktransformierten Schätzungen für den Median gearbeitet würde, veranschaulichen diese Mittelwertdifferenzen aber doch, wie sehr sich in der Beratungstätigkeit die Sorteneinschätzungen zwischen den Szenarien unterscheiden würden.

Eine für das Prüfsystem wesentliche Frage ist dabei, ob diese Differenzen im Zusammenhang mit dem Prüfstatus der Sorten (s. Tab. 8) stehen. Der Prüfstatus einer Sorte kennzeichnet ihre aktuell erreichte Stufe im Verlauf des Prüfsystems (Abb. 4). Ein fortschreitender Prüfstatus schlägt sich primär in einem mit der Prüfdauer abnehmenden SE_{Serie} (s. 5.3.9) nieder. Daher wurde als stetige Einflussgröße der SE_{Serie} (nach Sz 1) anstelle der nominalen Gruppierung der Sorten nach dem Prüfstatus‘ in die Darstellung einbezogen und der funktionale Zusammenhang dargestellt.



Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: mit Transformation
Sz 2: ohne Transformation, im Übrigen wie Sz 1 (s.a. 5.5)
* nur Sorten einbezogen, für die mindestens vier Versuchsergebnisse vorlagen

Abb. 25: Einfluss der Transformation auf die Mittelwertschätzung; ‚Minderertrag geringe Intensität‘ bei Winterweizen

Es zeigt sich erwartungsgemäß, dass die beobachteten Mittelwertunterschiede je Sorte zwischen den Szenarien nicht unabhängig vom Standardfehler der Sortenmittelwerte sind. Bei Sorten mit geringem Standardfehler ihres geschätzten Mittelwertes (also Sorten mit umfangreichem Prüfumfang) treten auch nur geringfügige Unterschiede zwischen den Schätzungen beider Szenarien auf. Allerdings kann von Varianzhomogenität der Differenzen eher nicht ausgegangen werden, so dass eine

Signifikanzaussage für den funktionalen Zusammenhang problematisch wäre und hier unterlassen wurde. Im Einzelfall wurden zwischen beiden Vorgehensweisen Differenzen bis zu 0,7 dt/ha gefunden. Ein Vergleich der SE_{Serie} zwischen beiden Szenarien wird hier im Gegensatz zu sonst ähnlichen Herangehensweisen in 6.2.2, 6.2.3 und 6.3.2 unterlassen, da die Bestimmung des ‚*Bestmodells*' ausschließlich aufgrund der Maximierung der Likelihood-Funktion erfolgt. Die Überlegenheit des Szenariums 1 wurde unter 6.1.1 lediglich über den entsprechenden Parameter für die Güte der Modellanpassung hergeleitet.

Der Korrelationskoeffizient für den Zusammenhang zwischen den Schätzwerten aus beiden Szenarien beträgt bei Einbeziehung aller Sorten etwa ‚1' $(r = 0{,}9991)$. Praktisch relevanter als der mittlere Zusammenhang ist aber die Frage, wie groß *im Einzelfall* Abweichungen bei Sorten im beratungsrelevanten Prüfstadium sein können. In Tab. 27 werden für die Kernsorten - gruppiert nach Prüfstatus - die mittleren und maximalen absoluten Differenzen zwischen Mittelwerten nach beiden Szenarien ausgewiesen.

Tab. 27: Relevanz einer Transformation für die Mittelwertschätzung im Kernsortiment; ‚Minderertrag geringe Intensität' bei Winterweizen

Prüfstatus der Sorten (Definition s. Tab. 8)	**N**	**|Sz1 - Sz2|** [dt/ha]	
		Mittelwert	**Maximum**
alte Sorten	7	0,03	0,04
LSV5-8	8	0,03	0,07
LSV4	2	0,06	0,06
LSV3	3	0,03	0,06
LSV2	6	0,02	0,03
LSV1	4	0,10	0,15
OV_WP	12	0,12	0,28
OV_EU	2	0,07	0,12

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: mit Transformation
Sz 2: ohne Transformation, im Übrigen wie Sz 1, s.a. 5.5
Definition der Szenarien: s. 5.5

6.1.3 Diskussion

Grundsatzdiskussion zu einer Vorabtransformation

Bartlett (1947) führte die Transformation nicht-normal verteilter Daten in die ANOVA-Methodologie ein. Eine Standard-ANOVA untransformierter Daten kann bei nicht-normal verteilten Daten bei ausreichendem Stichprobenumfang von der Wirkung des ‚Zentralen Grenzwertsatzes' gewinnen, wonach sich unabhängig von der Verteilungsfunktion der Einzelwerte die Verteilungsfunktion der Mittelwerte mit zunehmender Stichprobengröße der Normalverteilung annähert. Bei einem Stichprobenumfang von (ca.) n > 50 können auch bei Abweichung von der Normalverteilung t-Tests für den

Vergleich von Erwartungswerten erfolgen. Für pflanzenbauliche Einzelversuche mit typischen Wiederholungsanzahlen sieht Stroup (2014) die Stichprobe als zu klein an. Hinsichtlich einer alternativen Entscheidung zugunsten einer Transformation kommt er für unterschiedliche Konstellationen der Datenstruktur und Verteilung der Einzeldaten zum Ergebnis, dass bei Einzelversuchen eine Transformation häufig eher schädlich als nützlich ist. Diese Aussagen für typische Einzelversuche können nicht undifferenziert auf große Versuchsserien wie in dieser Arbeit übertragen werden. Für das Gros der Sorten des Kernsortimentes, insbesondere für die bereits im LSV befindlichen beratungsrelevanten Sorten kann der Mindeststichprobenumfang für die Zugrundelegung des Zentralen Grenzwertsatzes (s.o.) als gegeben angesehen werden. Vor diesem Hintergrund erscheint auch eine prinzipielle Kritik an der Alternative einer Vorab-Transformationen auf derartige Versuchsserien nicht pauschal übertragbar zu sein. Stroup sieht im Anwendungsfeld von Einzelversuchsauswertungen bei Abweichungen von Normalverteilung Vorzüge in der Anwendung von GLMMs (generalized linear mixed model; z.B. in der SAS® durch die PROC GLIMMMIX) gegenüber einer Varianz stabilisierenden Vorab-Transformation. Im aktuellen Stand der Softwareentwicklung zur Umsetzung der Hohenheim-Gülzower-Serienauswertung wurden allerdings nur gemischte Modelle eingesetzt und insofern konnte in diesem gesetzten Rahmen bislang nur die Option einer Vorab-Transformation verfolgt werden.

Box-Cox-Transformation

Bei der überwiegenden Anzahl der in dieser Arbeit untersuchten Merkmale wurde durch eine Transformation ($\varphi \neq 1$) eine signifikant bessere Modellanpassung erzielt, als bei Auswertung untransformierter Daten. Es ist von großem Interesse, ob dies primär mit der Verletzung einzelner oder mehrerer Modellvoraussetzungen erklärbar ist. Dabei muss allerdings berücksichtigt werden, dass Additivität der Effekte, Normalverteilung der Residuen und Varianzhomogenität der Zufallseffekte miteinander assoziiert sind und es somit problematisch ist, eine diese Voraussetzungen als die ursächlich primär verletzte auszumachen.

Besonders häufig wurde ein als *intermediär* interpretierbares Datenverhalten mit $0 < \varphi < 1$ gefunden, was darauf schließen lässt, dass das Verhalten untransformierter Daten häufig weder einem rein additiven noch einem rein multiplikativen Modellansatz adäquat ist. Dies wäre bezüglich des Datenverhaltens z.B. durch mit steigenden Erwartungswerten - merkmalspezifisch in unterschiedlichem Maße - steigende Varianzen erklärbar. Inwieweit sich dieser Erklärungsansatz bestätigt wird weiter unten in der Diskussion aufgegriffen.

Auch Schätzwerte für den Transformationsparameter mit $\varphi > 1$ wurden mehrfach gefunden. Thöni (2007) ging noch davon aus, dass der Wertebereich $\varphi > 1$ nicht so interessant sei, da ihm keine Beispiele bekannt seien, in denen eine Abnahme der Varianz mit steigendem Erwartungswert von Relevanz wären. Für eine Abnahme der Varianzen mit steigenden Erwartungswerten gibt es allerdings plausible Erklärungsbeispiele, die z.B. mit einer oberen Begrenzung des Parameterraumes in Zusammenhang zu bringen sind. Besonders häufig dürfte dies bei Prozentmerkmalen mit der oberen Grenze ‚100%' auftreten, wenn der Datensatz nahe dem oberen Grenzwert liegt. Bei den in dieser Arbeit untersuchten Merkmalen, die dieser Beschreibung entsprechen, hat sich dies auch so bestätigt. Von Interesse ist in diesen Fällen die weiter unten diskutierte Nutzung der ‚*folded exponential transformation*'.

Normalverteilung der Residuen

Es wurde für nahezu alle Merkmale eine signifikante Abweichung von der Normalverteilung - sowohl ohne als auch mit Transformation - gefunden. Dies sollte nicht überbewertet werden, da aufgrund der ausgesprochen umfangreichen Daten die Wahrscheinlichkeit hoch ist, das selbst sachlich relativ geringfüge Abweichungen von Modellvoraussetzungen als signifikant erkannt werden (große Test-Power). In Einzelversuchen oder auch in üblichen zeitlich befristeten Versuchsserien sind derartige Nachweise im Gegensatz zu diesen umfangreichen Datensätzen nur bei ausgeprägten Abweichungen von Modellvoraussetzungen zu erwarten.

Zudem sind Abweichungen von der Normalverteilung bei den in dieser Arbeit und dem entsprechenden Anwendungsgebiet vorliegenden umfangreichen Datensätzen u.U. weniger kritisch zu bewerten als andere Verletzungen von Modellvoraussetzungen, da entsprechend dem zentralen Grenzwertsatz die Verteilungsfunktion der Sortenmittelwerte bei den hier analysierten großen Stichprobenumfängen deutlich gegen die Verteilungsfunktion der Normalverteilung konvergiert, selbst wenn die Residuen nicht normalverteilt sein sollten.

Sehr augenfällig ist bei fast allen Merkmalen der große Betrag im Exzess - sowohl bei der Auswertung ohne Transformation als auch nach Auswertung mit Transformation. Es handelt sich um hochgipflige Verteilungen der Residuen mit einer gegenüber der Normalverteilung überhöhten Wahrscheinlichkeit für Extremwerte außerhalb der Grenzen $\mu \pm 3\sigma$. Alle Datensätze enthalten einige wenige extrem auffällige Einzelwerte. Offen bleibt, ob es sich bei diesen auffälligen Einzelwerten um regelrechte ‚Falschwerte' handelt (was in derart großen Datensätzen nicht auszuschließen und häufig nicht rückwirkend belegbar ist) oder um einzelne extreme Sorte $\times$ Umwelt - Interaktionseffekte in Ausnahmesituationen, die nicht im Variationsbereich einer gemeinsamen Verteilung liegen. Solche Fälle wären ein Hinweis auf Varianzinhomogenitäten bezüglich einzelner zufälliger Effekte im Modell als mögliche Ursache für die häufige Verbesserung der Modellanpassung durch Transformation. Eine einfache Datentransformation scheint dieses Problem des positiven Exzesses nicht hinreichend zu verbessern.

Im Gegensatz zum Exzess gab es für die Schiefe keine eindeutige Tendenz. Beim Großteil der in dieser Arbeit untersuchten Merkmale ist keine ausgeprägte Wirkung der Transformation auf eine Verringerung der Schiefe der Residuenverteilung erkennbar - lediglich bei einem Prozentmerkmal mit extremer negativer Schiefe der Verteilung der Residuen ist die erwartete negative Beziehung zwischen dem Transformationsparameter φ und der Schiefe der Residuenverteilung bei Auswertung untransformierter Daten klar zu beobachten.

Insofern kann für die in dieser Arbeit untersuchten Merkmale - mit Ausnahme ausgesprochen schief verteilter Merkmale - die Abweichung von der Normalverteilung, im Speziellen eine schiefe Verteilung, nicht als primäre Ursache für die Verbesserung der Modellanpassung durch Transformation ausgemacht werden.

Bei einem nicht unerheblichen Anteil der Merkmale wird die Anpassung an die Normalverteilung bezüglich der Parameter Schiefe wie auch Exzess nach Transformation gegenüber der Auswertung untransformierter Daten sogar deutlich schlechter. Q-Q-Plots zeigen, dass Verbesserungen wie auch Verschlechterungen der Anpassung an die Normalverteilung durch Transformation bei allen Merkmalen vorrangig mit einem sehr geringen Anteil aller Daten im äußersten Randbereich der Verteilung in Verbindung zu bringen sind, in einem Fall sogar nur mit einem einzelnen von über 7000 Werten. Plausibilitätsprüfungen in allen Phasen der Versuchs-

betreuung, Datenverarbeitung, Versuchs- und Serienauswertung kommt eine enorme Rolle zu. Insbesondere ist aber zumindest bei Merkmalen mit derartig problematischen Wirkungen einer Transformation auf die Normalverteilung der Residuen (im Speziellen auf Schiefe oder Excess) die Transformation nicht unkritisch zu bewerten, selbst wenn die Parameter für die Güte der Modellanpassung eine Transformation nahe legen.

Limpert und Stahel (2006) gehen davon aus, dass Normalverteilung auf Addition der Wirkung verschiedener Einflussfaktoren beruht, während in der Natur auch ein multiplikatives Zusammenwirken häufig ist (Zitat „Es erscheint daher sinnvoll, die Auswirkungen additiver und multiplikativer Kräfte auf die Entstehung quantitativer Variation zu betrachten."). Da sich die bei den meisten in dieser Arbeit untersuchten Merkmalen gefundene signifikante Verbesserung der Modellanpassung durch Transformation aber nicht mit schiefer Verteilung der Residuen in Verbindung bringen ließ, kann für diese Merkmale vermutet werden, dass auch eine Abweichung von Additivität der Effekte nicht als primäre Ursache für die Verbesserungen der Modellanpassung durch Transformation im Vordergrund steht.

Am ehesten kommt als Ursache für die häufig gefundene Verbesserung der Modellanpassung durch Transformation daher eine Verletzung der Annahme homogener Varianzen für alle Zufallseffekte in Frage. Das Auswertungsmodell lässt bereits Varianzheterogenität bezüglich der Fehler der Sortenmittelwerte aus Einzelversuchen und bezüglich der S $\times$ AG - Varianzen zu. Daher wurden im Rahmen vorliegender Arbeit Untersuchungen zu Varianzheterogenität bezüglicher weiterer Effekte des Auswertungsmodells zurückgestellt. Für künftige Projekte wäre es aber von großem Interesse, diese Fragestellung aufzugreifen.

Ein denkbarer Erklärungsansatz für ‚multiplikative Anteile' im Datenverhalten könnte für einzelne Merkmale auch darin liegen, dass es sich z.T. nicht um im Einzelversuch original erfasste Messwerte an der Einzelparzelle, sondern um durch multiplikative Verknüpfung z.B. einer Ertragsmessung mit dem ebenfalls ermittelten Trockensubstanzgehalt handelt (dadurch wird rechnerisch ein Flächenertrag bei einem einheitlichen definierten Feuchtegehalt berechnet). Erst diese multiplikativ gebildeten Werte werden dann varianzanalytisch ausgewertet. Dies trifft allerdings nur in den Fällen zu, bei denen außer dem Ertrag auch der Trockensubstanzgehalt für jede Parzelle und nicht als Mischprobe je Prüfglied erfasst wird. Auswertungsansätze für solche multiplikativ gebildeten Komplexmerkmale, insbesondere auch im Kontext zu einer zwischen den verknüpften Merkmalen unterschiedlichen Fehlwertestruktur, zeigen Piepho et al. (2014). Die multiplikative Verknüpfung zweier Merkmale wird in diesem Beitrag nach logarithmischer Transformation jedes dieser beiden Merkmale in eine additive Verknüpfung gewandelt, wodurch ein lineares Auswertungsmodell angepasst werden kann.

Unabhängigkeit der Varianzen vom Erwartungs- bzw. vom Vorhersagewert

Den Erklärungsansatz eines Zusammenhangs von Erwartungswert und Varianz für ein Transformationserfordernis verwendet z.B. Thöni (2007), wenn er Werte für $\varphi < 0$ damit erklärt, dass die Varianz proportional zur vierten Potenz des Erwartungswerts ansteigt (s. Tab. 23). Frensham et al. (1997) fanden, dass es in den meisten Fällen nötig sei, eine Versuchsniveau abhängige Genotyp $\times$ Umwelt - Interaktion zu berücksichtigen. Auch van Eeuwijk (2007) geht davon aus, dass die Plausibilität eines additiven Modells u.a. von der Spannweite der Umwelteffekte im Datensatz (z.B. durch Durchschnittswerte je Versuch charakterisiert) abhängen kann und dass mit

zunehmender Umweltvarianz die Annahme additiven Datenverhaltens bzw. konstanter, vom Niveau der einzelnen Umwelten unabhängiger Varianzen zunehmend unplausibel wird. Niederschlag findet ein solcher Gedankenansatz darin, dass viele Versuchsansteller subjektiv bzw. intuitiv eher zur Darstellung von Relativ- als von Absolutzahlen neigen. Letzteres steht dann allerdings nach Beobachtungen des Autors vorliegender Arbeit häufig dazu im Widerspruch, dass die den Mittelwerten zugrundeliegende Auswertung mit einem additiven Modell ohne Transformation erfolgte.

Nachfolgend werden anstelle der Erwartungswerte die mit den Residuen assoziierten, aus dem Modell geschätzten Vorhersagewerte (predicted values) genutzt und mit den Residuen in Beziehung gebracht.

Die Befunde der in dieser Arbeit untersuchten Merkmale unterstützen den theoretischen Ansatz, dass bei Merkmalen mit geschätztem $\varphi < 1$ die Restvarianz mit steigendem Vorhersagewert steigt, während sie bei $\varphi > 1$ sinkt. Insofern scheint sich in vielen Fällen eine Abhängigkeit der Varianz vom Versuchs- / bzw. Umweltniveau zu bestätigen. Entgegen der Erwartung schlägt sich dieser Zusammenhang nicht in einer entsprechenden Schiefe der Verteilung der Residuen nieder (s.o.), wie es z.B. von Thomas (2006) postuliert wird. Die Datensätze sind offensichtlich so umfangreich / die Power ist so groß, dass selbst relativ geringe Effekte einer Varianzstabilisierung durch Transformation als signifikante Verbesserung der Modellanpassung erkannt werden. Auch die große Anzahl zufälliger Effekte in der Modellgleichung kann Ursache dafür sein, dass die erwarteten Zusammenhänge bei nur einem der Zufallseffekte, in diesem Fall bei der Restvarianz, nicht entsprechend erkennbar sind (s.u.).

Einheitliches ‚Datenverhalten' bezüglich unterschiedlicher Effekte im Modell

Zur bisherigen Diskussion ist anzumerken, dass bezüglich der Normalverteilung und des Zusammenhangs zum Erwartungs- oder Vorhersagewert ausschließlich der Residual-Effekt betrachtet wurde (Definition der hier verwendeten Residuen in 6.1.1). Es ist aber nicht zwingend notwendig, dass sich andere in der Modellgleichung enthaltene Zufallseffekte analog zum Residualeffekt verhalten. Insbesondere erscheint es angesichts der dargelegten Ergebnisse fraglich, ob von Varianzhomogenität bezüglich jedes Effektes ausgegangen werden kann - eine wahrscheinlich zu starke Annahme. Die verwendeten Modellgleichungen 8a bzw. 8b enthalten drei fixe und vier bzw. fünf zufällige Effekte - verschiedenste Konstellationen für Varianzinhomogenitäten sind denkbar. Es ist z.B. ein durchaus plausibles Szenario, dass Sorteneffekte nicht unabhängig von Erwartungs- bzw. Vorhersagewerten sind, ohne dass dies für andere Effekte der Modellgleichung in gleichem Maße gilt. Künftig sollte auch in Betracht gezogen werden, dass die Varianz-Kovarianz von Sorteneffekten nicht homogen sein muss. Hu und Spilke (2011) fanden in chinesischen regionalen Versuchen mit Getreide, dass Heterogenität über alle Versuche vorlag.

Auch für den Residualeffekt wäre es denkbar, dass er sich anders verhält als andere Effekte der Modellgleichung. Plausibel wäre z.B., dass einerseits diejenigen Daten des Gesamtdatensatzes, die unter ungünstigen äußeren Umweltbedingungen erzielt wurden, häufig eher im unteren Skalenbereich liegen und dass anderseits in diesem unteren Skalenbereich ebenfalls infolge dieser ungünstigen Bedingungen hohe Restvarianzen auftreten, ohne dass andere Effekte im Modell von diesem Zusammenhang betroffen sind. In diesem Fall beträfe eine mögliche Abnahme der Varianz bei steigenden Erwartungswerten u.U. nur eine der Varianzkomponenten, nämlich die Restvarianz, während das Datenverhalten insgesamt im Zusammenwirken aller

Varianzursachen begründet liegt und vom Verhalten der Restvarianz abweichen kann. Dieser mögliche spezifische Einfluss der Restvarianz findet in der ‚Hohenheim-Gülzower Serienauswertung' quasi vorbeugende Berücksichtigung in der unter 5.3.3 vorgeschlagenen Nutzung der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen - hier wird also bezüglich eines Effektes der Modellgleichung Varianzheterogenität zugelassen.

Eine tiefergehende Analyse aller weiteren Effekte der Modellgleichung in Beziehung zur häufigen Verbesserung der Modellanpassung durch Transformation gestaltet sich komplex und problematisch in der Interpretation. Eine weitere Vertiefung erfolgte im Rahmen dieser Arbeit daher nicht, die Fragestellung ist aber für weitere Projekte von Bedeutung.

‚Folded exponential transformation' bei Prozentmerkmalen

Die bisherigen Diskussionsansätze basierten auf dem Vergleich der Nicht-Transformation mit der Box-Cox-Transformation. Nachfolgend wird dazu im Vergleich speziell für Prozentmerkmale die Nutzung der ‚*folded exponential transformation*' diskutiert.

Zwischen beiden Transformationsansätzen waren bei den untersuchten Pflanzenarten und Merkmalen keine relevanten Unterschiede auszumachen.

Ursache für die entgegen der theoretischen Überlegungen nicht gefundene Überlegenheit der ‚*folded exponential transformation*' bei Prozentmerkmalen dürften zwei Besonderheiten aller in dieser Arbeit untersuchten Prozentmerkmale sein:

- vom $\{0 \leq x \leq 1\}$ bzw. $\{0\% \leq x \leq 100\%\}$ - Parameterraum wurde bei allen Merkmalen jeweils nur ein kleiner Skalenabschnitt übergedeckt, der Wendepunkt der Transformationsfunktion bei $x=0{,}5$ (50%) wurde nie überlappt; dadurch könnte ein potenzieller / theoretischer Vorzug der sigmoiden Funktion abschwächt bzw. aufgehoben worden sein,
- die Original-Messwerte wiesen in keinem Fall die Grenzwerte des Parameterraums $x=0$ oder $x=1$ (0% oder 100%) auf.

Bei anders gearteten und erfassten Prozentmerkmalen z.B. aus dem Bereich der Phytopathologie sollte eine solche Überlegenheit häufiger zu erwarten sein, da sich bei breitem Skalenbereich und auftretenden 0- oder 1- Werten die unter 5.2.2 beschriebenen Vorteile eher realisieren können.

Effekte der Transformation auf die Mittelwerte

Eine Untersuchung der Wirkung der Transformation auf die Mittelwertunterschiede erfolgte nur an einem Merkmal, dem Minderertrag bei geringer Intensität bei Winterweizen. Auf der rücktransformierten Skala wurden zwischen den Auswertungsvarianten ‚mit' versus ‚ohne' Transformation Abweichungen im Median bis zu 0,7 dt/ha für Einzelsorten gefunden. In Bezug zum Kornertrag von Winterweizen ist die Relevanz als gering einzuschätzen, bezogen auf das mittlere Niveau des Minderertrages bei geringer Intensität von 13,6 dt/ha liegt diese Differenz allerdings über 5% und hat damit gewisse Relevanz für die Einschätzung einer betroffenen Sorte. Eine Relevanz in dieser Größenordnung ergab sich aber nur bei sehr wenigen Sorten, welche nur ausgesprochen selten geprüft wurden. Für Sorten, die aufgrund mehrjähriger Prüfung bereits Beratungsrelevanz erlangt haben kann der Einfluss auf die vergleichende Sorteneinschätzung zumindest bei diesem Merkmal nahezu vernachlässigt werden. Je größer die Abweichung des Transformationsparameters von

$\varphi = 1$ ist, desto eher sollten auch relevante Unterschiede möglich sein. Eine zunehmende Relevanz ist auch zu erwarten, wenn Sorten einseitig überproportional häufig an Umwelten mit besonders über- bzw. unterdurchschnittlichen Ausprägungen fehlen (Möhring et al., 2004b) und wenn die Variabilität der Umwelteffekte sehr hoch ist.

Abschlussbetrachtung - unter welchen Umständen kann eine Transformation empfohlen werden

Die übliche Regel, die Erfüllung der Modellvoraussetzungen für die untransformierten Daten als Ausgangshypothese (H0) anzusehen, welche nur bei signifikanter Ablehnung in Frage gestellt wird, erscheint bei den untersuchten Merkmalen zumindest für sehr umfangreiche Datensätze fraglich. Vielmehr sollte nach Ansicht des Autors dieser Arbeit in jedem Einzelfall, hier also Pflanzenart- und Merkmal-spezifisch, eine Transformation, welche zur besten Übereinstimmung mit den Modellvoraussetzungen führt, in Erwägung gezogen werden, ohne dass eine zu widerlegende Ausgangshypothese für das Datenverhalten zugrunde gelegt wird. Es handelt sich nach diesem Vorgehen nicht mehr um einen Test auf Einhaltung von Modellvoraussetzungen, sondern um die optimale Anpassung gegebener Daten an die Modelle.

Die Anwendung der in dieser Arbeit untersuchten Transformationsansätze bietet sich nur bei sehr großen Datensätzen an, welche es ermöglichen, den die Güte der Modellanpassung maximierenden Transformationsparameter φ hinreichend präzise zu schätzen (beurteilt z.B. an einem schmalen Vertrauensintervall). Insbesondere wenn sich bei der Box-Cox-Transformation keine signifikante Abweichung von $\varphi = 1$ zeigt, sich also keine signifikante Verbesserung der Modellanpassung gegenüber der unterlassenen Transformation ergibt, sollte die Durchführung der Transformation nur erfolgen, wenn auch andere, vorrangig grafisch-explorative Indizien dies eindeutig unterstützen (s.u.: Q-Q-Plots, schiefe Verteilung der Residuen, Zusammenhang von Varianz und Erwartungswert).

Entsprechend den oben diskutierten Einwendungen von Stroup (2014) sollte bei Abweichung von der Normalverteilung als Alternative zu einer Vorab-Transformation untersucht werden, ob die Anwendung von GLMMs (generalized linear mixed model; z.B. in der SAS® durch die PROC GLIMMIX) geeigneter ist. In der Softwareentwicklung zur Umsetzung der Hohenheim-Gülzower-Serienauswertung wurden bislang allerdings noch keine GLMMs eingesetzt und insofern konnte in diesem gesetzten Rahmen bislang nur die Option einer Vorab-Transformation verfolgt werden. Für die Weiterentwicklung der Verfahren besteht hier allerdings Untersuchungsbedarf.

In jüngerer Literatur werden Voraussetzungstests für eine Signifikanzprüfung auf Mittelwertunterschiede kritisch diskutiert, sofern die beiden Tests nicht an voneinander unabhängig erhobenen Datensätzen erfolgen. Kubinger et al. (2009) schlagen vor, für den Zwei-Stichproben t-Test ab einem Stichprobenumfang von 30 Erfassungen pro Stichprobe auf einen Vorab-Test auf Normalverteilung zu verzichten. Schucany und Tony Ng (2006) nennen hierfür einen Stichprobenumfang über 50. Sie zeigen, dass Vorab-Tests auf Erfüllung von Modellvoraussetzungen zu einem fehlerhaften Signifikanzlevel z.B. im anschließenden t-Test führen. Explorative, grafische Analysen (wie Quantile-Quantile-Plots) seien eher akzeptabel. Rochon und Kieser (2011) zeigen, dass bei einer Abweichung von der Normalverteilung ein Vorab-Test auf Voraussetzungen die Verteilung der Mittelwerte bei den für den t-Test selektierten Stichproben beeinflusst. Zimmerman (1996) und (2004) sieht eine Verzerrung des Signifikanz-Levels für Mittelwertvergleiche durch Vorab-Tests auf

Varianzhomogenität besonders dann, wenn die Varianzen sich nur leicht unterscheiden.

In der Anwendung der Hohenheim-Gülzower-Serienauswertung ist die Kernauswertung - die Mittelwertschätzung - klar als Schätz- und nicht als Testverfahren definiert. Insofern ist die Problematik der exakten bzw. hinreichend robusten Einhaltung eines vorgegebenen Signifikanzlevels beim Mittelwertvergleich nicht unmittelbar relevant. Vielmehr wird ein Transformationsansatz gesucht, der die Modellvoraussetzungen am besten erfüllt und damit zur besten Anpassung des Modells an die Daten führt. In diesem Vorgehen sieht der Autor vorliegender Arbeit Parallelen zur zunehmend akzeptierten Heranziehung von Modell-Anpassungskriterien bei der Modellwahl bzw. Modellreduktion, wie es z.B. durch Richter et al. (2007) aufgezeigt wurde.

Dessen ungeachtet beeinflusst die Entscheidung zur Nutzung des geschätzten Transformationsparameters die nachfolgende Varianzkomponenten- und Mittelwertschätzung. Insofern schlägt der Autor der vorliegenden Arbeit aufgrund der o.g. Argumente aus der Literatur vor, die Entscheidung zur Transformation nicht ausschließlich davon abhängig zu machen, ob eine signifikante Abweichung zu $\varphi = 1$ besteht. Diese Entscheidung sollte nach Ansicht des Autors auch kein Automatismus aufgrund der Schätzung des Transformationsparameters, der zur besten Modellanpassung führt, sein. Vielmehr sollten die bereits in der vorliegenden Arbeit genutzten Diagnosegrafiken (Histogramm und Anpassung an Normalverteilung für die Residuen, Q-Q-Plot für die Residuen, Plot der Residuen gegen die geschätzten Vorhersagewerte) für die Entscheidung zur Transformation in den Vordergrund gerückt werden. Führt die Transformation zu einer Verschlechterung der Übereinstimmung mit Modellvoraussetzungen, so sollte die Transformation ggf. eher unterlassen werden.

Das weiter oben beschriebene Vorgehen des Autors vorliegender Arbeit im Sortenwesen in Mecklenburg-Vorpommern führt immerhin dazu, dass die Daten zur Untersuchung auf das Transformationserfordernis sich nur anteilig mit den Daten der eigentlichen Auswertung zur Schätzung der Sortenleistungen und -differenzen decken. Die Vorab-Analyse erfolgt an einem neunjährigen Datensatz und wird nur alle drei Jahre wiederholt. Die eigentliche Mittelwertschätzung erfolgt dagegen jährlich mit den Daten der letzten sechs Jahre. Eine komplette Datensatztrennung ist allerdings in der praktischen Umsetzung unrealistisch.

Einschränkend muss auch zum Ansatz der Vorab-Transformation gesagt werden, dass man durch diese Transformation ggf. zwar erreicht, dass die Daten durch ein lineares, homoskedastisches Modell am besten beschrieben werden können (Thöni, 2007), die Gültigkeit der Modellvoraussetzungen ist damit allerdings nicht zwangsläufig gewährleistet (was nach Einschätzung des Autors von realen Daten ohnehin nicht im absoluten Sinne, sondern bestenfalls hinreichend erwartet werden kann). Immerhin schätzt Thomas (2006) ein, dass es mit derartigen varianzstabilisierenden Transformationen meistens möglich ist, gleichzeitig mehrere Verletzungen der Modellvoraussetzungen zu beheben.

6.2 Varianzkomponenten- und Mittelwertschätzung

6.2.1 Typische Muster in der Struktur der Varianzkomponenten

Auf Basis der beschriebenen Modelle 7a bzw. 7b (s. 5.3.2) wurden für alle Pflanzenart × Merkmal - Kombinationen die Varianzkomponenten für den einbezogenen Raum diluvialer Standorte in Nordostdeutschland geschätzt. Bei allen nachfolgend ausgewiesenen Varianzen mit positivem Betrag (>0) liegt ein signifikanter Unterschied zu null vor. Wie unter 5.1.1 für den Transformationsparameter gezeigt, steht auch hier die Breite der Vertrauensintervalle in hohem Maße im Zusammenhang mit dem Datenumfang und der Datenstruktur je Merkmal, was aber hier nicht weiter analysiert und dargestellt werden soll.

Zunächst sollen die Größenordnungen der Sorte × Umwelt - Gesamtvarianzen analysiert und mit der Fehlervarianz der Sortenmittelwerte in Einzelversuchen verglichen werden (Tab. 28). Die S×T(J×O) - Varianz lässt sich nur bei Anwendung von Modell 7b von der Fehlervarianz in Einzelversuchen isolieren. Sie liegt in ihrer sachlichen Bewertung zwischen Sorte × Umwelt - Varianzen und der Fehlervarianz der Sortenmittelwerte in Einzelversuchen. Wohin diese Bewertung eher tendiert hängt davon ab, ob sie eher als Schlag spezifische Reaktion an einem Ort (Ort als Mikro-Region mit Schlägen / Unter - Orten) oder eher als nicht näher bestimmbare Fehlerkomponente an einer Ort × Jahr - Kombination interpretiert werden sollte. In Auswertungen ohne Gewichtung und damit automatisch auch bei allen Qualitätsmerkmalen, welche im Einzelversuch als Mischprobe je Sorte ohne Wiederholungen bestimmt wurden, geht diese Komponente in der Serienauswertung in den Restfehler ein. Aus diesem formalen Grund wird die S×T(J×O) - Varianz im Weiteren nicht mit den Komponenten der Sorte × Umwelt - Varianzen, sondern im Kontext zur Fehlervarianz dargestellt und diskutiert. In Tab. 28 wird dies durch verbundene Zellen deutlich, in welchen die S×T(J×O) - Varianz und die Fehlervarianz subsumiert ist - hier kam also das Modell 7a zur Anwendung. Durch die Ausweisung der Varianzkomponenten als Variationskoeffizienten soll Vergleichbarkeit über alle Merkmale erzielt werden. Dies erfolgt in Analogie zum Vorgehen von Laidig et al. (2008).

Alle hier dargestellten Varianzen stehen in Bezug zum Faktor ‚Sorte', sind Interaktionsvarianzen mit dem Faktor Sorte bzw. es handelt sich um Fehlervarianzen. Reine Umweltvarianzen werden im Unterschied zum Vorgehen bei Laidig et al. (2008) nicht geschätzt (entsprechend der Modelle 7a und 7b).

Sehr geringe relative Gesamtvarianzen (Summe aller Varianzkomponenten in Tab. 28) weisen einige Qualitätsmerkmale auf, insbesondere die ‚Sortierung > 2.5 mm (%)' und das Hektolitergewicht bei Sommerbraugerste (SORTGR25 und HLG), die Energiekonzentration bei Silomais (NELKGX), der Ölgehalt bei Raps (FETT91) und der Trockensubstanzgehalt bei Körnermais (TS_KORN). Hohe relative Gesamtvarianzen zeigen der Ertrag bei Lupine, der Minderertrag bei Triticale und unter den Qualitätsmerkmalen der Glucosinolatgehalt bei Raps (GLUC91) sowie die Tausendkornmasse bei Lupinen (TKG_86).

Sorte × Umwelt - Varianz (SU in Tab. 28):

Die Sorte × Umwelt - Varianz hat bei Ertragsmerkmalen mit einem durchschnittlichen Variationskoeffizienten um 5% eine herausragend große Bedeutung. Einen besonders hohen Wert weist hier die Lupine auf - eine Pflanzenart die überwiegend für extreme Trockenstandorte empfohlen wird und daher auch in der Sortenprüfung überwiegend auf Extremstandorte gestellt wird. Erbse, Kartoffel und Hafer zeigen einen hohen Anteil der Sorte × Umwelt - Varianz an der Gesamtvarianz (Summe der drei in Tab. 28 aufgeführten Varianzen). Bei der Merkmalskategorie ‚Minderertrag' (bei geringer Intensität) fallen Roggen und Raps mit geringen Sorte × Umwelt - Varianzen auf. Dagegen zeigt Wintertriticale beim Minderertrag eine hohe Sorte × Umweltinteraktionen - Ursache dürften die aktuell starken Brüche und Veränderungen im Resistenzgeschehen dieser Pflanzenart sein.

Die einzelnen Komponenten dieser Varianz werden in Tab. 29 aufgeführt.

Sorte × Typ (Jahr × Ort) - Varianz (ST):

Auf die relativen Größenordnungen der S×T(J×O) - Varianz kann in Tab. 28 indirekt aus der Differenz zwischen den Spalten SE^2 und ($ST+SE^2$) geschlossen werden. Sie sind erstaunlicherweise z.T. nicht unerheblich. Bereits in unmittelbarer Nachbarschaft angelegte Versuche, bei Bewirtschaftung durch die gleichen Personen mit gleichem Management und bei gleicher Witterung treten also offensichtlich unterschiedliche Sortenrelationen bzw. -rangfolgen auf, welche nicht nur im Rest-Fehler begründet sind. Einen besonders hohen relativen Anteil an der Gesamtvarianz weist diese Varianzkomponente beim Ertrag des Rapses auf.

Bei gut besetzen Merkmalen war durchweg eine von null verschiedene Varianz schätzbar. In einigen Fällen wurde diese Varianz allerdings auf null geschätzt (in Tab. 28 erkennbar an den Fällen mit $SE^2=ST+SE^2$). Da es sich in diesen Fällen aber regelmäßig um Merkmale mit geringem Datenumfang, also einer vergleichsweise kleinen Anzahl von beobachteten Stufen für diesen Effekt handelt, sollte davon ausgegangen werden, dass dies eher in dem zu geringen Datenumfang als in einer tatsächlich in der Grundgesamtheit fehlenden Varianz begründet war.

Tab. 28: Sorte × Umwelt - Interaktionsvarianzen im Vergleich zu Fehlervarianzen der Sortenmittelwerte aus den Einzelversuchen

Pflanzenart	Merkmal	Median des Merkmals	Einheit	Variationskoeffizient (%) *		
				SU	SE²	ST+SE²
WW	Ertrag	96,7	dt/ha	4,0	2,6	3,0
WS	Ertrag	61,0	dt/ha	5,2	3,2	3,2
GW	Ertrag	86,7	dt/ha	4,3	3,0	3,7
GS_F	Ertrag	56,8	dt/ha	4,8	3,4	4,8
GS_B	Ertrag	62,9	dt/ha	3,8	3,6	3,6
RW	Ertrag	82,5	dt/ha	4,0	3,2	3,9
TIW	Ertrag	87,9	dt/ha	5,4	3,9	4,8
HA	Ertrag	63,8	dt/ha	4,9	2,1	2,1
RAW	Ertrag	51,9	dt/ha	3,2	3,0	4,2
EF	Ertrag	48,4	dt/ha	6,3	3,1	3,1
LUB	Ertrag	28,5	dt/ha	10,5	4,2	6,5
MK	Ertrag	107	dt/ha	4,2	2,4	3,2
MS	Ertrag	179	dt/ha	3,3	2,9	3,4
KA	Ertrag	557	dt/ha	7,6	2,9	3,9
Mittelwert Merkmalskategorie *Ertrag*				**5,1**	**3,1**	**3,8**
WW	MinderE	§	dt/ha	2,4	3,8	4,1
WS	MinderE		dt/ha	4,6		1,6
GW	MinderE		dt/ha	2,4	4,3	4,7
GS_F	MinderE		dt/ha	3,1		2,3
GS_B	MinderE		dt/ha	2,5	5,0	5,0
RW	MinderE		dt/ha	0,6	4,1	4,6
TIW	MinderE		dt/ha	7,5	5,6	7,5
RAW	MinderE		dt/ha	0,7		5,7
Mittelwert Merkmalskategorie Minderertrag				**3,0**		**4,4**
MS	TSNIGEPF	33,8	%	2,5	2,1	2,7
MK	TS_KORN	71,9	%	1,5	0,6	0,8
RAW	GLUC91	14,3	%	5,9		7,9
RAW	FETT91	42,5	%	1,3		1,5
EF	RP__86	19,9	%	2,3		1,6
LUB	RP__86	30,0	%	3,8		1,8
MS	STAEGEH	32,8	%	3,4	3,5	4,1
KA	STAEGEKN	19,2	%	3,6		3,5
GS_B	SORTGR25	96,4	%	0,9		1,1
GS_B	RP__I_TM	11,5	%	2,7		3,1
GS_B	TKG_LUTR	49,3	g	3,9		3,2
GS_B	HLG	70,7	kg	1,1		1,3
LUB	TKG_86	154	g	5,2		4,4
MS	NELKGX	6,5	MJ	0,8	1,2	1,3
Mittelwert Merkmalskat. *Qualität [%]* und *sonst. Qualität*				**2,8**		**2,7**

SU = S×O + S×J + S×J×AG + S×O×J (für Sorte × Umwelt; ohne S×AG)

ST = S×T(O×J) (für Sorte × Typ an O×J-Kombinationen); nur bei Modell 7b isolierbar von $SE_{Versuch}^2$

SE² hier für $SE^2_{Versuch}$ = mittlere Fehlervarianz der Sortenmittelwerte in Einzelversuchen

* Variationskoeffizient = √(Varianzkomponente) * 100 / Median

§ für ‚Minderertrag' ist bei Bestimmung des s% der ‚Ertrag' (nicht der ‚Minderertrag') sachliche und rechnerische Bezugsgröße (Minderertrag der Stufe I bemisst sich in fachlicher Bewertung in Relation zum Ertrag in der praxisüblichen Intensitätsstufe II)

Tab. 29: Relativer Anteil der Komponenten der Sorte × Umwelt - Varianz an dieser Umweltvarianz

Pflanzenart	**Merkmal**	**Sorte × Umwelt**			
		S × J	**S × J × AG**	**S × O**	**S × J × O**
		Varianzen relativiert auf Gesamtvarianz$_{(Sorte \times Umwelt)}$ (%)			
WW	Ertrag	33	11	9	48
SW	Ertrag	20	0*	15	65
GW	Ertrag	27	10	16	48
GS_F	Ertrag	22	0*	26	53
GS_B	Ertrag	21	3	6	71
RW	Ertrag	65	0*	12	23
TIW	Ertrag	45	5	4	46
HA	Ertrag	26	0*	11	63
RAW	Ertrag	26	13	6	56
EF	Ertrag	45	1	7	47
LUB	Ertrag	20	0*	8	72
MK	Ertrag	44	16	0*	40
MS	Ertrag	36	1	18	45
KA	Ertrag	39	0*	22	39
Mittelwert *Ertrag*		**33**	**4**	**11**	**51**
WW	MinderE	19	7	12	62
SW	MinderE	14	0*	14	72
GW	MinderE	9	5	14	73
GS_F	MinderE	0*	0*	43	57
GS_B	MinderE	14	14	0*	73
RW	MinderE	0*	50	50	0*
TIW	MinderE	55	0*	13	32
RAW	MinderE	66	0*	34	0**
Mittelwert *Minderertrag*		**22**	**9**	**22**	**46**
MS	TSNIGEPF	42	9	15	34
MK	TS_KORN	36	0*	5	59
RAW	GLUC91	49	0*	1	50
RAW	FETT91	48	0*	21	31
EF	RP__86	50	0*	1	50
LUB	RP__86	31	0*	14	55
MS	STAEGEH	45	9	20	26
KA	STAEGEKN	27	0*	0*	57
GS_B	SORTGR25	28	0*	0*	72
GS_B	RP__I_TM	69	12	0*	19
GS_B	TKG_LUTR	5	28	0*	67
GS_B	HLG	10	15	1	75
LUB	TKG_86	13	0*	6	81
MS	NELKGX	36	9	18	37
Mittelwert *Qualität [%]* und *sonst. Qualität*		**35**	**6**	**8**	**51**

* keine signifikante Abweichung von null (durchweg bei geringem Datenumfang)

Fehlervarianz der Sortenmittelwerte aus Einzelversuchen (SE^2 für $SE_{Versuch}^2$ in Tab. 28):

Diese Fehlervarianz liegt komplett nur für die Kategorie ‚Ertrag' vor. Ihre Variationskoeffizienten unterscheiden sich hier nicht außerordentlich, sie liegen um 3%.

Bei den meisten Merkmalen ist diese Fehlervarianz ($SE^2_{Versuch}$) nicht von der S×T(J×O) - Varianz isolierbar (ST als Kürzel für S×T(J×O) in Tab. 28). Die Summe beider Varianzen und der zugehörige Variationskoeffizient (ST+SE^2 in Tab. 28) charakterisieren gewissermaßen die Gesamt-Fehlervarianz an Ort×Jahr-Kombinationen. Diese ist im Verhältnis zur Sorte × Umwelt - Interaktionsvarianz beim Minderertrag von Sommerweizen und beim Rohproteingehalt von Lupinen besonders gering.

<u>Komponenten der Sorte × Umwelt - Varianz:</u>

Der relative Anteil der Komponenten der Sorte × Umwelt - Varianz an dieser Varianz ist für alle Pflanzenart × Merkmal - Kombinationen in Tab. 29 ausgewiesen. Nachfolgend werden diese Komponenten in der Reihenfolge ihrer relativen Bedeutung besprochen.

Sorte × Jahr × Ort - Varianz:

Diese Varianzkomponente macht über alle Merkmale etwa die Hälfte der gesamten Sorte × Umwelt - Varianz aus und ist damit in den meisten Fällen die größte Varianzkomponente.

Sorte × Jahr - Varianz und Sorte × Jahr × Anbaugebiet - Varianz:

Diese beiden Varianzkomponenten beschreiben die Interaktion von Sorte und Jahr im gesamten einbezogenen Raum einerseits (S × J) bzw. innerhalb von Anbaugebieten in diesem ‚Großraum' andererseits (S × J × AG). Mit durchschnittlich 37% stellen diese beiden Varianzkomponenten zusammen überwiegend die zweitgrößte Komponente der Sorte × Umwelt - Varianz dar. Bei Ertrag und Minderertrag von Roggen, Minderertrag bei Wintertriticale und Raps sowie Rohproteingehalt und Tausendkorngewicht bei Sommerbraugerste ist es sogar die dominierende Komponente. Insbesondere ist von Interesse, dass die Interaktion der Sorten mit Jahren in der Regel deutlich größer ist als mit Orten innerhalb von Anbaugebieten (in 31 von 36 Fällen). Nur bei der Merkmalskategorie ‚Minderertrag' fällt diese Relation nicht so deutlich und weniger einheitlich aus. Außer bei den Merkmalen Minderertrag bei Roggen und Tausendkornmasse bei Sommerbraugerste überwiegt die S × J - Varianz deutlich über der S × J × AG - Varianz.

Sorte × Ort - Varianz:

Die S × O - Varianzen (innerhalb von Anbaugebieten!) hatten überwiegend die geringste Größenordnung.

Eine weitergehende, für den Regionalisierungsansatz der *Hohenheim-Gülzower-Serienauswertung* wesentliche Frage ist die Analyse der Sorten-Varianz und ihrer Anbaugebietsspezifik:

Sortenvarianz (S) und ihre Anbaugebietsspezifik (S × AG) (Tab. 30):

Der Variationskoeffizient für die Sorten-Variabilität (innerhalb von Sorten-Gruppen) liegt bei Ertragsmerkmalen bei 2,6 % (Median der s% über die Merkmale in Tab. 30). Bei Sommerfuttergerste (sehr kleiner Datensatz!) konnte bei Ertrag und Minderertrag keine Varianz nachgewiesen werden, weder insgesamt noch regional (zwischenzeitlich wurde für den Zuständigkeitsbereich der LFA bereits entschieden, eine eigenständige regionale Prüfung einzustellen; diese Entscheidung erweist sich aufgrund der hier vorgelegten Ergebnisse als begründet). Unter den Ertragsmerkmalen zeigen Kartoffeln und Hafer die höchsten Variationskoeffizienten. In der Merkmalsgruppe *Minderertrag* fällt Triticale mit dem höchsten Variationskoeffizienten auf, wahrscheinlich aufgrund der bereits oben diskutierten Umbrüche im Resistenzgeschehen bei dieser Pflanzenart. Besonders hohe Variationskoeffizienten weisen die Qualitätsmerkmale Stärkegehalt bei Kartoffeln und Glucosinolatgehalt bei Raps auf.

Von besonderem Interesse sind die Sorte × Anbaugebiet - Interaktionsvarianzen (S × AG) in ihrer Relation zur Sortenvarianz (S). Die unterschiedlich zusammengefassten Zellen in Tab. 30 sind den Pflanzenart spezifischen Regionalisierungen geschuldet (entsprechend Tab. 15 in 5.1.3). Bei fast allen Pflanzenart × Merkmal - Kombinationen war für jeweils einzelne Anbaugebiete keine Sortenspezifik nachweisbar, die S × AG - Varianz wurde hier auf null geschätzt (u.U. zufallsbedingt vor dem Hintergrund nicht ausreichender Schätzgenauigkeit). Trotzdem gab es bei denjenigen Pflanzenart × Merkmal - Kombinationen, bei denen mehrere Anbaugebiete gebildet wurden, in 20 von 24 Fällen jeweils mindestens ein Anbaugebiet mit Sortenspezifik ($Varianz(S\times\mathrm{AG}) \neq 0$). Allein dies ist schon ein Hinweis darauf, dass die Bildung von Anbaugebieten in diesen räumlichen Dimensionen prinzipiell sinnvoll, wenn auch im Einzelfall überprüfenswert ist.

Unter den Ertragsmerkmalen treten einzelne besonders hohe Interaktionsvarianzen bei Körnermais, Erbsen, Winterweizen und Raps auf. Des Weiteren zeigen sich hohe Interaktionsvarianzen beim Minderertrag-Wintergerste, bei Sommerbraugerste in den Qualitätsmerkmalen Rohproteingehalt und ‚Sortierung > 2.5 mm (%)' sowie beim Trockensubstanzgehalt bei Körnermais.

Während die westlich vom Zielgebiet gelegenen Nachbargebiete häufig starke Spezifik zeigten, kann dies für südlich ausgerichtete (im wesentlichen *D-Süd*) nur beim Kornertrag von Raps klar ausgewiesen werden.

Die hier diskutierten S - und S × AG - Varianzen sind die Basis für die Berechnung der genetischen Korrelationen und für die Gewichtung von Ergebnissen aus Nachbargebieten. Eine vertiefte Diskussion der genetischen Korrelationen zwischen Zielgebieten und einbezogenen Nachbargebieten erfolgt im Abschnitt 6.3.1.

Tab. 30: Sortenvarianz und ihre Anbaugebietsspezifik

Pflanzenart	**Merkmal**	**Varianzkomponenten**						
		S	**S × AG** relativ zu **S** ** (%)					
		s(%)* **	**spezifisch für AG**					
			D-Nord	**MV-Süd**	**Ṣ**	**ṢẆ**	**Ẉ**
WW	Ertrag	2,7	0*	0*	0*	0*	**41**
SW	Ertrag	2,2	entfällt				
GW	Ertrag	2,6	**8**		**1**		0*
GS_F	Ertrag	0*	entfällt				
GS_B	Ertrag	3,9	0*			0,2	
RW	Ertrag	2,5	0*		0*		
TIW	Ertrag	2,7	**6**		0*		
HA	Ertrag	6,6	0*			0*	
RAW	Ertrag	2,5	0*	**28**	**27**	0*	0*
EF	Ertrag	1,9	**100**			0*	
LUB	Ertrag	4,5	entfällt				
MK	Ertrag	1,9	0*			**135**	
MS	Ertrag	3,2	**13**		0*		**21**
KA	Ertrag	6,9	entfällt				
WW	MinderE	1,6	**26**		0*	0*	
SW	MinderE	0,9	entfällt				
GW	MinderE	1,2	**118**		0*		**9**
GS_F	MinderE	0*	entfällt				
GS_B	MinderE	1,2	**18**			0*	
RW	MinderE	0,8	0*		0*		
TIW	MinderE	4,9	**19**		0*		
RAW	MinderE	0,7	entfällt				
MS	TSNIGEPF	1,5	0*		0*		**34**
MK	TS_KORN	0,6	0*			**78**	
RAW	GLUC91	14,1	entfällt				
RAW	FETT91	2,4	entfällt				
EF	RP__86	2,6	**7**			0*	
LUB	RP__86	2,8	entfällt				
MS	STAEGEH	4,2	0*		**1**		**3**
KA	STAEGEKN	14,5	entfällt				
GS_B	SORTGR25	0,4	**49**			**77**	
GS_B	RP__I_TM	1,5	**195**			0*	
GS_B	TKG_LUTR	3,6	0*			0*	
GS_B	HLG	2,2	0*			**9**	
LUB	TKG_86	9,9	entfällt				
MS	NELKGX	1,4	**3**		0*		**6**

* keine signifikante Abweichung von null

** spezifisch für bis zu fünf Anbaugebiete (zwei Zielgebiete und drei Nachbargebiete)

*** $s\% = \sqrt{Varianz} * 100 / Median$; s. Tab. 28

6.2.2 Effekte der langjährigen Auswertung und der Einbeziehung vorgelagerter Versuche auf Mittelwerte und ihre Genauigkeit

Traditionell wurden Landessortenversuche in den einzelnen Bundesländern als dreijährige Versuchsserien ausgewertet. Dabei wurden orthogonale Strukturen geschaffen, die nicht unter Verwendung komplexer Modelle sondern mit Methoden der Tabellenkalkulation verdichtet wurden. Dabei wurden i.d.R. orthogonale Vergleiche jeweils innerhalb aller drei-, zwei- bzw. einjährig geprüften Sorten angestellt. Vorgelagerte Versuche wie Wertprüfungen und EU-Sortenversuche wurden nicht einbezogen, da die Unbalanciertheit dies methodisch noch nicht zuließ. Mit den hier beschriebenen Methoden kann die Datenbasis erheblich erweitert bzw. besser genutzt werden, womit allerdings zwangsläufig eine massive Zunahme der Unbalanciertheit einhergeht. Entsprechend der in 6.2.1 vorgestellten Relationen der Varianzkomponenten (hohe S × J - Varianzen) ließ gerade die Einbeziehung möglichst vieler Jahre deutliche Vorteilseffekte erwarten.

Nachfolgend soll - exemplarisch für den Kornertrag bei Winterweizen - zunächst untersucht werden, wie sich dies auf die Genauigkeit (SE_{Serie}, Definition s. 5.3.5) bei Sorten unterschiedlichen Prüfstatus‘ (Definition s. Tab. 8) auswirkt. Des Weiteren soll eingeschätzt werden, wie stark sich die Neuerungen auf die Mittelwertschätzungen selbst - wiederum je nach Prüfstatus - auswirken können.

Für diese Analysen wurden Szenarien (Sz) unter 5.5 definiert und in diesem Abschnitt typische Auswertungsstrategien (Sz 1 bis Sz 4) in Tab. 31 nochmals zum besseren Verständnis systematisiert. Dabei stellt das Sz 1 wiederum die derzeitige durch den Autor eingeführte Vorgehensweise in Mecklenburg-Vorpommern dar, während Szenarium 5 eine nach wie vor verbreitete Vorgehensweise in denjenigen Bundesländern, welche noch nicht auf die *Hohenheim-Gülzower-Serienauswertung* umgestellt haben, ist. Die Szenarien 3 und 4 sind markante Zwischenstufen, bei denen zum einen die vorgelagerten Prüfungen ohne Verlängerung des Auswertungszeitraumes einbezogen werden (Sz 3), zum anderen die Anzahl einbezogener Jahre erhöht wird, aber die vorgelagerten Versuche nicht genutzt werden (SZ 4).

Um für den Vergleich der Szenarien einheitliche Randbedingungen zu gewährleisten, erfolgte nicht für jedes Szenarium eine eigenständige Schätzung des Transformationsparameters und der Varianzkomponenten, sondern es wurden für alle Szenarien gleichermaßen die in 6.1.1 und 6.2.1 dargestellten Parameter eingesetzt.

Tab. 31: Definition von Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen

Szenarien	**Anzahl einbezogener Jahre**	**Einbeziehung vorgelagerter Versuchstypen (WP, EUSV u.a.)**
Sz 1	6	ja
Sz 3	3	ja
Sz 4	6	nein
Sz 5	3	nein

WP Wertprüfung EUSV EU-Sortenversuch
Definition der Szenarien: s. 5.5

Die Auswirkung der Szenarien auf die Genauigkeit der geschätzten Serienmittelwerte (SE_{Serie}) ist in Tab. 32 und in der entsprechenden Abb. 26 gezeigt.

Tab. 32: Auswirkung von Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen

Prüfstatus (Definition s. Tab. 8)	**Anzahl Sorten**	**Sz 1**	**Sz 3**	**Sz 4**	**Sz 5**
		SE_{Serie} [dt/ha]			
alte Sorten	7	1,26	1,77	1,44	1,83
LSV5-8	8	1,12	1,50	1,14	1,52
LSV4	2	1,12	1,41	1,22	1,41
LSV3	3	1,21	1,50	1,47	1,52
LSV2	6	1,36	1,57	1,69	2,01
LSV1	4	1,65	1,77	2,70	2,70
OV_WP	3 *	1,78	1,93	3,02	3,02
OV_EU	2	1,94	1,94	4,10	4,10

* von 12 Sorten dieses Prüfstatus‘ konnten mit Sz4 und Sz5 mangels Datenbasis nur für drei Sorten Mittelwerte geschätzt werden, während dies bei Sz 1 und Sz 3 für alle 12 Sorten möglich wäre; hier wurde über alle vier Szenarien nur der Kern dieser drei Sorten orthogonal ausgewertet

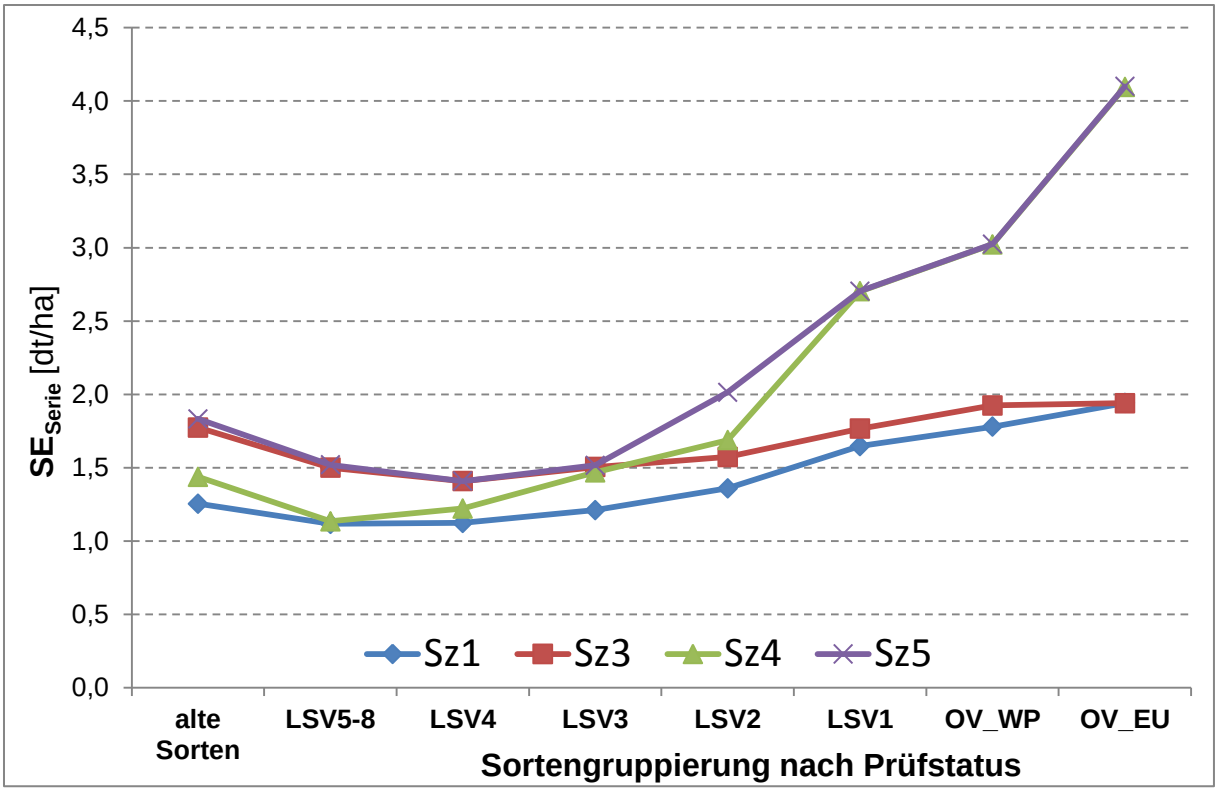
Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: sechs Versuchsjahre zur Mittelwertschätzung und Einbeziehung vorgelagerter Versuche

Sz 3: nur drei Versuchsjahre einbezogen, im Übrigen wie Sz 1

Sz 4: keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1

Sz 5: nur drei Versuchsjahre und keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1

Definition der Szenarien: s. 5.5



Definition des Prüfstatus‘ s. Tab. 8

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: sechs Versuchsjahre zur Mittelwertschätzung und Einbeziehung vorgelagerter Versuche
Sz 3: nur drei Versuchsjahre einbezogen, im Übrigen wie Sz 1
Sz 4: keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1
Sz 5: nur drei Versuchsjahre und keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1
Definition der Szenarien: s. 5.5

Abb. 26: Einfluss der Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen

Sz 1 ist über alle Sorten unterschiedlichen Prüfstatus‘ (Def. s. Tab. 8) die Bestvariante. Schon sehr junge Sorten, die erst in den LSV aufgenommen werden sollen, weisen vergleichsweise niedrige SE_{Serie} auf (im Sinne einer ersten Früheinschätzung), welche beim konservativen Sz 5 erst nach dem LSV2 unterschritten werden. Dies ist der Einbeziehung der vorgelagerten Versuche geschuldet, wodurch diese Sorten dann bereits mit vierjährigen (nach *OV_WP*) bzw. dreijährigen (nach *OV_EU*) Ergebnissen einfließen. Im Ergebnis wirkt sich dies erheblich auf die Schätzgenauigkeit aus, obgleich die Anzahl Versuche je Jahr in den vorgelagerten Versuchen noch deutlich niedriger liegt, als nach der Aufnahme in den LSV. Aber auch ‚*alte Sorten*‘, die nicht mehr im aktuellen LSV stehen, können noch mit hoher Vorhersagegenauigkeit bewertet werden, da durch die langjährige Auswertung viele zurückliegende Versuchsergebnisse einbezogen werden können. Lediglich beim Prüfstatus ‚OV_EU‘ erreicht ein anderes Szenarium, das Sz 3, etwa die gleiche Vorhersagegenauigkeit. Ursache hierfür ist, dass bei diesem Prüfstatus per Definition erst drei Prüfjahre vorliegen und somit diese beiden Szenarien für diese Sorten eine identische Datenbasis aufweisen.

Sz 3 zeichnet sich also erwartungsgemäß dadurch aus, dass die sehr jungen Sorten durch die Einbeziehung der vorgelagerten Versuche vergleichsweise schon relativ gut eingeschätzt werden können, vergleichbar mit Sz 1. Eine längerjährige

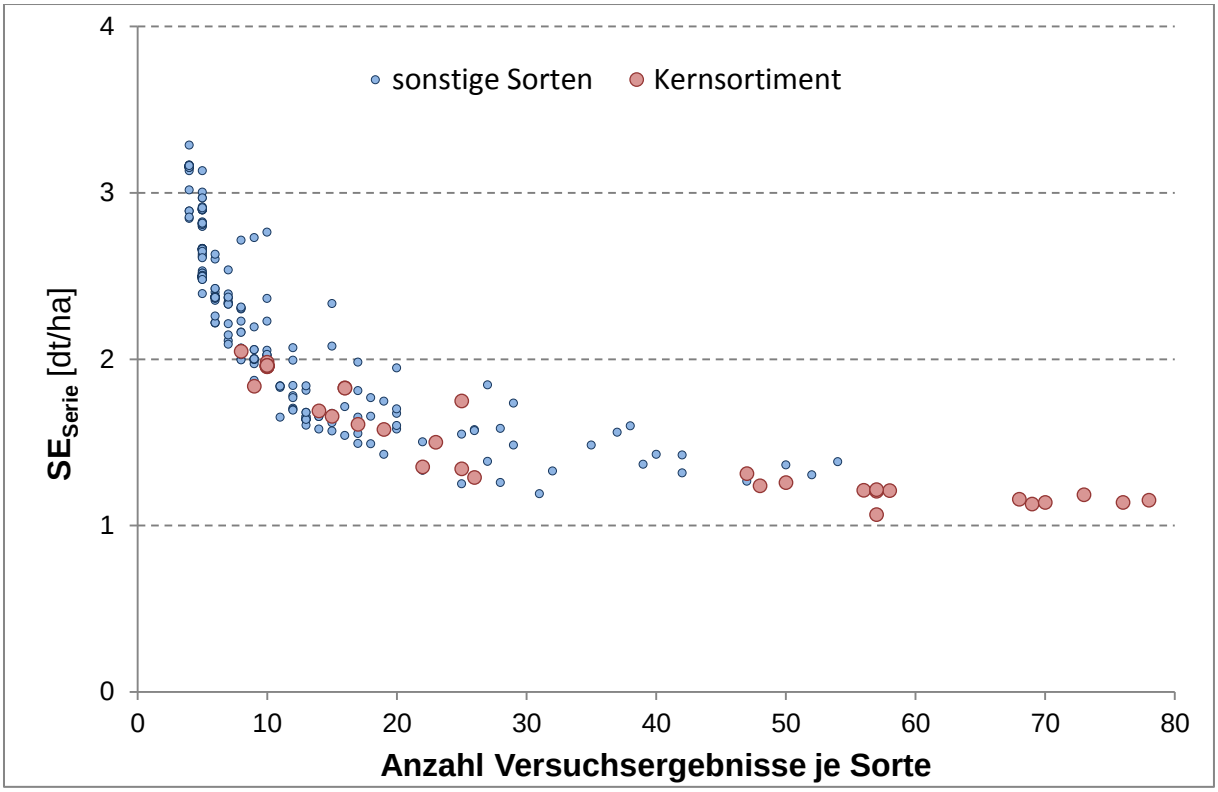
Auswertung erbringt bei diesen Sorten noch keine (bei *OV_EU*) bzw. geringe (bei *OV_WP*) Vorteile. Mit zunehmender Prüfdauer verschlechtert sich die relative Güte von Sz 3 gegenüber Sz 1. Im Prüfstatus LSV2 und LSV3, wenn die Entscheidung zur regionalen Empfehlung ansteht, ist die Genauigkeit deutlich eingeschränkt. Zu keinem Zeitpunkt wird mit Sz 3 eine Präzision erreicht, die Sz 1 bereits nach LSV2 erreicht. Hier schlägt sich das willkürliche Abschneiden der Prüfungen, die länger als drei Jahre alt sind, deutlich negativ nieder. Die Einschätzung ‚*alter Sorten*' ist besonders beeinträchtigt, da besonders diese eine Einbeziehung zurückliegender Jahre erfordert.

Sz 4 verhält sich - auch erwartungsgemäß - gegenläufig zu Sz 3. Hier weisen die bereits lange geprüften Sorten nahezu ebenso geringe Fehler (SE_{Serie}) auf wie nach Sz 1. Beim Prüfstatus LSV5-8 ist der Abstand zu Sz 1 nur noch minimal. Bei den langjährig geprüften Sorten schlagen sich viele LSV-Jahre positiv nieder, während die vorgelagerten Versuche einen immer geringer werdenden Beitrag leisten bzw. zeitlich verdrängt werden. Dagegen weisen in Sz 4 die jungen Sorten außerordentlich große Fehler (SE_{Serie}) auf bzw. sind gar nicht schätzbar (s. Anmerkung in Tab. 32). Nach LSV3 wird mit Sz 4 noch nicht die Genauigkeit erreicht, die Sz 1 bereits nach LSV2 aufweist. Nach LSV4 erreicht Sz 4 die Präzision von Sz 1 nach LSV3.

Sz 5 vereinigt die jeweiligen Nachteile von Sz 3 und Sz 4 und wird somit für jeden Prüfstatus zur schlechtesten Variante. Nach LSV2 wird noch nicht die Präzision erreicht, die bei Sz 1 bereits vor der Aufnahme in den LSV vorliegt. Eine inzwischen in Deutschland übliche vorläufige Sortenempfehlung nach LSV2 ist mit diesem Szenarium noch zu unsicher und kaum zu verantworten.

Die Genauigkeit der geschätzten Sortenmittelwerte steigt erwartungsgemäß mit zunehmender Ergebnisdichte (Abb. 27; für Szenarium 1). Die Abbildung darf allerdings nicht im Sinne eines zu erwartenden durchgängigen funktionalen Zusammenhangs interpretiert werden. Die Wertepaare sind maßgeblich durch systematisch differenzierte Effekte infolge unterschiedlichen Prüfstatus' der Sorten geprägt. Hinzu kommt insbesondere die formale Begrenzung des Auswertungszeitraumes auf sechs Jahre. Im unteren Teil des Definitionsbereiches bis zu etwa 20 Versuchsergebnissen je Sorte ist der scheinbare Kurvenverlauf vor allem durch die zunehmende Anzahl Jahre bis zu maximal sechs Jahren geprägt (s.a. Tab. 10 und Abb. 5). Im weiteren Verlauf hat es - trügerisch - den Anschein, als nähere sich die Punktewolke mit steigender Versuchsanzahl einer Asymptote oberhalb der als Asymptote zu erwartenden ‚Nulllinie'. Eine zunehmende Versuchsanzahl im oberen (rechten) Teil des Definitionsbereiches ist bei diesen Sorten aber ausschließlich in großer Anzahl Orte je Jahr und in der häufigen Mehrfachprüfung an gleichen Ort × Jahr - Kombinationen (in unterschiedlichen Versuchstypen) bei willkürlich auf sechs Jahre begrenzter Jahresanzahl begründet. Um diesen Effekt nicht noch weiter zu übersteigern, wurden alle Sorten nicht in die Darstellung einbezogen, die in mindestens einem Jahr den Status einer Verrechnungs- oder Vergleichssorte inne hatten (VRS bzw. VGL). Diese formale Begrenzung der Jahre begrenzt im Kontext mit der gezeigten hohen Sorte × Jahr - Interaktionsvarianz die Möglichkeit, den Standardfehler weiter spürbar zu senken. Eine extrem einseitige Erhöhung der Anzahl Orte (und insbesondere der häufigen Mehrfachprüfung an gleicher Ort × Jahr - Kombination in verschiedenen Versuchstypen) - wie sie sich im Zuge des Prüfsystems für einzelne wenige Sorten automatisch ergab - bringt keinen sichtbaren Zugewinn an Genauigkeit, solange sich die Anzahl Jahre nicht auch erhöht. Sehr hohe Versuchsanzahlen betreffen besonders Sorten mit dem Status einer Verrechnungs- oder Vergleichssorte. Die

häufige Prüfung dieser Sorten hatte nicht das Ziel einer Präzisionserhöhung für diese Sorten, sondern diente der Absicherung einer halbwegs orthogonalen Kernstruktur, die die zusammenfassende Auswertung vieler Versuche mit erheblich unterschiedlicher Sortenzusammensetzung ermöglicht. Sorten mit sehr hoher Versuchsanzahl bei harmonischem Verhältnis der Anzahl Orte zur Anzahl Jahre (N_{Jahre} >> 6) fehlen in der Darstellung wie im Datensatz überhaupt.



Sorten, die im Prüfzeitraum mindestens in einem Jahr den Status einer Verrechnungssorte (VRS) oder Vergleichssorte (VGL) inne hatten (16 Sorten), sind hier nicht aufgenommen

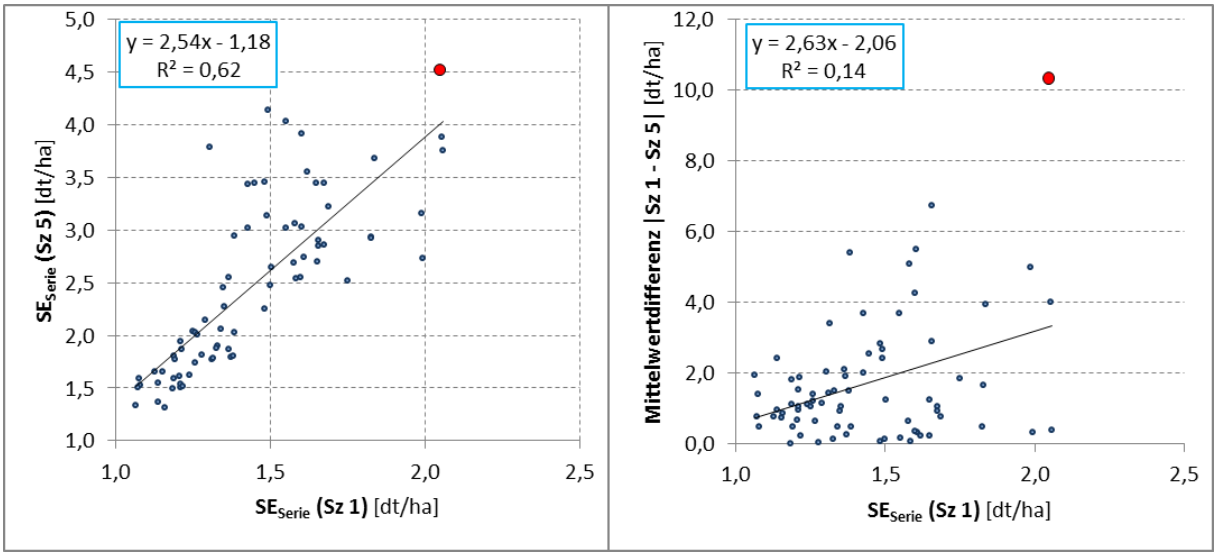
Abb. 27: Beziehung zwischen Datenumfang und Genauigkeit bei Begrenzung des Auswertungszeitraumes auf sechs Jahre; Kornertrag Winterweizen

Die Schwankung der SE_{Serie} bei Sorten mit gleicher Anzahl Versuchsergebnisse ist der differenzierten Datenstruktur (Orte, Jahre, Typen) geschuldet. Diese Schwankung ist im Bereich der rapiden Änderung des Anstiegs zwischen etwa 20 und 40 Versuchsergebnissen je Sorte besonders groß. Bei Begrenzung auf sechs Auswertungsjahre wurde in keinem Fall ein Vorhersagefehler unter 1,0 dt/ha erreicht.

Nachfolgend wird für die beiden Eck-Szenarien Sz 1 und Sz 5 die Wirkung auf die Mittelwerte für alle auch mit Sz 5 schätzbaren Sorten (N=75) analysiert und ebenfalls in Bezug zum Prüfstatus dargestellt. Eine für das Prüfsystem wesentliche Frage ist dabei, ob diese Differenzen im Zusammenhang mit dem Prüfstatus der Sorten (Def. s. Tab. 8) stehen. Der Prüfstatus einer Sorte kennzeichnet ihre aktuell erreichte Stufe im Verlauf des Prüfsystems (Abb. 4). Er steht in Bezug zum Faktor ‚Versuchstyp' (Tab. 11), in dem diese Sorte aktuell geprüft wird und zur Dauer der Prüfung dieser Sorte (Faktor ‚Jahr'). Ein fortschreitender Prüfstatus schlägt sich primär in einem mit der Prüfdauer abnehmenden SE_{Serie} (s. 5.3.5) nieder. Daher wurde als stetige Einflussgröße der SE_{Serie} (nach Sz 1) anstelle des Prüfstatus' in die Darstellung einbezogen.

Dazu werden in Abb. 28 (links) zum einen die Standardfehler beider Szenarien gegenübergestellt. Erwartungsgemäß lässt sich ein Zusammenhäng erkennen, der aber von deutlichen Ausschlägen im Einzelfall geprägt ist. Letzteres ergibt sich daraus, dass die Datenbegrenzung im Sz 5 die Datenstruktur der Sorten unterschiedlichen Prüfstatus‘ sehr divergierend beeinflusst (s. Ausführungen zu Abb. 26). Bei den Sorten mit der schwächsten Datenbasis (großer SE, also i.d.R junge Sorten) liegt der Standardfehler bei Sz 5 durchschnittlich etwa doppelt so hoch wie in Sz 1.

Zum anderen werden in Abb. 28 (rechts) die SE_{Serie} nach Sz 1 den Absolutwerten der Mittelwertdifferenzen zwischen beiden Szenarien gegenübergestellt. Auch hier wird deutlich, dass mit steigendem Standardfehler bzw. geringerer Versuchsbasis von Sorten die mittleren Abweichungen zwischen den Mittelwerten beider Szenarien tendenziell steigen und schnell in für die Sortenbewertung relevante Bereiche (> 1,0 dt/ha) gelangen. Für eine nicht unwesentliche Anzahl von Sorten (überwiegend Sorten, die nicht zum Kernsortiment gehören) ergeben sich Abweichungen von über 4,0 dt/ha. Dabei muss nochmals betont werden, dass die nach Sz 1 geschätzten Mittelwerte geringere SE_{Serie} aufweisen. Je geringer die Datenbasis ist, desto schwerwiegender wirken sich die willkürlichen Datenbeschränkungen durch Sz 5, im Analogieschluss aber abgeschwächt ebenso bei Sz 3 und 4, aus.



Hervorhebung eines einbezogenen Extremwertes

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: sechs Versuchsjahre zur Mittelwertschätzung und Einbeziehung vorgelagerter Versuche

Sz 5: nur drei Versuchsjahre und keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1

Abb. 28: Einfluss von Begrenzungen der Datenbasis auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen

Der Pearsonsche Korrelationskoeffizient für den linearen Zusammenhang zwischen den Schätzwerten aus beiden Szenarien weist für das Kernsortiment nur einen mäßigen Betrag auf $(r = 0{,}82)$ - für das Gros der Restsorten war eine Mittelwertschätzung mit dem Sz 5 sogar nicht möglich. Relevanter als der mittlere Zusammenhang ist aber auch hier die Frage, wie groß *im Einzelfall* Abweichungen bei Sorten in Abhängigkeit vom Prüfstadium sein können. Dazu wird in Tab. 33 dargestellt, wie sich dies bei den konkret definierten Kernsorten je nach Prüfstatus auswirkte.

Tab. 33: Relevanz von Begrenzungen der Datenbasis für die Mittelwertschätzung im Kernsortiment; Kornertrag Winterweizen

Prüfstatus der Sorten (Definition s. Tab. 8)	**N**	**|Sz 1 - Sz 5|** [dt/ha]	
		Mittelwert	**Maximum**
alte Sorten	7	1,3	1,9
LSV5-8	8	1,0	2,4
LSV4	2	1,0	1,9
LSV3	3	0,7	1,1
LSV2	6	0,6	1,5
LSV1	4	1,4	2,9
OV_WP	3 *	1,0	1,7
OV_EU	2	7,1	10,3

* von 12 Sorten dieses Prüfstatus‘ konnte Sz 5 nur für drei Sorten Mittelwerte bilden, während dies bei Sz1 für alle 12 Sorten möglich war

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: sechs Versuchsjahre zur Mittelwertschätzung und Einbeziehung vorgelagerter Versuche

Sz 5: nur drei Versuchsjahre und keine vorgelagerten Versuche einbezogen, im Übrigen wie Sz 1

Bereits die mittlere absolute Abweichung ist bei Sorten jeden Prüfstatus‘ nicht irrelevant. Noch bei langjährig geprüften Sorten (LSV5-8) wurde im Einzelfall eine erhebliche Differenz von 2,4 dt/ha gefunden, nach LSV1 sogar von 2,9 dt/ha. Zu extremen Abweichungen kann es bei sehr jungen Sorten kommen. Als Extremwert wurde im Prüfstatus *OV_EU* eine Abweichung von 10,3 dt/ha gefunden (s.a. hervorgehobener Datenpunkt in Abb. 28). Hierzu ist anzumerken, dass für diese Sorte nach Sz 5 nur ein Ergebnis vorlag, welches nicht einmal im Zielgebiet sondern in einem Nachbargebiet gemessen wurde. In der Realität wäre dieser Mittelwert nicht in Beratungsmaterial dargestellt worden. Dagegen lagen für diese Sorte nach Sz 1 bereits acht Versuchsergebnisse aus zwei Jahren vor, davon die Hälfte aus dem Zielgebiet. Mit einem daraus resultierenden Vorhersagefehler von 2,0 dt/ha kann diese Sorte mit der neuen Methode nach Sz 1 bereits grob eingeschätzt werden.

6.2.3 Effekte einer Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen auf Serien-Mittelwerte und ihre Genauigkeit

Analog zu den unter 6.1.2 und 6.2.2 ausgewerteten Szenarien wird hier untersucht, welche Effekte die Basisvariante (Sz 1) mit Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen gegenüber einer Auswertung ohne diese Gewichtung (Sz 6) hat (Definition der Szenarien s. 5.5).

Die Gewichtung war u.a. durch die ausgesprochen große Variation bei Rechtsschiefe der Einzelversuchsfehler motiviert (Abb. 29, Abb. 2). Speziell bei jungen Sorten mit relativ geringer Datenbasis wären bei einer derartigen Variation der Einzelversuchsfehler a priori Unterschiede zwischen Sz 1 (mit Gewichtung) und Sz 6 (ohne Gewichtung) zu erwarten. In Abb. 29 fallen insbesondere sehr hohe $SE_{Versuch}$ oberhalb von etwa 4,0 bis 5,0 dt/ha auf. Dabei handelt es sich überwiegend um Versuch spezifische Standardfehler, die also alle Mittelwerte jeweils eines Versuches betreffen (fette Kreise und höhere Säule). Z.T. betraf dies aber auch abweichende Standardfehler einzelner Mittelwerte innerhalb von Versuchen, i.d.R. durch Fehlwerte verursacht (Kreis nicht fett, keine Säule).

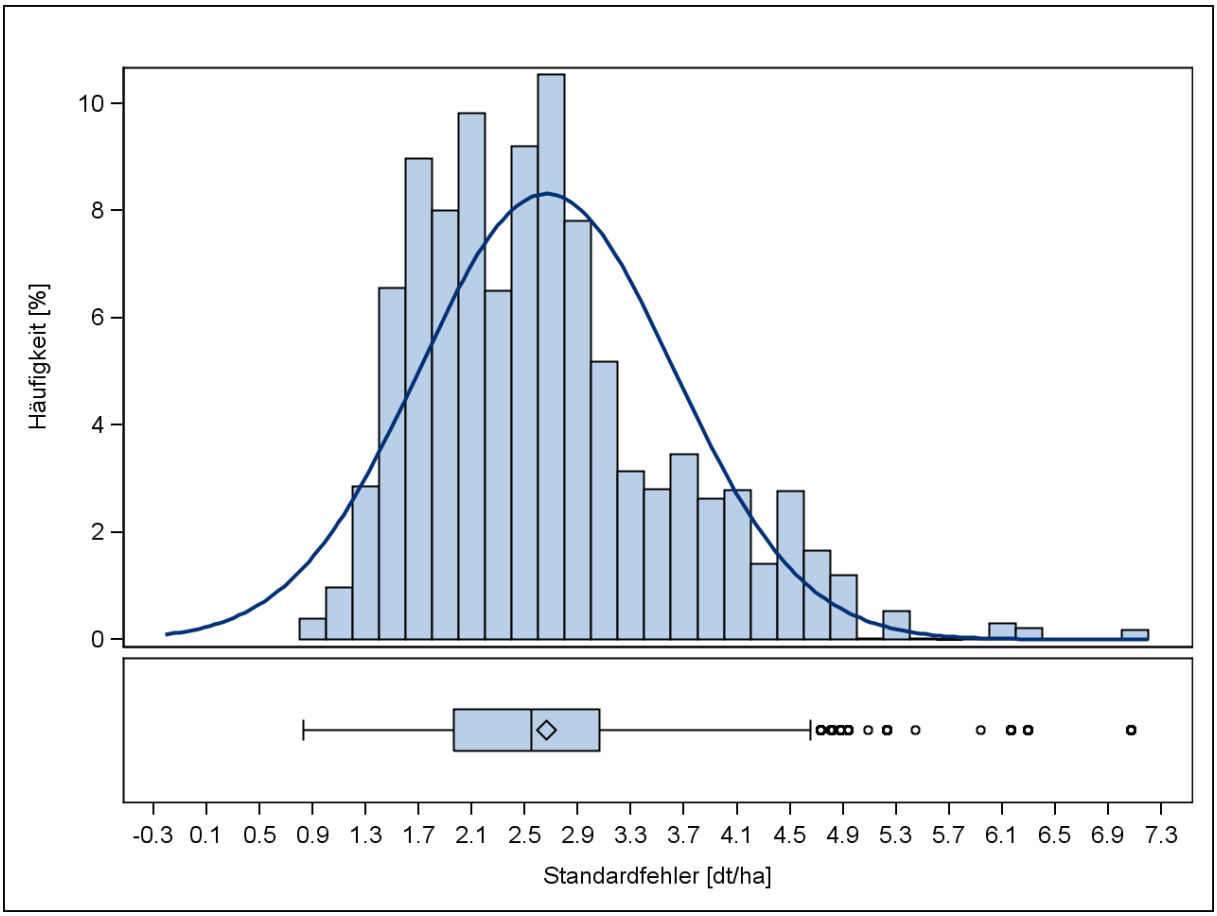
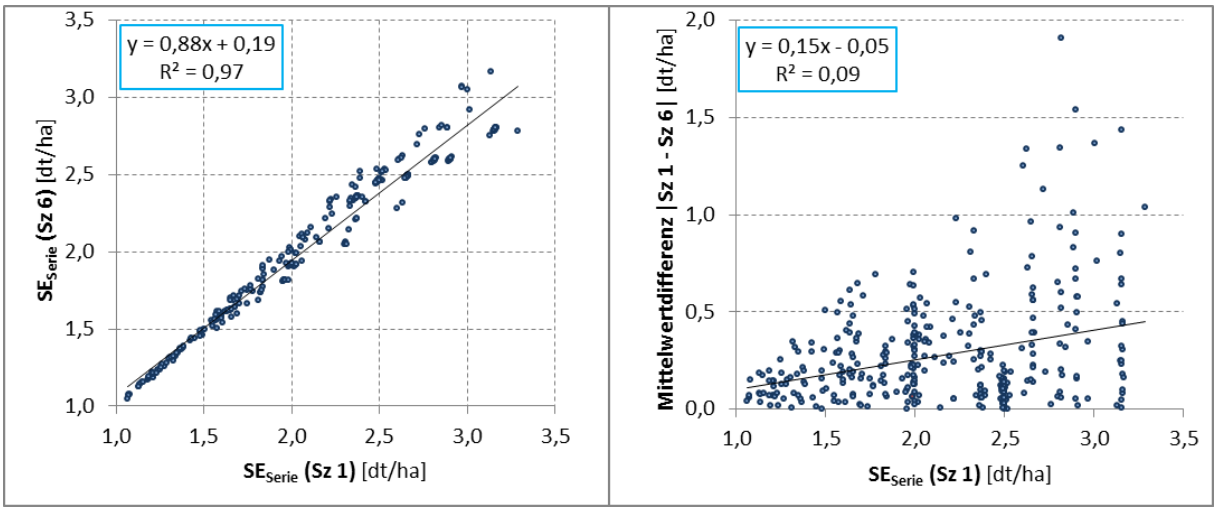


Abb. 29: Häufigkeitsverteilung der Standardfehler der Sortenmittelwerte in Einzelversuchen ($SE_{Versuch}$); Kornertrag Winterweizen

In Abb. 30 (links) ist zunächst zu ersehen, dass die Gewichtung nicht zu verringerten Fehlern der Serienmittelwerte führt. Dies ist im Mittel der Sorten auch nicht zu erwarten. Die Schwankungen um die Trendlinie nehmen bei geringerer Datenbasis

(also höherem SE) allerdings zu: die Frage der Gewichtung (ja/nein) wirkt sich bei schwächerer Datenbasis also zunehmend auf die Genauigkeit der Mittelwertschätzung für einzelne Sorten aus.



Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: mit Gewichtung …
Sz 6: ohne Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen, im Übrigen wie Sz 1

Definition der Szenarien: s. 5.5

Abb. 30: Einfluss der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte der Fehlervarianzen auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen

Zwischen beiden Szenarien treten Unterschiede in der Mittelwertbildung auf, wobei die Ausschläge mit Verringerung der Datenbasis (charakterisiert durch hohe Fehler SE_{Serie}, s. 5.3.5) deutlich zunehmen (Tab. 29 rechts). Lediglich unter den sehr wenig geprüften *sonstigen* Sorten (nicht *Kernsortiment*) traten relevante Abweichungen über 1,0 dt/ha bis knapp unter 2,0 dt/ha auf.

Der Pearsonsche Korrelationskoeffizient für den linearen Zusammenhang zwischen den Schätzwerten aus beiden Szenarien beträgt ca. eins $(r = 0{,}996)$. Relevanter als der mittlere Zusammenhang ist aber die Frage, wie groß *im Einzelfall* Abweichungen bei Sorten in Abhängigkeit vom Prüfstadium sein können. Dazu wird nachfolgend (Tab. 34) dargestellt, wie sich dies bei den Kernsorten je nach Prüfstatus (Definition s. Tab. 8) auswirkt. Alle Kernsorten weisen Differenzen zwischen den Szenarien unter 1,0 dt/ha auf.

Tab. 34: Relevanz einer Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen für die Mittelwertschätzung in der Serienauswertung; Kornertrag Winterweizen

Prüfstatus der Sorten (Definition s. Tab. 8)	**N**	**|Sz1 - Sz 6|** [dt/ha]	
		Mittelwert	**Maximum**
alte Sorten	7	0,11	0,29
LSV5-8	8	0,09	0,18
LSV4	2	0,12	0,20
LSV3	3	0,11	0,14
LSV2	6	0,22	0,51
LSV1	4	0,18	0,48
OV_WP	12	0,22	0,51
OV_EU	2	0,40	0,48

nach Auswertung transformierter Daten werden rücktransformierte Mittelwerte, also Schätzwerte für den Median verglichen

6.2.4 Diskussion

Typische Muster in der Struktur der Varianzkomponenten

Die Struktur der Varianzkomponenten bei mehrortigen Versuchsserien sollte immer im Zusammenhang mit dem einbezogenen Raum und ggf. dessen innerer Regionalisierung diskutiert werden. In der vorliegenden Arbeit liegt der Fokus auf standortkundlich begründeten Anbaugebieten innerhalb des Großraums der diluvialen Standorte Nordost-Deutschlands. Von besonderem Interesse ist nachfolgend die vergleichende Betrachtung zu den Ergebnissen von Laidig et al. (2008) zur Varianzkomponentenschätzung für den Ertrag von 30 Pflanzenarten über einen 16-jährigen Zeitraum für den Großraum Deutschland. Der Großraum Deutschland ist die regionale Basis für die nationale Sortenzulassung durch das Bundesamt, während ein Anbaugebiet die Basis für eine kleinräumiger, standortkundlich begründet regionalisierte Sortenempfehlung durch Länderdienststellen innerhalb von Deutschland ist.

In den Modellen der Hohenheim-Gülzower-Serienauswertung werden nur Interaktionsvarianzen mit dem Faktor Sorte geschätzt, keine reinen Umweltvarianzen. Dahingegen erfolgte bei Laidig et al. (2008) auch eine Schätzung der Umweltvarianzen, welche erwartungsgemäß ausgesprochen dominant gegenüber der Sortenvarianz, der Sorten $\times$ Umwelt - Varianz und der Fehlervarianz ausfielen. Diese Dominanz fanden auch Piepho und Michel (2001) an Rapsdaten. Sie schlussfolgerten daraus, dass bei erforderlicher Modellvereinfachung reine Umwelt -Effekte in der Modellgleichung im Unterschied zu allen Sorte • Umwelt - Interaktionseffekten formal als fix betrachtet werden können, wenn das Ziel in der Schätzung von Sortenkontrasten liegt.

Sorte × Umwelt - Varianz insgesamt:

Es wurden große Beträge für die Summe über alle Interaktionsvarianzen zwischen Sorte und Umwelten gefunden, insbesondere bei Ertragsmerkmalen. Diese großen Interaktionen begrenzen die Reproduzierbarkeit von Versuchsergebnissen für den Einzelfall, also einen Schlag in einem Jahr erheblich. Eine *gute* Reproduzierbarkeit für das langjährige Mittel vieler Schläge erfordert eine Serienauswertung mit entsprechend *vielen* Versuchen. Dabei sollte die Dimensionierung eines Prüfsystems besonders von der Präzisionsanforderung an die Ertragseinschätzung bestimmt werden, da diese einerseits überdurchschnittliche Sorte × Umwelt - Interaktionen aufweisen (s.o.) und andererseits i.d.R. eine herausragende Bedeutung in der Gesamtbewertung von Sorten besitzen. Auch Laidig et al. (2008) zeigen, dass niedrige Sorte × Umwelt - Varianzen, insbesondere niedrige Sorte × Jahr - Varianzen, zu niedrigen Standardfehlern von Sortendifferenzen führen.

Sorte × Typ (Jahr × Ort) - Varianz:

Bemerkenswert ist, dass es sogar innerhalb von J × O - Kombination, also bei gleicher Witterung und Betreuung in unmittelbarer Nähe, zu relevanten Interaktionen der Sorten über Versuche kommt. Dabei stellen verschiedene Versuche an gleichem Ort und Jahr i.d.R. verschiedene Versuchstypen dar (z.B. LSV, WP2, WP3 etc.). Dies unterstreicht die sehr eingeschränkte Verallgemeinerungsfähigkeit eines Einzelversuchsergebnisses bzw. von Beobachtungen auf einzelnen Schlägen. Die z.T. mehrfache Besetzung einzelner Sorten (z.B. Verrechnungssorten) in Versuchen am gleichen Ort und Jahr wird durch das gemischte Modell im Adjustierungsprozess der Mittelwertschätzung bestmöglich berücksichtigt.

Fehlervarianz der Sortenmittelwerte aus Einzelversuchen:

Die Genauigkeit der Sortenmittelwerte aus der Serienanalyse wird in weitaus stärkerem Maße von den Sorte × Umwelt - Interaktionen als von der Fehlervarianz in Einzelversuchen bestimmt. Ein ausreichendes Versuchsstellennetz und langjährige Auswertungen können nicht durch Intensivierung der Arbeit in Einzelversuchen bzw. an Einzelstandorten (z.B. Erhöhung der Wiederholungsanzahl) ersetzt werden. Ein Mindestmaß an Wiederholungen bleibt allerdings erforderlich bzw. nützlich zur Kontrolle und Bewertung der Plausibilität und Qualität der Daten wie auch ggf. zur Nutzung der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen. Insofern sollte aus den Varianzkomponenten nicht leichtfertig der Schluss gezogen werden, dass zusätzliche Versuche die Anlage von Wiederholungen im Einzelversuch überflüssig machen würden. Hinzukommt aus Kosten-Nutzen-Betrachtungen, dass zusätzliche Versuche an zusätzlichen Standorten erheblich teurer bzw. aufwändiger sind, als Wiederholungen im ohnehin angelegten und betreuten Versuch.

Komponenten der Sorte × Umwelt - Varianz:

Aufgrund der Größenordnungen und der über alle Merkmale auffallend großen Stabilität der Relationen der Komponenten der Sorte × Umwelt - Varianzen zueinander kann festgestellt werden, dass eine Vereinfachung der Sorte × Umwelt -Interaktionsstruktur z.B. mit der FITCON-Methode nach Yates (1933) eine relevante Beeinträchtigung der Übereinstimmung von Modell und Realität verursachen würde. Versuche im gleichen Jahr bzw. Versuche am gleichen Ort können nicht als unabhängige Umwelten angesehen werden. Hierin bestätigen sich die Aussagen von Piepho und Michel (2001). Auch Ouk et al. (2006) fanden an Reisversuchen in

Kambodscha, dass die Genotyp × Umwelt - Interaktion untersetzt werden sollte in die Komponenten Genotyp × Ort, Genotyp × Jahr und Genotyp × Ort × Jahr.

Sorte × Jahr × Ort - Varianz:

Die Sorte × Jahr × Ort - Varianz ist bei fast allen untersuchten Merkmalen die größte Komponente der Sorte × Umwelt - Varianz insgesamt. Dies steht in Übereinstimmung mit den Ergebnissen von Ouk et al. (2006) an Reis in Kambodscha. Die Sorten reagieren an einzelnen Ort × Jahr - Kombinationen also spezifisch auf Umwelt-Ereignisse, die weder für den Ort (in mehrjähriger Betrachtung) noch für das Jahr (in mehrortiger Betrachtung) als typisch angesehen werden können. Dieser Umstand erschwert die beratungsrelevante Interpretation der Sorten-Reaktionen auf Umwelteinflüsse. Bei Betrachtung von Versuchsergebnissen einzelner oder weniger Orte oder Jahre besteht erfahrungsgemäß ein großes Risiko, dass Sortenreaktionen fälschlich oder überzogen als Orts- oder Jahreseffekte interpretiert und Empfehlungen daraus abgeleitet werden. Die unterschiedliche Besetzung der Sorten an Ort × Jahr - Kombinationen wird durch das gemischte Modell im Adjustierungsprozess der Mittelwertschätzung bestmöglich berücksichtigt.

Sorte × Jahr - Varianz und Sorte × Jahr × Anbaugebiet - Varianz:

Die Interaktion von Sorten mit Jahren ist überwiegend die zweitgrößte Komponente der Sorte × Umwelt - Interaktionen. Bereits Bätz (1984) schätze ein, dass diese Varianzkomponente in der Regel vergleichsweise groß ist, was besonders durch Jahre mit extremem Witterungsverlauf verursacht würde. Die Interaktion der Sorten mit Jahren vollzieht sich eher einheitlich über alle einbezogenen Anbaugebiete hinweg als spezifisch in Anbaugebieten. Demnach beeinflusst Jahres-spezifisches Witterungsgeschehen (oder auch Befallsgeschehen u.ä.) die Sorteneffekte eher großräumig als Anbaugebiets-spezifisch. Großwetterlagen u.ä. scheinen also i.d.R. bedeutsamer als kleinräumigere Witterungsereignisse für die Sorteneffekte zu sein. Die vergleichsweise große Interaktion von Sorten mit Jahren zeigt, wie wichtig die Einbeziehung möglichst vieler Jahre für die Reproduzierbarkeit der Mittelwerte ist. Junge Sorten, die zur Überleitung des Züchtungsfortschrittes besonders interessant sind, haben naturgemäß zunächst nur wenige Versuchsjahre. Umso wichtiger ist die zusammenfassende Einbeziehung aller Sortenversuche des integrierten Prüfsystems, speziell die Einbeziehung der den LSV vorgelagerten WP und EUSV. Letzteres ist trotz eines entsprechenden Beschlusses der Agrarministerkonferenz (2004) derzeit noch nicht in allen Bundesländern umgesetzt. Die unterschiedliche Besetzung der Sorten in Jahren, welche Prüfsystem bedingt sehr extrem ist, wird durch das gemischte Modell im Adjustierungsprozess der Mittelwertschätzung berücksichtigt.

Sorte × Ort - Varianz:

Die gefundene vergleichsweise geringere Relevanz der Interaktion von Sorten mit Orten (innerhalb von Anbaugebieten) steht in Übereinstimmung mit den Ergebnissen von Cullis et al. (2000), die für Sortenversuche in Australien ebenso fanden, dass ‚statische' Komponenten der Sorte × Umwelt - Varianz erheblich geringer sind, als ‚saisonale' Komponenten. Dieses Ergebnis stützt den zugrunde gelegten Regionalisierungsansatz, wonach die Homogenität innerhalb eines vordefinierten Anbaugebietes so hinreichend ist, dass es nur ausnahmsweise bzw. nur geringe Ort spezifische Sortenreaktionen innerhalb von Anbaugebieten geben sollte. Dass speziell in der Merkmalskategorie ‚Minderertrag' einige hohe Varianzanteile auftreten,

könnte u.U. bzw. teilweise darin begründet sein, dass diese Behandlungseffekte stark vom Pflanzenschutzmanagement des Versuchsbetreuers am einzelnen Standort abhängen. Abweichend zu den Ergebnissen vorliegender Arbeit für Anbaugebiete stellen Laidig et al. (2008) für die Offizialprüfung Gesamt-Deutschlands fest, dass diese Varianzkomponente größer als die Sorte × Jahr - Varianz ist. Dies zeigt, dass die Gesamtheit der Versuchsstandorte der Sortenprüfung in Deutschland und damit die Grundgesamtheit ‚*alle Ackerbaustandorte in Deutschland*' als Regionalisierungsebene sehr groß und in sich ausgesprochen uneinheitlich ist. Vor dem Hintergrund der in der vorliegenden Arbeit genutzten vergleichsweise kleinräumigen Regionalisierung nach Anbaugebieten überrascht die geringe Bedeutung der S × O - Varianz in Anbaugebieten gegenüber der großen Bedeutung in Gesamt-Deutschland nicht. Es wird deutlich, dass die nationale Zulassung in Deutschland aufgrund der Auswertung der Wertprüfungen über alle Standorte in Deutschland eine relativ globale Bewertungsebene, aber noch keine relevante Regionalisierungsebene darstellt. Damit bestätigt sich die Notwendigkeit einer Regionalprüfung in Anbaugebieten im Anschluss an die nationale Zulassung. Über den idealen räumlichen Zuschnitt und die Größe von Anbaugebieten ist damit allerdings noch keine Aussage getroffen.

Cullis et al. (2000) schlussfolgern, dass die Anzahl Jahre im Vordergrund stünde während eine hohe Ortsanzahl wenig nützlich, aber kostspielig sei. Trotz dieser vergleichsweise geringen Interaktion zwischen Sorte und Ort soll durch den Autor vorliegender Arbeit doch auf die große Bedeutung einer ausreichenden Ortsanzahl je Anbaugebiet hingewiesen werden, da besonders bei den jungen Sorten noch zu wenige Jahre zur Verfügung stehen und damit nur über eine ausreichende Versuchsanzahl insgesamt ein hinreichend reproduzierbares regionales Ergebnis entstehen kann. Der Prüfprozess im regionalen Freiland ist auf natürliche Weise an den Jahreszyklus gebunden und kann nicht beschleunigt werden. Des Weiteren lassen sich die Vorzüge des Gesamtmodells der Hohenheim Gülzower-Serienauswertung nur realisieren, wenn die Varianzkomponenten auch innerhalb von Anbaugebieten hinreichend gut geschätzt und von Komponenten zwischen Anbaugebieten separiert werden können.

In Ihren Analysen zum genetisch und zum nicht-genetisch bedingten Trend von Erträgen über einen längeren Zeitraum kommen Laidig et al. (2014) zu der Empfehlung, dass Modelle für Langzeitanalysen im Bereich der Züchtung und Sortenprüfung auch Effekte der ‚Alterung' von Sorten und der Selektion berücksichtigen sollten. Diese Fragen können nach Ansicht des Autors dieser Arbeit auch für die Modelle der Hohenheim-Gülzower-Serienauswertung Relevanz besitzen und sollten für die Weiterentwicklung in Betracht gezogen werden. Für das Phänomen einer möglichen Alterung von Sorten, also einer Leistungsveränderung über die Zeit, kommen einige Ursachen in Betracht, z.B. Veränderung einer Sorte im Zuge der Erhaltungszüchtung, Anpassung der Schaderreger an Anbau bestimmende Sorten in der landwirtschaftlichen Praxis oder auch Veränderungen in den Anbauverfahren durch technische Entwicklungen, politische Vorgaben u.v.m. Diesem Aspekt wird bereits jetzt dadurch (suboptimal) etwas Rechnung getragen, dass die Ertragsauswertung auf sechs Jahre begrenzt wird, obgleich die Interpretation der Varianzkomponenten eine möglichst noch längere Auswertungsperiode nahe legt. Im Zuge der Weiterentwicklung der Auswertungsverfahren sollten Untersuchungen zur Einbeziehung der Einflüsse von Alterung von Sorten und der Selektion in die Modellbildung einfließen.

Die - Prüfsystem bedingt - unterschiedliche Besetzung der Sorten an Orten (Standorte der Länderdienststellen (LDS), des Bundessortenamtes oder der Züchter

weisen unterschiedliche Anteile der Versuchstypen LSV, WP oder EUSV auf) wird durch das gemischte Modell im Adjustierungsprozess der Mittelwertschätzung bestmöglich berücksichtigt.

Sortenvarianz und ihre Anbaugebietsspezifik:

Bei fast allen Merkmalen war für mindestens eines der bis zu fünf einbezogenen Anbaugebiete Sortenspezifik nachweisbar. Dies ist ein Hinweis darauf, dass die Bildung von Anbaugebieten in diesen räumlichen Dimensionen zumindest im Ansatz sinnvoll ist. Allerdings war für relativ viele andere Anbaugebiete - differenziert nach Merkmalen - keine Sortenspezifik nachweisbar. Insofern scheinen einige Anbaugebiete entweder nicht optimal zugeschnitten zu sein oder die derzeit verfügbare Datenbasis hat z.T. bereits das für eine hinreichende Parameterschätzung und -eingrenzung erforderliche Maß unterschritten. Eine Schlussfolgerung, die Größe der Anbaugebiete prinzipiell zu vergrößern würde zwar seitens der dann erhöhten Datenbasis je Gebiet die Aussichten vergrößern, etwaige Spezifik nachzuweisen und zu quantifizieren, allerdings mit dem Risiko, dass genau diese Spezifik durch regionale Vergröberung verloren geht. Immerhin gibt es Veranlassung, den Zuschnitt und die Größe der Anbaugebiete zu hinterfragen. Hierfür sollten nicht wie bisher ausschließlich standortkundliche Erwägungen in Betracht kommen. Zur Validierung unterschiedlicher standortkundlich begründeter Annahmen für eine geeignete Regionalisierung bietet sich die Nutzung der Hohenheim-Gülzower-Serienauswertung geradezu an.

Effekte der langjährigen Auswertung und der Einbeziehung vorgelagerter Versuche auf Mittelwerte und ihre Genauigkeit

Die Ergebnisse zeigen, dass die Einbeziehung vorgelagerter Versuche speziell für die frühe Einschätzung junger Sorten und eine langjährige Auswertung besonders für die zur Empfehlung anstehenden älteren Sorten ausgesprochen vorteilhaft ist. Die in Mecklenburg-Vorpommern eingeführte Vorgehensweise der Kombination beider Maßnahmen erbringt hier Synergien für Sorten unterschiedlichsten Prüfstatus‘. Die Erweiterung der Datenbasis führt zu einer Verbesserung der Genauigkeit der Mittelwertschätzungen / -vorhersagen.

Tendenziell wirkt eine ausgewogene Verteilung der Versuche (wie auch der Versuchsergebnisse je Sorte im Speziellen) auf Jahre und Orte positiv auf die Schätzgenauigkeit, wobei in diesem Fall - entsprechend der Relation zwischen den Varianzkomponenten - zusätzliche Jahre zunächst (bis zu einem Gleichgewichtspunkt) wirksamer sind als zusätzliche Orte.

Als etwas pauschalisierte ‚Faustzahlen‘ für sechsjährige Auswertungen können folgende Aussagen getroffen werden: Weniger als 10 Versuchsergebnisse über alle Jahre ergeben bestenfalls eine grobe Voreinschätzung von Sorten; bis zu etwa 20 Versuchsergebnissen bringt jedes zusätzliche Ergebnis einen relevanten Gewinn an Genauigkeit; im Bereich von etwa 20 bis 40 Versuchsergebnissen flacht der Zugewinn zunehmend ab. Ab ca. 40 Versuchsergebnissen bringen weitere Ergebnisse keinen nennenswerten Zugewinn an Genauigkeit mehr, wenn nicht auch die Anzahl Jahre erhöht wird. Letzteres bestätigt die Vorgehensweise in Mecklenburg-Vorpommern, langjährig durchgeprüfte aber noch interessierende ‚alte Sorten‘ nicht mehr weiter zu prüfen (im Sinne der Kosteneffizienz und der Versuchsgröße), diese Sorten aber weiterhin im Auswertungsdatensatz zu behalten und die (auf den

Versuchen der Vorjahre basierenden) adjustierten Mittelwerte zu nutzen. Im Zusammenhang mit diesen Faustzahlen soll betont werden, dass es sich um die Voraussetzungen für eine hinreichend präzise Schätzung der mittleren Ertragsleistung handelt. Für die präzise Einschätzung der Ertragsstabilität einzelner Sorten sind nach Mühleisen et al. (2014b) mindestens 40 Test-Umwelten erforderlich (bei Wintergerste).

Die Nutzung langjähriger Daten im Zusammenspiel mit der Einbeziehung den LSV vorgelagerter Versuche hat von allen Optimierungsansätzen der Hohenheim-Gülzower-Serienauswertung die stärksten Auswirkungen auf die Mittelwerte und deren Genauigkeit. Die Korrelation zu den Mittelwerten nach traditioneller Herangehensweise, ausschließlich LSV-Ergebnisse der letzten drei Jahre zu nutzen, ist vergleichsweise gering (beim Kornertrag Winterweizen: $r = 0{,}82$). Es wurden z.T. erhebliche fachlich ausgesprochen relevante Unterschiede der Sorteneinschätzungen zwischen dem neuen und dem traditionellen Ansatz gefunden, besonders ausgeprägt bei jungen Sorten mit noch relativ geringer Datenbasis.

Damit ist die Einbeziehung langjähriger Ergebnisse und den LSV vorgelagerter Versuche eine ausgesprochen wirksame Maßnahme, um die Genauigkeit allgemein zu erhöhen, insbesondere aber um junge Sorten frühzeitig einzuschätzen. Dies steht in Übereinstimmung zu Piepho und Möhring (2005), die ebenfalls besonders auf die Notwendigkeit langjähriger Datensätze hinwiesen. Die zwangsläufig damit einhergehende Zunahme der Unbalanciertheit muss dafür als systemimmanent in Kauf genommen werden. Das Festhalten an vollständig balancierten Strukturen mittels ‚*Schneiden*' orthogonaler Kerne (Verwerfen von existierenden geeigneten Daten) muss als nachteilig angesehen werden.

Effekte einer Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen auf Mittelwerte in der Serienauswertung und ihre Genauigkeit

Diese Gewichtung wirkt sich nur bei Sorten mit schwächerer Datenbasis in relevantem Maße auf die Schätzung der Serienmittelwerte und deren Genauigkeit aus. Bei wenigen Versuchsergebnissen je Sorte kann es im Einzelfall spürbar auf die Genauigkeit durchschlagen, ob die Versuchsergebnisse tendenziell gehäufter aus problematischen bzw. eher aus sehr präzisen Versuchen stammen. Es kann davon ausgegangen werden, dass ohne Gewichtung die Genauigkeit eines Sortenmittelwertes der Serienauswertung speziell in den Einzelfällen überschätzt wird, bei denen wenige Versuchsergebnisse für diese Sorte vorlagen, von denen überproportional viele mit einem hohen Einzelversuchsfehler belastet waren. Relevante Mittelwertveränderungen über 1,0 dt/ha traten bei Sorten im Kernsortiment nicht auf, sondern lediglich bei noch seltener geprüften ‚sonstigen' Sorten, deren Leistung allerdings durchaus gelegentlich für sekundäre Fragestellungen und grobe Früheinschätzungen abgefragt wird.

Die im Allgemeinen relativ geringen Auswirkungen der Gewichtung beschreiben bereits Piepho und Michel (2001) an einem langjährigen Rapsdatensatz. Möhring und Piepho (2009) kommen bei Auswertungen mit Winterweizen zu dem Ergebnis, dass eine Zweischrittanalyse i.d.R. bereits ohne Gewichtung zu akzeptablen Ergebnissen gelangt, dass Gewichtung mit geeigneten Wichtungsmethoden aber meist noch etwas leistungsfähiger ist und dass die Unterschiede zwischen diesen Wichtungsmethoden marginal und Datensatz-spezifisch sind. Dessen ungeachtet sollte die in den PIAFStat-Verfahren für die Hohenheim-Gülzower-Serienauswertung imple-

mentierte Gewichtung künftig methodisch vervollkommnet werden, um eine adäquate Einschrittanalyse möglichst vollständig zu reproduzieren. Grundlagen dafür haben Piepho et al. (2012) dargelegt.

Der potenzielle Vorteil der gewichteten Auswertung liegt darin, dass die Informationen zu den Einzelversuchsfehlern nicht verloren gehen, so dass die konkrete Datenlage je Sorte bestmöglich berücksichtigt wird. Aus prinzipiellen Überlegungen heraus erscheint es insofern sinnvoll, die Genauigkeit der Mittelwerte im Einzelversuch zu berücksichtigen. Die Nutzung der Gewichtung erleichtert die Entscheidung, auch Ergebnisse aus Versuchen mit Fehlwerten bzw. mit erhöhtem Versuchsfehler (insofern der Versuch ansonsten, z.T. subjektiv, als fachlich relevant, repräsentativ und ‚auswertbar' angesehen wird), einzubeziehen. Das Risiko, dass besonders ‚*gestörte*' Einzeldaten zu stark auf die Mittelwertschätzungen durchschlagen, wird reduziert und ihr Informationsbeitrag wird angemessen gewichtet genutzt. Letzteres ist auch im Sinne des Aufwand-Nutzen-Verhältnisses bezüglich der für die Versuchsdurchführung eingesetzten erheblichen Ressourcen begründet - die Nichtwertung eines Versuches besitzt nicht nur einen fachlichen sondern ebenso einen ökonomischen Aspekt. Für eine routinemäßige Auswertung wird vom Autor vorliegender Arbeit eine Gewichtung vorgeschlagen, um ggf. sehr ungleiche Fehlervarianzen angemessen zu berücksichtigen.

Die Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen erfordert bei vielen PIAF-Anwendern eine rückwirkende Schätzung der Standardfehler der Mittelwerte ($SE_{Versuch}$) in älteren Versuchen und die Speicherung dieser Standardfehler im PIAF-System. Sofern diese Arbeitskapazität nicht zur Verfügung steht, sollte zumindest bei aktuellen Versuchen diese ohnehin verfügbare PIAFStat-Option der Berechnung und automatischen Speicherung genutzt werden. Dadurch wird eine künftige Nutzung unschwer vorbereitet.

6.3 Gewichtung von Zielgebiet und Nachbargebieten

6.3.1 Typische Muster in der Struktur der genetischen Korrelationen

Die genetischen Korrelationen zwischen Anbaugebieten sind in Tab. 35 ausgewiesen. Die Tabellenstruktur entspricht weitgehend Tab. 30, wobei Pflanzenart × Merkmal - Kombinationen mit nur einem einzigen Anbaugebiet in Tab. 35 nicht mehr aufgenommen wurden. Des Weiteren soll nochmals darauf aufmerksam gemacht werden, dass bei Winterweizen und Raps die Ertragsauswertung in Mecklenburg-Vorpommern für zwei unterschiedliche Zielgebiete erfolgt und daher hier in zwei Zeilen dargestellt wird.

Die am deutlichsten unter eins liegenden genetischen Korrelationen zum jeweiligen Zielgebiet konnten für Nachbargebiete in der regionalen Ausrichtung W bzw. (W+SW) ausgemacht werden. Fast alle Korrelationen zu diesen Gebieten lagen unter 1,0. Das Nachbargebiet S zeigt relativ selten eine erheblich von eins abweichende genetische Korrelation zum Zielgebiet.

Auffällig ist, dass sich beim Raps die Unterteilung in die beiden räumlich kleinen Zielgebiete D-Nord und MV-Süd als sinnvoll bestätigt hat $(r=0,85)$. Dagegen war dies beim Winterweizen nicht der Fall $(r=1,0)$.

Vergleichsweise niedrige genetische Korrelationen für einzelne Nachbargebiete wurden bei folgenden Pflanzenart × Merkmal - Kombinationen gefunden: Ertrag-Körnermais (r_{min}=0,65), Minderertrag-Wintergerste (r_{min}=0,65), Rohproteingehalt-Sommerbraugerste (r_{min}=0,58) und ‚Sortierung > 2.5 mm (%)' bei Sommerbraugerste (r_{min} =0,62).

Bezüglich der Merkmalskategorien zeigten die Mindererträge mit einem Median von r=0,89 tendenziell die niedrigsten genetischen Korrelationen. Bei Erträgen und Qualitätsmerkmalen liegt der Median bei r=0,96.

Für Ertrag- und Minderertrag-Roggen, Ertrag-Hafer, Minderertrag-Triticale und Tausendkornmasse-Sommerbraugerste wurden keinerlei S × AG - Interaktion nachgewiesen.

Tab. 35: Genetisch Korrelation der Sorteneffekte von Nachbargebieten zum jeweiligen Zielgebiet

Pflanzenart	**Merkmal**	**Koeffizient der genetischen Korrelation von Nachbargebieten zum Zielgebiet (ZG)**				
		D-Nord	**MV-Süd**	**S**	**SW**	**W**
WW	Ertrag	ZG	1		1	0,84
WW	Ertrag	1	ZG	1	1	
GW	Ertrag	ZG		0,95		0,96
GS_B	Ertrag	ZG			1	
RW	Ertrag	ZG		1		
TIW	Ertrag	ZG		0,97		
HA	Ertrag	ZG			1	
RAW	Ertrag	ZG	0,85		1	1
RAW	Ertrag	0,85	ZG	0,75	0,85	
EF	Ertrag	ZG			0,71	
MK	Ertrag	ZG			0,65	
MS	Ertrag	ZG		0,94		0,85
WW	MinderE	ZG		0,89	0,89	
GW	MinderE	ZG		0,68		0,65
GS_B	MinderE	ZG			0,92	
RW	MinderE	ZG		1		
TIW	MinderE	ZG		0,92		
MS	TSNIGEPF	ZG		1		0,87
MK	TS_KORN	ZG			0,75	
EF	RP__86	ZG			0,97	
MS	STAEGEH	ZG		0,99		0,98
GS_B	SORTGR25	ZG			0,62	
GS_B	RP__I_TM	ZG			0,58	
GS_B	TKG_LUTR	ZG			1	
GS_B	HLG	ZG			0,96	
MS	NELKGX	ZG		0,98		0,96

aus standortkundlichen Erwägungen nicht einbezogenes Nachbargebiet (s. 4.1.3)

6.3.2 Typische Muster in der Gewichtung von Nachbargebieten

Die Analyse erfolgt beispielhaft für das Merkmal Kornertrag bei Winterweizen für das Zielgebiet D-Nord.

In Tab. 36 ist die Verteilung der Gewichte bei der Ertragsschätzung auf das Zielgebiet und die Nachbargebiete dargestellt (oberstes Tabellen-Drittel), wobei für die Darstellung die Summenrestriktion zur Anwendung kam (Summe der Anbaugebiets-Gewichte je Sorte = eins). Da diese Gewichtung allerdings nicht nur von der genetischen Korrelationsstruktur zwischen den Anbaugebieten abhängt, sondern auch von der jeweiligen Besetzung einer Sorte in den einzelnen Anbaugebieten, ist ergänzend auch die Anzahl Versuchsergebnisse je Sorte und Anbaugebiet dargestellt (mittleres Tabellen-Drittel). Der Quotient beider Werte ergibt die durchschnittliche Gewichtung je Versuch (unteres Tabellen-Drittel). Die Darstellung für alle Sorten des Kernsortimentes erschien hier überzogen. Zusammenfassende Parameter wie Mittelwert oder Spannweite für die Sorten je eines Prüfstatus‘ (Def. s. Tab. 8) sind allerdings auch schwer interpretierbar, da die Sorten innerhalb eines Prüfstatus‘ z.T. auch noch sehr unterschiedlich besetzt waren. Ursache dafür war z.B., dass Sorten temporär den Status ‚Verrechnungssorte‘ (VRS) inne hatten und somit in den Jahren dieser Einordnung eine untypische Zunahme des Prüfumfangs erfuhren. Andererseits wurden im Zuge des Prüfprozesses einzelne Sorten in ausgewählten Anbaugebieten reduziert oder eliminiert, in anderen ausgeweitet. Als Kompromiss wurde daher je Prüfstatus eine in der Besetzung über die Anbaugebiete *typische* Sorte ausgewählt und dargestellt (Kriterium: minimale Summe der quadrierten Abweichungen der Ist-Besetzungen dieser Sorte über die Anbaugebiete vom entsprechenden Median aller Sorten dieses Prüfstatus‘).

Für Sorten des Kernsortimentes ergab sich in jedem Prüfstatus, dass der Effekt im Zielgebiet D-Nord jeweils eine größere (Eigen-) Gewichtung erhielt, als die Effekte jedes einzelnen Nachbargebietes (Tab. 36 und 37; Parameter W_{AG}). Hierzu muss allerdings angemerkt werden, dass unter den dargestellten Sorten außer im Falle der Sorte *Smaragd* das Zielgebiet immer auch die stärkste Besetzung hatte. Dies hängt z.T. mit dem hier angewandten Zuschnitt der Nachbargebiete zusammen, wonach standortkundlich und geografisch vom Zielgebiet tendenziell eher entfernte Boden-Klima-Räume eines Anbaugebietes nicht als Teil des entsprechenden Nachbargebietes einbezogen wurden (s. 5.1.3). Betrachtet man allerdings jede einzelne Sorte, so treten auch (hier nicht dargestellte) Einzelfälle auf, bei denen ein besonders stark besetztes Nachbargebiet höher gewichtet wird, als das schwächer besetzte Zielgebiet. Bemerkenswert ist, dass die Summe der Gewichte über alle Nachbargebiete überwiegend größer war, als das Gewicht des Zielgebietes selbst (W_{AG} des ZG < 0,5).

In Tab. 37 erfolgt über alle Sorten des Kernsortimentes eine Verdichtung in vier Klassengruppen - dadurch wird die Dominanz des Zielgebietes einerseits und die geringe Gewichtung besonders der Nachbargebiete Ẉ und ṢẈ andererseits noch deutlicher als in Tab. 36.

Die S × AG - Effekte eines Anbaugebietes werden höher gewichtet, je stärker die Datenbesetzung in diesem Anbaugebiet ist. Das wird beim Vergleich der Nachbargebiete MV-Süd und ṢẈ ersichtlich, die sich nicht in der genetischen Korrelation zum Zielgebiet unterscheiden. Beim Nachbargebiet Ẉ wird deutlich, dass die geringere genetische Korrelation zum Zielgebiet die Gewichtung dämpft (*Discus* und *Tabasco* erhalten für ṢẈ eine höhere Gewichtung als für Ẉ, obgleich die Anzahl Versuche in Ẉ höher ist.

Gegenläufig zur Gewichtung der S × AG - Effekte verhält sich die durchschnittliche Gewichtung je Einzelversuch. Je höher die Ergebnisdichte in einem Anbaugebiet, desto geringer wird die mittlere Gewichtung jedes Einzelergebnisses - dies in Abhängigkeit der Varianzkomponenten: je stärker die Korreliertheit der Versuche innerhalb von Anbaugebieten ist, desto geringer fällt der Informationsgewinn eines zusätzlichen Versuches aus. So erbringt z.B. im Extrem die Besetzung der Sorte *Mercato* in MV-Süd mit nur einem Versuch erwartungsgemäß zwar eine vergleichsweise geringe Gewichtung des S × AG - Effektes, dieser Einzelversuch erbringt aber andererseits die höchste hier bestimmte Einzelgewichtung. Die höchste Einzelversuchsgewichtung wurde im Prüfstatus OV_EU (*Mercato*) im Nachbargebiet MV-Süd mit 17% gefunden, die geringste mit nur 0,2% im Prüfstatus LSV5-8 bei der langjährigen Verrechnungssorte *Tommi* im Nachbargebiet W.

Im Verlauf des ‚*normalen*' Prüfprozesses (über die deutsche Wertprüfung) von OV_WP bis etwa LSV4 *e*rfolgt automatisch eine zunehmende Eigengewichtung des Zielgebietes von 36% bis 51% der Summe der Gewichte. Die Abweichungen hierzu bei OV_EU und bei Sorten älteren Prüfstatus' sind Resultat sortenspezifischer vom ‚Regelprüfzeitraum' abweichender Verteilungen der Daten über die Anbaugebiete.

Tab. 36: Verteilung der Gewichte bei der Ertragsschätzung für ein Zielgebiet auf einbezogene Ziel- und Nachbargebiete; Kornertrag Winterweizen, Zielgebiet D-Nord

Prüfstatus der Sorten (Definition s. Tab. 8)	**typische Sorte**	**D-Nord** Zielgebiet	**MV-Süd** r_g=1	**SW** r_g=1	**W** r_g=0,84
			N a c h b a r g e b i e t e		
		W_{AG} - Gewichtung der S × AG - Effekte ($\Sigma_{AG} w_{AG} = 1$)			
alte Sorten	Impression	0,38	0,30	0,24	0,09
LSV5-8	Brilliant	0,44	0,39	0,06	0,11
LSV4	Potenzial	0,51	0,43	-	0,05
LSV3	Discus	0,40	0,30	0,18	0,12
LSV2	Tabasco	0,38	0,30	0,18	0,14
LSV1	Global	0,36	0,29	0,13	0,22
OV_WP	Smaragd	0,36	0,15	0,21	0,28
OV_EU	Mercato	0,56	0,17	-	0,28
		N_{AG} - Anzahl Versuchsergebnisse je Sorte im AG			
alte Sorten	Impression	19	13	11	7
LSV5-8	Brilliant	30	21	3	14
LSV4	Potenzial	25	18	0	14
LSV3	Discus	21	15	10	12
LSV2	Tabasco	19	16	8	12
LSV1	Global	9	7	3	6
OV_WP	Smaragd	5	2	3	6
OV_EU	Mercato	4	1	0	3
		mittlere relative Gewichtung je Versuchsergebnis (%)			
alte Sorten	Impression	2,0	2,3	2,2	1,3
LSV5-8	Brilliant	1,5	1,9	2,0	0,8
LSV4	Potenzial	2,0	2,4	-	0,4
LSV3	Discus	1,9	2,0	1,8	1,0
LSV2	Tabasco	2,0	1,9	2,3	1,2
LSV1	Global	4,0	4,1	4,3	3,7
OV_WP	Smaragd	7,2	7,5	7,0	4,7
OV_EU	Mercato	14,0	17,0	-	9,3

r_g genetische Korrelation (hier zum Zielgebiet D-Nord)
W Gewichtung (0 ≤ W ≤ 1)

Tab. 37: Häufigkeitsverteilung für die Gewichtung der S × AG - Effekte (W_{AG}) in den Anbaugebieten; Kornertrag Winterweizen-Kernsortiment, Zielgebiet D-Nord,

Häufigkeits-Klassen der Gewichtung ($W_{AG\ \%}$)	**Anzahl Fälle (Sorten) in Anbaugebieten**			
	D-Nord	MV-Süd	ṢẆ	Ẉ
0% - 14 %	0	0	28	25
15% - 29%	0	23	13	19
30% - 44%	21	20	3	0
> 45%	23	1	0	0

6.3.3 Effekte der Einbeziehung von Nachbargebieten auf Mittelwerte im Zielgebiet und ihre Genauigkeit

Die Analyse erfolgt wie in 6.3.2 beispielhaft für das Merkmal Kornertrag bei Winterweizen für das Zielgebiet D-Nord.

Nachfolgend soll zunächst untersucht werden, wie sich typische Regionalisierungsstrategien auf die Genauigkeit bei Sorten unterschiedlichen Prüfstatus‘ (Def. s. Tab. 8) auswirken. Des Weiteren soll eingeschätzt werden, wie stark sich die Strategien auf die Mittelwertschätzungen - wiederum je nach Prüfstatus - auswirken können.

Auch für diese Analysen wurden Szenarien (Sz) unter 5.5 definiert und in diesem Abschnitt alternative Auswertungsstrategien (Sz 7 und Sz 8) in Tab. 38 nochmals zum besseren Verständnis systematisiert. Dabei stellt das Sz 1 wiederum die derzeitige durch den Autor eingeführte Vorgehensweise in Mecklenburg-Vorpommern dar. Die Sz 7 und 8 sind konträre Strategien, bei denen zum einen nur das Zielgebiet (SZ 7), zum anderen ein Großraum um das Zielgebiet D-Nord (das Zielgebiet einschließend) (Sz 8), ausgewertet wird. Dieser Großraum kann fachlich-standortbezogen als ‚*Diluviale Standorte Nordost-Deutschlands*‘ bezeichnet werden, das dahinter stehende Konzept ist unter 5.1.3 dargelegt.

Tab. 38: Definition von Szenarien für Regionalisierungsstrategien (s. 5.5)

Szenarien	**Definition des Zielgebietes entsprechend 4.1.3**	**Einbeziehung von Nachbargebieten mit abgestufter Gewichtung**
Sz 1	D-Nord	ja*
Sz 7	D-Nord	nein
Sz 8	‚Großraum‘ um D-Nord**	nein

* Nachbargebiete MV-Süd, ṢẆ, Ẉ (s. Tab. 15 und Abb. 9);

** Einbeziehung aller Standorte der Anbaugebiete D-Nord, MV-Süd, ṢẆ, Ẉ (s. Abb. 9)

Um den Vergleich der Szenarien nicht durch zufällige Abweichungen in der Schätzung des Transformationsparameters und der Varianzkomponenten zu beeinflussen wurden für Sz 1 und Sz 7 übereinstimmende Parameter entsprechend 6.1.1 und 6.2.1 zugrunde gelegt. Aufgrund der grundsätzlich anderen räumlichen Struktur wurden für das Sz 8 bei unverändertem Transformationsparameter die Varianzkomponenten neu geschätzt.

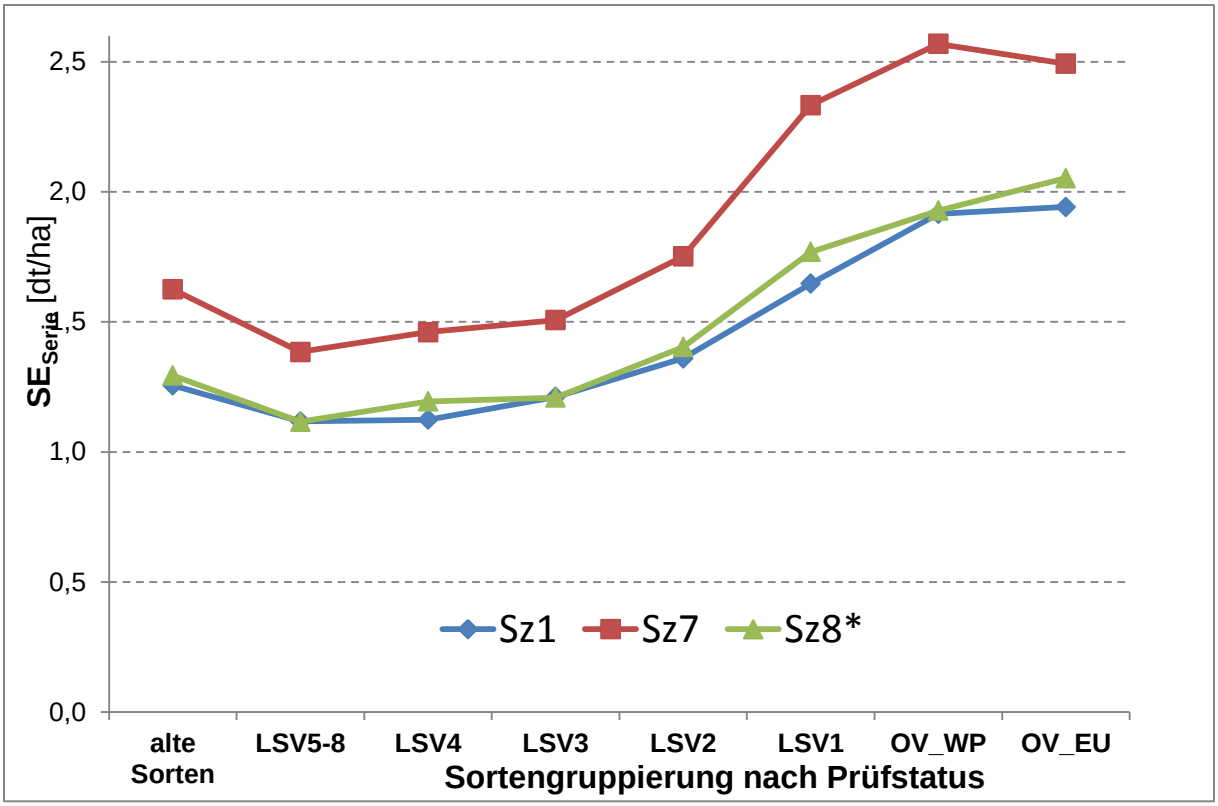
Tab. 39 und Abb. 31 zeigen die Unterschiede in der erzielten Genauigkeit bei den unterschiedlichen Szenarien. Dabei ist zu berücksichtigen, dass Sz 8 hierin nicht direkt mit Sz 1 und Sz 7 verglichen werden kann, da die Grundgesamtheit / das Zielgebiet anders definiert ist (nicht D-Nord, sondern ‚Großraum' um D-Nord).

Tab. 39: Auswirkung von Szenarien für die Regionalisierung auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen

Prüfstatus (Definition s. Tab. 8)	**N**	**Sz 1**	**Sz 7**	**Sz 8***
		SE_{Serie} [dt/ha]		
alte Sorten	7	1,26	1,63	1,29
LSV5-8	8	1,12	1,38	1,12
LSV4	2	1,12	1,46	1,19
LSV3	3	1,21	1,51	1,21
LSV2	6	1,36	1,75	1,40
LSV1	4	1,65	2,33	1,77
OV_WP	3	1,92	2,57	1,93
OV_EU	2	1,94	2,49	2,05

Definition des Prüfstatus' s. Tab. 8

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: abgestuft gewichtete Einbeziehung von Nachbargebieten
Sz 7: ohne Einbeziehung von Nachbargebieten, im Übrigen wie Sz 1
Sz 8: Nachbargebiete ohne abgestufte Gewichtung einbezogen, im Übrigen wie Sz 1
* Standardfehler gelten für Großraum, nicht für definiertes Zielgebiet, nur bedingt vergleichbar!
Definition der Szenarien: s. 5.5



Definition des Prüfstatus‘ s. Tab. 8

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: abgestuft gewichtete Einbeziehung von Nachbargebieten
Sz 7: ohne Einbeziehung von Nachbargebieten, im Übrigen wie Sz 1
Sz 8: Nachbargebiete ohne abgestufte Gewichtung einbezogen, im Übrigen wie Sz 1
* Standardfehler gelten für Großraum, nicht für definiertes Zielgebiet, nur bedingt vergleichbar!
Definition der Szenarien: s. 5.5

Abb. 31: Einfluss von Regionalisierungsstrategien auf die Genauigkeit; Kornertrag Winterweizen

Die räumliche Erweiterung der Datenbasis in Sz 1 und Sz 8 führt zu einer deutlichen Erhöhung der Genauigkeit der Mittelwertschätzungen gegenüber der traditionellen räumlich eng begrenzten Strategie (Sz 7). Besonders auffällig sind die hohen Standardfehler beim Sz 7 im frühen Prüfstadium bis etwa LSV1. Beim LSV1 liegt der Standardfehler (SE_{Serie}, s. 5.3.5) für das Sz 7 um das 1,4-fache über dem für das Sz 1, beim LSV3 liegt dieses Verhältnis noch beim 1,2-fachen.

Der Pearsonsche Korrelationskoeffizient für den linearen Zusammenhang zwischen den Schätzwerten aus Sz 1 und Sz 7 weicht bei Einbeziehung aller Sorten leicht, aber nicht irrelevant von eins ab $(r = 0{,}954)$. Dagegen ist Sz 1 mit Sz 8 sehr hoch korreliert $(r = 0{,}995)$. Wie in den vorherigen Abschnitten diskutiert, ist aber die Frage, wie groß *im Einzelfall* Abweichungen bei Sorten sein können, relevanter als der mittlere Zusammenhang (Tab. 40). Stellt sich z.B. für eine einzelne konkrete Sorte eine positive oder negative spezifische Reaktion in einem Anbaugebiet dar, so ist diese individuelle Information für einen Landwirt in diesem Gebiet u.U. hochgradig relevant für Entscheidungen, auch wenn für die Gesamtheit der Sorten / also im Durchschnitt der Sorten kaum Gebietsspezifik auszumachen ist. Insofern sind einzelne ‚Korrelationsbrecher‘ unabhängig von der Durchschnittsreaktion aller Sorten von hoher Relevanz.

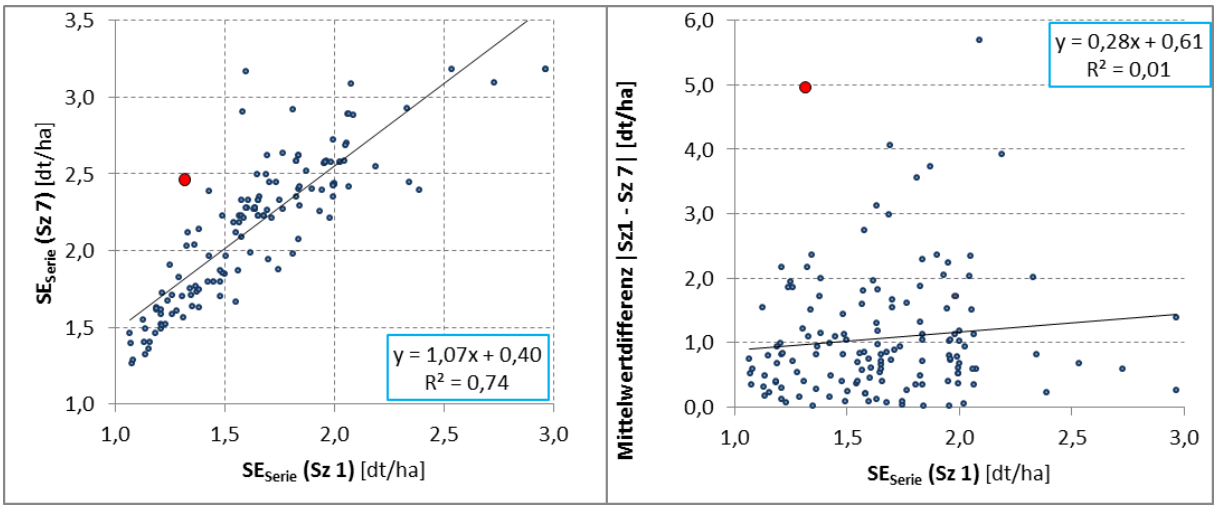
Tab. 40: Auswirkung der Regionalisierungsstrategien auf die Mittelwerte; Kornertrag Winterweizen

Prüfstatus (Definition s. Tab. 8)	**N**	**|Sz1 - Sz 7|** [dt/ha]		**|Sz1 - Sz 8|** [dt/ha]	
		Mittelwert	**Maximum**	**Mittelwert**	**Maximum**
alte Sorten	7	0,72	1,86	0,24	0,78
LSV5-8	8	0,58	1,55	0,14	0,56
LSV4	2	0,56	0,74	0,39	0,50
LSV3	3	1,04	2,17	0,29	0,43
LSV2	6	0,98	1,71	0,09	0,13
LSV1	4	0,48	0,86	0,22	0,31
OV_WP	12	1,24	2,99	0,43	1,46
OV_EU	2	2,16	2,29	0,27	0,53

Definition des Prüfstatus‘ s. Tab. 8

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: abgestuft gewichtete Einbeziehung von Nachbargebieten
Sz 7: ohne Einbeziehung von Nachbargebieten, im Übrigen wie Sz 1
Sz 8: Nachbargebiete ohne abgestufte Gewichtung einbezogen, im Übrigen wie Sz 1
nach Auswertung transformierter Daten werden rücktransformierte Mittelwerte, also Schätzwerte für den Median verglichen
Definition der Szenarien: s. 5.5

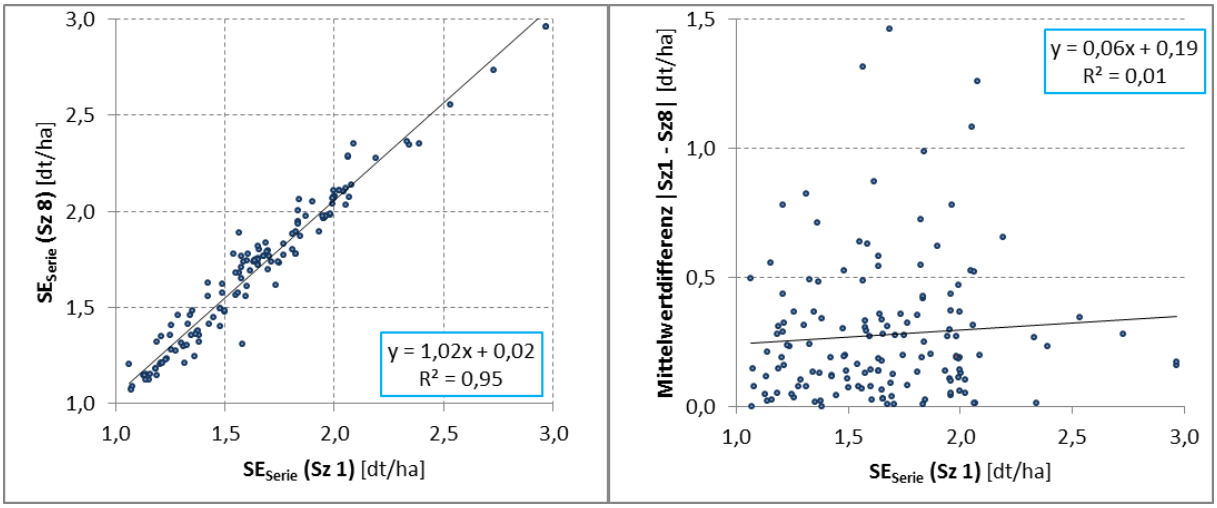
Ergänzend für die Auswertungen zum Kernsortiment (Tab. 40) wird nachfolgend für das gesamte Sortiment für die Vergleiche Sz 1 versus Sz 7 sowie Sz 1 versus Sz 8 die Wirkung auf die Genauigkeit und insbesondere auf die Mittelwerte analysiert. Eine für das Prüfsystem wesentliche Frage ist dabei wiederum, ob diese Differenzen im Zusammenhang mit dem Prüfstatus der Sorten stehen. Ein fortschreitender Prüfstatus schlägt sich primär in einem mit der Prüfdauer abnehmenden SE_{Serie} nieder. Daher wurde als stetige Einflussgröße der SE_{Serie} (nach Sz 1) anstelle des Prüfstatus‘ in die Darstellung einbezogen und ein funktionaler Zusammenhang dargestellt (Abb. 32 und 33). Aufgrund formaler Vorgaben in PIAFStat (s. 5.4) werden nur Mittelwerte für Sorten ausgegeben, für die mindestens drei Werte im Gesamtdatensatz vorliegen. Dadurch wurde diese Auswertung auf 144 Sorten im Sz 7 beschränkt - diese Beschränkung wurde daraufhin einheitlich für alle drei Szenarien orthogonal angewendet.



Hervorhebung eines einbezogenen Extremwertes der Datenreihe

Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: abgestuft gewichtete Einbeziehung von Nachbargebieten
Sz 7: ohne Einbeziehung von Nachbargebieten, im Übrigen wie Sz 1
Definition der Szenarien: s. 5.5

Abb. 32: Einfluss der gewichteten Einbeziehung von Nachbargebieten auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen



Sz 1: Standard der Hohenheim-Gülzower-Serienauswertung, hier speziell: abgestuft gewichtete Einbeziehung von Nachbargebieten
Sz 8: Nachbargebiete ohne abgestufte Gewichtung einbezogen, im Übrigen wie Sz 1
Standardfehler gelten für Großraum, nicht für definiertes Zielgebiet, nur bedingt vergleichbar
Definition der Szenarien: s. 5.5

Abb. 33: Einfluss der Gewichtungsabstufung von Anbaugebieten gegenüber einer Großraumauswertung auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen

Die Einbeziehung von Nachbargebieten hat einen erheblichen Effekt auf die Mittelwertschätzungen (Sz 1 versus Sz 7). Bei dem in Abb. 32 hervorgehobenen Wert handelt es sich konkret um eine Sorte, die zwar nicht zum Kernsortiment zählt, aber aufgrund ihres relativ geringen Standardfehlers durchaus mit dem Kernsortiment verglichen werden könnte. Die Besonderheit dieser speziellen Sorte *Winnetou* liegt in einer äußerst untypischen Stichprobenverteilung über die Regionen. Aufgrund bereits im Prüfprozess beobachteter Regionalspezifik wurde sie über die Anbaugebiete sehr unterschiedlich besetzt. Die Einbeziehung der Nachbargebiete bei Sz 1 führt bei dieser Sorte mit insgesamt 42 Versuchsergebnissen zu einem sehr großen absoluten Mittelwertunterschied (ca. 5,0 dt/ha) gegenüber dem Sz 7, bei dem nur sieben Versuchsergebnisse (per Definition nur aus dem Zielgebiet) verfügbar waren.

Die Mittelwertunterschiede zwischen Sz 1 und dem Großraumansatz (Sz 8) sind in diesem konkreten Fall geringer (< 1,5 dt/ha) (Abb. 33).

6.3.4 Diskussion

genetische Korrelationen

Westlich von den Zielgebieten in Mecklenburg-Vorpommern (MV) gelegene Nachbargebiete zeigten tendenziell geringere genetische Korrelationen zum jeweiligen Zielgebiet, als das südlich gelegene Nachbargebiet. Ursache für diese Spezifik der westlichen Gebiete könnten die dort tendenziell bevorteilten Niederschlagsverhältnisse, besseren Böden und in Folge dessen auch höheren Anbauintensitäten sein. Im Ergebnis treten in diesen Anbaugebieten abweichende Sorteneffekte gegenüber *trockeneren* Zielgebieten in MV auf. Für die gegenüber dem Zielgebiet noch trockeneren Standorte im Süden von MV war vom Autor der vorliegenden Arbeit eine hohe Spezifik erwartet worden. Ursachen für die dahingegen gefundene eher geringe Spezifik könnten insbesondere in erhöhten S × J - Interaktionen in Trockengebieten liegen. Der Wechsel von niederschlagsreicheren und -armen Jahren lässt gerade auf diesen Standorten hohe S × J - Interaktion erwarten, so dass Anbaugebiets-Spezifik im langjährigen Mittel verwischt würde. Varianzheterogenität bezüglich des Effektes S • J • AG in der Modellgleichung, differenziert für unterschiedliche Anbaugebiete, ist im Auswertungsmodell bislang nicht vorgesehen, sollte allerdings bei künftigen Weiterentwicklungen in Betracht gezogen werden. Zur Bewertung z.B. der Trockentoleranz von Sorten ist also das mehrjährige Mittel in von Trockenheit häufiger betroffenen Anbaugebieten (hier insbesondere Ṣ bzw. D-Süd) nicht hinreichend geeignet. Hierfür sind weiterführende Analysen unter Einbeziehung von Parametern zur Charakterisierung von Trockenheitsstress bzw. von Parametern der Ökostabilität erforderlich.

Geringe bzw. nicht signifikante Anbaugebiets-Spezifik sollte aus o.g. Gründen eher im Sinne von „mit dem vorliegenden Datensatz und zugrunde gelegten Modell nicht hinreichend nachweisbar“ als im Sinne von „nicht vorhanden“ interpretiert werden.

Nur bei den Ertragsauswertungen der beiden Pflanzenarten mit der umfangreichsten Datenbasis - Winterweizen und Winterraps - wurden zwei Zielgebiete (statt einem) für Standorte in (und um) Mecklenburg-Vorpommern gebildet. Beim Raps wurde ein deutlicher Unterschied in der Sortenspezifik zwischen diesen beiden Gebieten gefunden. Beim Winterweizen war dies nicht zu erkennen und insofern sollte geprüft werden, ob diese Unterteilung zu Gunsten eines größeren Zielgebietes mit dadurch entsprechend größerem Datensatz aufgegeben oder anderweitig verändert werden sollte.

Bei einigen Merkmalen wurden für einzelne Nachbargebiete vergleichsweise niedrige genetische Korrelationen zum Zielgebiet im Bereich $0{,}5 < r < 0{,}6$ gefunden. Es traten aber keine derart niedrigen genetischen Korrelationen auf, dass die Einbeziehung eines Nachbargebietes prinzipiell als ungeeignet betrachtet werden könnte. Der Ausschluss bestimmter Nachbargebiete ist demnach eher standortkundlich zu entscheiden als über ggf. sehr geringe genetische Korrelationen. So werden z.B. durch den Autor vorliegender Arbeit aus fachlichen Gründen wie auch aus Gründen der Akzeptanz der Ergebnisse in der Branche prinzipiell nur Anbaugebiete mit diluvialer geologischer Herkunft einbezogen, jedoch keine Löß-, Marsch, -Verwitterungs- oder andere Böden. Diese Grundsatzentscheidung erfolgte unabhängig von den potenziellen genetischen Korrelationen zwischen Böden völlig unterschiedlicher geologischer Herkunft.

Bei Merkmalen ohne jegliche gefundene Anbaugebietsspezifik der Sorteneffekte ist zu überdenken, auf die Unterteilung in Anbaugebiete innerhalb des einbezogenen Raumes künftig zu verzichten.

Przystalski et al. (2008) verwenden ebenfalls den Parameter ‚genetische Korrelation' zur Beurteilung der Ähnlichkeit der Leistung von Getreidesorten im ökologischen gegenüber dem konventionellen Landbau. Diese Untersuchung erfolgte über verschiedene europäische Länder. Auch hier wurden relativ hohe genetische Korrelationen gefunden (0,79 bis 1,0), wobei es sich hier nicht um die Korrelation zwischen Gebieten sondern zwischen Anbausystemen handelte.

Für Subregionen des südlichen und östlichen Afrikas untersuchten Weber et.al. (2012) geeignete Regionalisierungsansätze für Mais. Sie fanden für geografische und klimatische Parameter kaum Ansätze - Mais schien sehr breit adaptiert zu sein. Dagegen fanden sie deutliche Genotyp × Ertragslevel Interaktionen. Sie schlagen anstelle von Regionen eine Unterteilung in die zwei Standortgruppen ‚niedriger Ertrag' und ‚hoher Ertrag' vor, wobei eine gewichtete Einbeziehung des jeweils anderen Gebietes erfolgen sollte. Die genetische Korrelation zwischen diesen beiden ‚Regionen' betrug für zwei unterschiedliche Reifegruppen 0,81 bzw. 0,69. Aus Sicht des Autors dieser Arbeit stellt dieses Ergebnis aber alternativ auch einen Hinweis auf erfolgversprechende Untersuchungen zur Umweltstabilität der Genotypen unter Berücksichtigung der ‚Umweltleistung' jedes Einzelstandortes bzw. Einzelversuches dar (Ökoregression).

Der überwiegend genutzte Modellansatz für die Schätzung der genetischen Kovarianzmatrix lässt zwar für jedes Anbaugebiet eine eigene S × AG - Varianz zu, die damit im Zusammenhang stehenden genetischen Korrelationen sind aber nur bedingt als tatsächlich paarweise interpretierbar. Hierfür wäre z.B. ein faktoranalytischer Ansatz (Piepho, 1997) weiter führender, welcher aber für die vorliegenden Datensätze bei der Komplexität des Modells nicht hinreichend geeignet war. Der heterogene Modellansatz erkennt für zwei AG mit höheren Interaktionsvarianzen z.B. nicht, wenn beide AG eher gleichgerichtet vom *Durchschnittsverhalten* im Großraum abweichen - in solchen Fällen wird die paarweise Korrelation unterschätzt. So treten z.B. beim Ertrag von Raps die beiden höchsten S × AG - Varianzen in den beiden Anbaugebieten mit dem bekanntermaßen höchsten Risiko von Trockenheitsstress auf: MV-Süd und Ṣ (6.2.1; Tab. 30). Die hohen Varianzen erscheinen daher sehr plausibel, die daraus rechnerische hergeleitete vergleichsweise geringe Paar-Korrelation $(r = 0{,}75$; Tab. 35) aus standortkundlichen Erwägungen aber weniger. Daher sollten die paarweisen genetischen Korrelationen nach den in dieser Arbeit genutzten Schätzverfahren hinsichtlich eines fachlichen Vergleichs der Sorteneffekte zweier Anbaugebiete nicht überbewertet werden. Entscheidend für derartige Sekun-

därauswertungen wäre eher eine Analyse der tatsächlich geschätzten Sorteneffekte in diesen Gebieten.

Muster in der Gewichtung von Nachbargebieten

In der Analyse des Kornertrages bei Winterweizen zeigte sich, dass das jeweilige Zielgebiet fast ausnahmslos höher gewichtet wurde als jedes einzelne Nachbargebiet. Andererseits gingen die Sorteneffekte im Zielgebiet in die gewichteten Mittelwerte über alle einbezogenen Gebiete nur mit einem Gewicht unter 50% ein. Einzelfall-Abweichungen hiervon sind ausnahmslos mit sehr Sorten spezifischen Datenbesetzungen in den unterschiedlichen Anbaugebieten zu erklären. Somit erbringt die Summe der einbezogenen Nachbargebiete einen sehr hohen Beitrag für die Sorteneinschätzungen in einem Zielgebiet. Dies zeigt die praktische Relevanz des neuen Konzeptes in der Regionalprüfung in Deutschland, Nachbargebiete bei der Beurteilung von Sorten in einem Zielgebiet einfließen zu lassen.

Im frühen Prüfstadium bestimmen einzelne Versuchsergebnisse auch aus Nachbargebieten in relevantem Maße den Gesamtmittelwert einer Sorte. Im Verlauf des Prüfprozesses steigt die Eigengewichtung des Zielgebietes, Einzelergebnisse, insbesondere aus Nachbargebieten, verlieren zunehmend an Gewicht und die Mittelwerte werden somit robuster. Dies entspricht einer Herangehensweise, bei der die regionale Bewertung der Sorten mit fortschreitender Prüfdauer - beginnend mit einer großräumigen Sicht - zunehmend räumlich feiner skaliert und auf das Zielgebiet fokussiert wird. Es gibt keine starre vordefinierte Gewichtungsrelation, sondern letztlich erfährt jede Sorte vor dem Hintergrund ihrer individuellen Prüfsituation eine adäquate / optimale Gewichtungs-Konstellation. Darin liegt vor dem Hintergrund der erheblichen Unbalanciertheit und Komplexität der Datenstruktur ein nicht zu unterschätzendes, durch das Auswertungsmodell bereitgestelltes Optimierungspotenzial. Diese Optimierung berücksichtigt, wie bereits betont, sowohl die regionale Verteilung in der Datenbesetzung als auch die regionale Korreliertheit.

Effekte der Einbeziehung von Nachbargebieten auf Mittelwerte im Zielgebiet und ihre Genauigkeit

Der Prüfumfang je Zielgebiet ist derzeit bereits zu gering, um ‚*aus eigener Kraft*', also nur mit Daten des Zielgebietes, sehr frühzeitig genaue Sorteneinschätzungen zu erzielen - ein Strategiewechsel zur Effizienzsteigerung stand insofern tatsächlich an.

Es konnte gezeigt werden, dass eine gewisse Verzerrung der Sorteneffekte im Zielgebiet durch die Einbeziehung von Sorteneffekten in Nachbargebieten zugunsten einer deutlich gesteigerten Präzision in Kauf genommen werden kann. Mit der neuen Strategie (Sz 1) wird mit dem LSV1 bereits eine Genauigkeit erreicht, die nach dem traditionellen Herangehen (Sz 7) erst nach dem LSV2 erzielt werden kann. Nach LSV2 erbringt Sz 1 eine Genauigkeit, die Sz 7 im Sinne der Zielstellung (frühzeitige Überleitung von Züchtungsfortschritt) erst nach einer irrelevant langen Prüfdauer erreichen würde. Der Zeitgewinn allein durch die Einbeziehung der Nachbargebiete beträgt mindestens ein Jahr. Diese Ergebnisse bestätigen die Aussagen von Curnow (1988) und Atlin et al. (2000), dass die Information aus Nachbargebieten einen nützlichen Beitrag für ein Zielgebiet liefern können.

Kleinknecht et al. (2013) analysierten für fünf Maisanbaugebiete in Indien zwei vergleichbare Szenarien. Zum einen erfolgten BLUP-Auswertungen, wobei analog zum hier in 5.3.6 dargelegten Ansatz für jedes Anbaugebiet basierend auf den genetischen Korrelationen die Informationen der vier Nachbargebiete einbezogen wurden.

Zum anderen wurde analog zum hier definierten Szenarium 7 eine BLUE-Schätzung separat für jedes Anbaugebiet ohne Nutzung der Nachbargebiete durchgeführt. Es erfolgten Simulationen für die Kalkulation der Korrelationen der vorhergesagten/ geschätzten genotypischen Werte zu den simulierten ‚wahren' Werten in einem Zielgebiet. Bei variierenden genetischen Korrelationen wurden in 38 von 40 Fällen wie auch im Durchschnitt erhebliche Vorteile von BLUP unter Einbeziehung von Nachbargebieten gegenüber einer BLUE-Schätzung ohne Einbeziehung von Nachbargebieten gefunden. Die Autoren sehen dies ursächlich vor allem in der Einbeziehung der Nachbargebiete begründet, erst nachrangig aber auch in Vorteilen von BLUP gegenüber BLUE.

Die neue Strategie (Sz 1) ist auch einer Großraumstrategie (Sz 8), bei der die einbezogenen Anbaugebiete aufgelöst und zu einem Großraum zusammengefasst würden, in allen Phasen der Prüfung überlegen. In Fällen, bei denen die genetischen Korrelationen zum Zielgebiet geringer ausfallen als in diesem Beispiel des Kornertrages von Winterweizen könnte sich durchaus ein geringerer Vorhersagefehler für den Großraum als für kleinere Zielgebiete ergeben. In diesen Fällen ist aber die Verzerrung des Großraummittelwertes je Sorte gegenüber dem Sortenmittelwert für ein kleineres Zielgebiet nach wie vor zu berücksichtigen, ohne dass diese bei der alleinigen Großraumauswertung messbar wäre. Für ein jeweiliges Zielgebiet wäre auch dann das Sz 1 optimal, solange die einzelnen Anbaugebiete wenigstens so gut besetzt sind, dass die Varianzkomponenten und Korrelationen ‚hinreichend' gut schätzbar sind und solange nicht alle genetischen Korrelationen 1,0 betragen. Die neue Strategie ermöglicht eine frühzeitige relativ genaue Einschätzung für einzelne Zielgebiete bereits in einem relativ frühen Prüfstadium.

Die abgestuft gewichtete Einbeziehung von Nachbargebieten führt beim Kornertrag von Winterweizen zu relevanten Veränderungen der Sortenmittelwerte und -differenzen. Von allen eingeleiteten Neuerungen hat sie hierin nach dem Kombinationseffekt ‚langjährig + Einbeziehung vorgelagerter Versuche' die zweitgrößte Bedeutung. Im Unterschied zu Letzterem ist der Zusammenhang zum Prüfstatus bzw. zum Umfang der Datengrundlage der einzelnen Sorten aber schwächer. Im Einzelfall können erhebliche Mittelwertabweichungen (> 2,0 dt/ha) gegenüber der getrennten Auswertung je Anbaugebiet auftreten.

Dagegen ist die Übereinstimmung der neuen Strategie mit einer Großraumstrategie deutlich größer. Allerdings wurde mit dem Kornertrag von Winterweizen auch ein Beispiel gewählt, bei dem bei zwei von drei Nachbargebieten eine genetische Korrelation von 1,0 zum Zielgebiet geschätzt wurde (Tab. 35). Bei den meisten Pflanzenart × Merkmal - Kombinationen sind aufgrund geringerer genetischer Korrelationen größere Abweichungen zu erwarten.

Insofern sollte die Strategie der Einbeziehung von Nachbargebieten mit optimal abgestuftem Gewicht zur Mittelwertschätzung für ein Zielgebiet die Priorität haben, solange ein ausreichender Datenumfang zur hinreichend genauen Schätzung der Varianzkomponenten (und damit auch der Korrelationsstrukturen) vorhanden ist. Vor diesem Hintergrund soll an dieser Stelle auch davor gewarnt werden, das Versuchsstellennetz weiter zu reduzieren.

In der vorliegenden Arbeit ist die Regionalisierung rein standortkundlich vordefiniert, während die Gewichtung der Anbaugebiete Versuchsdaten basiert erfolgt. Ramburan et al. (2012) beschreiben für die Sortenprüfung bei Zuckerrohr in Südafrika Klassifizierungsansätze für Versuchsstandorte, die auf der gemeinsamen Analyse beider fachlicher Grundlagen erfolgt. Sie fanden für Bodenparameter deutliche Übereinstimmungen zur Versuchsdaten basierten Clusterung von Versuchsstandorten,

während dies bei Klimaparametern nicht in dem Maße beobachtet wurde. Ein derartig kombinierter Ansatz wäre für die Weiterentwicklung der Regionalisierungsansätze für die Anwendung der Hohenheim-Gülzower-Serienauswertung künftig auch in Betracht zu ziehen.

Die hier diskutierten Zusammenhänge werfen die Frage nach der geeignetsten Definition der Anbaugebiete allgemein und speziell auch die der Nachbargebiete nochmals auf. Die getrennte Berücksichtigung der Nachbargebiete MV-Süd und ṢẆ für das Zielgebiet D-Nord hat sich bei dem in dieser Arbeit untersuchten Merkmal Kornertrag von Winterweizen nicht bewährt. Für beide Nachbargebiete wurde zum Zielgebiet und auch zueinander eine genetische Korrelation von 1,0 geschätzt. Dies führt bei der vorliegenden Datenstruktur dazu, dass die Einzelversuche der Nachbargebiete überwiegend ein höheres Gewicht erhalten, als wären diese Nachbargebiete miteinander oder auch mit dem Zielgebiet zusammengefasst worden. Erfahrungsgemäß neigen viele Versuchsansteller dazu, Nachbargebiete aus sachlichen Gründen eher zu verkleinern als zu vergrößern. Erzielt wird damit eher ein nicht angestrebter und subjektiv nicht erwarteter Effekt. Die aus Sicht der Regionalität verbreitete konservative Strategie, das Zielgebiet im Zweifelsfall eher über- als unterzugewichten, wird - in Abhängigkeit von den regionalen Korrelationen - eher durch ausreichend große, ggf. zusammengelegte Nachbargebiete erzielt, als durch Verkleinerung und Trennung. Eine Situation, in der die einzelnen Nachbargebiete deutlich geringer besetzt sind als das Zielgebiet, wird daher vom Autor als suboptimal bis kontraproduktiv angesehen. Wenn es die genetischen Korrelationen und standortkundlichen Betrachtungen sinnvoll erscheinen lassen, sollten einzelne sich berührende Nachbargebiete in der Weise zusammengefasst werden, dass der Prüfumfang im Vergleich zum Zielgebiet ausgewogen, im Zweifelsfall eher etwas größer als im Zielgebiet ausfällt. In diesem Sinne sollten im Ergebnis dieser Arbeit im Nachgang auch die hier zunächst vordefinierten Regionalisierungen (s. 5.1.3) überdacht und angepasst werden.

Alternative Herangehensweisen - jüngste Literaturquellen

In australischen Pflanzenzüchtungsprogrammen werden zunehmend faktoranalytische Ansätze für die Modellierung additiver Genotyp × Umwelt - Interaktionen genutzt. Dabei werden genetische Korrelationen zwischen allen Paaren von Versuchen bzw. Orten unter Einbeziehung der Abstammungs- / Verwandtschaftsinformation der Genotypen geschätzt (Cullis et al., 2014).

Auch für das nationale australische Sortenprüfungs-Programm sehen Smith et al. (2015) Ansätze zur Verbesserung der Schätzung von Sorteneffekten durch Nutzung faktoranalytischer Ansätze zur Modellierung der Genotyp × Umwelt - Interaktion. Sie beschreiben das ‚*Varianzkomponentenmodell*' für die Langzeitauswertung von Sortenversuchen als nicht adäquat zur Modellierung der Genotyp × Umwelt - Interaktionen. Bei Ergebnisdarstellung auf einem regionalen Level (wie in der deutschen Regionalprüfung) würden dadurch wichtige lokale Genotyp × Umwelt - Interaktionen maskiert werden. Sie schlagen ein nutzerfreundliches Informationssystem vor, dass deutlich stärker lokale Informationen für den individuellen Landwirt beinhaltet.

Inwieweit diese Ansätze für das Sortenprüfsystem in Deutschland geeignet sind, sollte Gegenstand nachfolgender Untersuchungen werden. Von Interesse ist dabei insbesondere, ob die paarweisen Korrelationen zwischen Orten / Versuchen hinreichend präzise schätzbar sind und ob diese Informationen tatsächlich zu reproduzierbareren Informationen für den lokalen Landwirt führen.

6.4 Diskussion der Problematik: Faktor ‚Sorte' als fix oder als zufällig definieren?

6.4.1 BLUE- und BLUP- Ansatz für Sorteneffekte

Patterson (1997) betrachtet in *Sortenprüfsystemen* jede Sorte als distinktes ‚Wesen'. Für einen anderen Anwendungsbereich - die *Pflanzenzüchtung* - diskutiert Van Eeuwijk (2007) für Versuchsserien die Wahl der Faktoren im Modell als *fix* oder als *zufällig*. Obgleich er für beide Ansätze gute Gründe anerkennt, hält er auch für den Faktor *Genotyp* bzw. *Sorte* die Entscheidung ‚*zufällig*' für besser. Piepho und Möhring (2006) zeigen für pflanzenzüchterische Versuche Effizienzgewinne von BLUP gegenüber BLUE. Kleinknecht et al. (2013) sehen Vorteile von BLUP umso ausgeprägter, je größer der Grad der Unbalanciertheit ist. Dagegen zeigte sich bei geringer Unbalanciertheit in einer Auswertung deutscher Wertprüfungen mit Mais nur ein marginaler Unterschied zwischen BLUP und BLUE (Kleinknecht et al., 2011). In der Tierzüchtung erfolgen Zuchtwertschätzungen bereits routinemäßig über die Nutzung von Informationen verwandter Tiere, BLUP wird dafür routinemäßig genutzt (Henderson 1984). Auch in der Pflanzenzüchtung gibt es zunehmend Anwendungen für BLUP, u.a. bei der genomischen Selektion auf quantitative polygen bedingte Merkmale. Ziel ist es dabei häufig, für ein derartiges Merkmal wirksame Chromosomenregionen bzw. Genorte (Quantitative Trait Loci bzw. QTL) in ihrer relativen Position zu genetischen Markern zu lokalisieren und mittels dieser Marker die Selektion zu verbessern und zu beschleunigen. Fernando und Grossmann (1989) demonstrieren die Einbeziehung von Markerinformationen in die Zuchtwertschätzung durch BLUP. Meuwissen et al. (2001) vergleichen Kleinst-Quadrat-, BLUP- und Bayessche-Analysen in der Vohersagegenauigkeit der Zuchtwertschätzung in Situationen mit wenigen Individuen und vielen Markern. Für gewisse Situationen beschreiben sie dabei Probleme mit BLUP infolge überhöhter Regression gegen null (durch Schrumpfung). Heffner et al. (2009) kommen zum Ergebnis, dass sehr gute Vorhersagen des Zuchtwertes auch für quantitative Merkmale mit geringer Heritabilität möglich sind, wenn hoch aufgelöste Markerkarten vorliegen (hohe Markerdichte) und das Auswertungsmodell alle Markerinformationen einbezieht.

In der Sortenprüfung können im Unterschied zu o.g. Anwendungen Verwandtschafts- oder Markerinformationen nicht genutzt werden. Die in dieser Arbeit vorgestellte Nutzung der Information benachbarter Anbaugebiete stellt allerdings einen vergleichbaren Fall dar: es werden Informationen anderer Subjekte aufgrund von Ähnlichkeit genutzt, in diesem Fall nicht Ähnlichkeit von Ausprägungen aufgrund genetischer Verwandtschaft, sondern aufgrund standortkundlicher Gegebenheiten.

Eine immanente Eigenschaft von BLUP ist die Schrumpfungsschätzung. ‚Rohschätzer' werden durch Nutzung zusätzlicher Informationen im Hinblick auf die mittleren quadrierten Fehler der Vorhersage (MSEP) ‚verbessert', wobei sie zu einer fixen Konstanten hin geschrumpft werden (hier z.B. jede Sorte zum Mittelwert aller Sorten oder einer Sortengruppe). In den folgenden Abschnitten soll insbesondere hinterfragt werden, inwieweit Schrumpfung mit den Erfordernissen und Gegebenheiten speziell der offiziellen Sortenprüfung, ggf. im Unterschied zur Züchtung, vereinbar ist.

Im Zusammenhang mit der in dieser Arbeit diskutierten Problematik stehen die in Tab. 41 aufgeführten Unterschiede zwischen beiden Ansätzen in Auswertung und Interpretation an.

Tab. 41: Unterschiede zwischen BLUE- und BLUP- Ansatz für Sorteneffekte in der Hohenheim-Gülzower-Serienauswertung

Sorte als ‚*fix*' definiert	**Sorte als ‚*zufällig*' definiert**
Effekte ‚S' und 'S • AG' im Modell zur Mittelwertschätzung als ‚*fix*' gesetzt	Effekte 'S' und 'S • AG' im Modell zur Mittelwertschätzung als ‚*zufällig*' gesetzt
BLUE-Schätzer für 'S' und 'S × AG' - Effekte	BLUP-Schätzer für 'S' und 'S × AG' - Effekte
Keine Schrumpfungsschätzung der 'S' und 'S × AG' - Effekte	Schrumpfungsschätzung der 'S' und 'S × AG' - Effekte zum jeweiligen Mittel hin
Summenrestriktion für die Gewichtung der Anbaugebiete bei der Schätzung der 'S × AG' - Effekte; $\Sigma_r w_r = 1$ (wobei r: Zahl der Gebiete; w_r: Gewicht des Gebietes r)	Keine Summenrestriktion für die Gewichtung der Anbaugebiete bei der Schätzung der 'S × AG' - Effekte; $\Sigma_r w_r \leq 1$ (wobei r: Zahl der Gebiete; w_r: Gewicht des Gebietes r)
Nutzung des Makros *‚%auswertung*' entsprechend Tab. 20	Nutzung des Makros *‚%lsm*' entsprechend Tab. 20
Jede einzelne Sorte wird als Individuum betrachtet und ohne Schrumpfung eingeschätzt und bewertet, ggf. zu Lasten optimaler Entscheidungen über die Gesamtheit aller Sorten	Sorten werden als Elemente einer Grundgesamtheit bzw. einer Population betrachtet. Im Focus stehen optimale Entscheidungen über die Gesamtheit aller Sorten hinweg (z.B. Selektion), ggf. zu Lasten der ‚gerechten' Einstufung des Einzel-'Individuums'.

6.4.2 Differenzierte Anforderungen in den Phasen der züchterischen Selektion und der nachfolgenden amtlichen Sortenprüfung

Methodische Untersuchungen zu Versuchsserien mit Genotypen landwirtschaftlicher Nutzpflanzen werden häufig an Problemstellungen aus der Züchtung durchgeführt, wie auch bei van Eeuwijk (2007). Als Zielstellung wird dann in der Regel eine optimale Selektionsentscheidung über alle Genotypen hinweg unterstellt. Häufig handelt es sich bei den Genotypen um Kreuzungsnachkommenschaften, um Elemente einer Population, für die u.a. eine Verteilungsfunktion hinreichend angenommen werden kann. Die bestmögliche Gesamtentscheidung, die Minimierung des MSEP hat Vorrang vor einzukalkulierenden falsch-positiven und falsch-negativen Selektionsentscheidungen bezüglich einzelner Genotypen. Die Gesamtentscheidung trifft ein Versuchsansteller, z.B. ein Züchtungsunternehmen, für sich und es gibt keinerlei Interessenkonflikte.

Potenzielle Vorteile von BLUP gegenüber BLUE können bei derartigen Anwendungen heute als unstrittig gelten. Der Ansatz ‚zufällig' für den Faktor *Sorte* oder

Genotyp erscheint in der Selektionsphase der Pflanzenzüchtung fachlich relativ unproblematisch. Der Selektionserfolg lässt sich durch BLUP gegenüber BLUE steigern (s.a. 6.4.1).

Wenn die Wahl der im ausgewerteten Merkmal oder Merkmalskomplex ‚besten' Sorte als alleiniges oder primäres Ziel definiert ist, so wäre dieses auch auf die amtliche Sortenprüfung weitgehend übertragbar. Entscheidungen für die Sortenprüfung sollten allerdings nicht unkritisch aus Empfehlungen für die Pflanzenzüchtung übernommen werden. Die offizielle Sortenprüfung unterscheidet sich vom züchterischen Selektionsprozess u.a. dadurch, dass die ‚gerechte' Schätzung je Individuum (Sorte) zunehmende Priorität erhält, z.B. im offiziellen Zulassungsverfahren je Sorte beim Bundessortenamt. Wenn auch aus Sicht der landwirtschaftlichen Praxis einer optimalen Gesamtentscheidung eine gewisse Priorität beigemessen werden kann, so hat doch jeder einzelne Antragsteller, also z.B. ein Züchtungsunternehmen, einen rechtlichen, juristisch einklagbaren Anspruch auf eine ‚gerechte' Beurteilung einer ganz konkreten Sorte. Die Schrumpfungsschätzung könnte erhebliche, insbesondere bei Spezialsorten mit spezifischem Eigenschaftsprofil auch ungerechtfertigte Auswirkungen auf einzelne Unternehmen haben. Eine Schrumpfung z.B. der in einem Einzelmerkmal ‚besten' Sorten zum Sortenmittel hin wäre ebenso wenig vermittelbar und justiziabel wie im Schulsystem eine graduelle Korrektur von Noten der ‚Bestschüler' in Richtung Klassendurchschnitt.

Mit der Sortenzulassung in Deutschland wurde einer Sorte „landeskultureller Wert" bestätigt, sie liegt aufgrund der Anforderungen an die Zulassung im Spitzenbereich der vertriebsfähigen Sorten. Die weitere Prüfung in den Regionen hat nur noch anteilig und sehr bedingt das Ziel einer straffen Selektion. Das Ziel verschiebt sich mit fortschreitendem Prüfprozess zunehmend weg von der Selektion hin zu einer vergleichenden Sortenbeschreibung über z.T. sehr viele univariat ausgewertete Sorteneigenschaften. Dabei soll insbesondere auch das spezifische Profil jeder einzelnen anbaurelevanten Sorte über alle Eigenschaften hinweg herausgearbeitet werden. Dem Landwirt wird nicht vordergründig eine ‚beste Sorte' empfohlen, sondern er bekommt für eine Palette sehr guter regional empfohlener Sorten eine detaillierte Beschreibung der wichtigsten Sorteneigenschaften an die Hand. Dies ist eine Grundlage nicht nur für die Wahl der Sorten, sondern insbesondere auch für eine gezielte Verteilung des betrieblichen Sortenspektrums auf unterschiedliche Anbaubedingungen (z.B. in Abhängigkeit von Saattermin, Vorfrucht, Bodenbearbeitung, Bodengüte, administrativen Anbauauflagen u.v.m.). Zudem geben Sortenbeschreibungen im Beratungsmaterial Hinweise, welche u.U. kritischen Einzelmerkmale einer insgesamt sehr guten Sorte sortenspezifisch in der Anbautechnik berücksichtigt werden sollten. Die Komplexität von Sortenbeurteilung, Sortenwahl und Einordnung von Sorten in eine betriebliche Situation ist u.a. bei Michel und Pienz (2007) und bei Michel und Zenk (2010) dargelegt.

Diese Verschiebung der Einordnung des Faktors Sorte im Verlauf des Züchtungs- und Prüfsystems ist grob schematisch in Abb. 34 veranschaulicht. Das Schema soll neben dem Prioritätenwandel aber auch deutlich machen, dass es in keiner Phase eine im absoluten Sinne eindeutige Aussage zur Definition ‚fix' versus ‚zufällig' für den Faktor ‚Sorte' gibt.

Prozess	**Vorrangige Sicht auf den Faktor ‚Sorte' und Aufgaben und Ziel der Prüfung**	
	Genotyp als Element einer Population	Sorte als ‚Individuum'
	Selektion	Schätzung
	Rangfolgen	ungeschrumpfte Mittelwerte und Differenzen
	beste Entscheidung über gesamte Population	beste Einschätzung jeder Einzelsorte
Selektion im Züchtungsunternehmen		
nationale Wertprüfung		
Zulassung von Sorten mit ‚landeskulturellem' Wert		
Regionalprüfung zugelassener Sorten (LSV)		
begleitende anbautechnische Sortenversuche		

Abb. 34: Prioritätenwandel bei der Einordnung des Faktors ‚Sorte' im Verlauf des Prüfprozesses - schematisch

In der Phase der amtlichen Sortenprüfung zugelassener Sorten wird die Definition ‚zufällig' für den Faktor Sorte zunehmend problematisch. Im balancierten Fall (unrealistisch!) würde eine Schrumpfungsschätzung mit BLUP die Rangfolgen in einem Einzelmerkmal nicht verändern. Basis für eine Gesamtbewertung sind aber nicht die Ränge je Einzelmerkmal, sondern die geschätzten Effekte auf der metrischen Skala. Nur diese Effekte sind auch wirtschaftlich bewertbar und können im Vergleich zu Effekten in anderen Merkmalen gewichtet werden und in eine Merkmal übergreifende Gesamtbewertung einfließen. Geschrumpfte Effekte je Einzelmerkmal würden die vergleichende Gesamtbewertung über alle Merkmale maßgeblich beeinflussen. Im unbalancierten, also realistischen Fall kommt die Problematik nicht unerheblicher Rangfolgeänderungen auch in jedem Einzelmerkmal hinzu. Im Ergebnis wäre mit erheblichen Akzeptanzproblemen von BLUP bei Sorteninhabern wie auch bei Informationsnutzern in der landwirtschaftlichen Praxis zu rechnen.

Jeder Versuchsansteller sollte sich die konkrete Zielstellung einer Versuchsserienauswertung vor Augen zu führen und nicht pauschal eine der beiden Sichtweisen - fix versus zufällig bzw. BLUP versus BLUE - übernehmen.

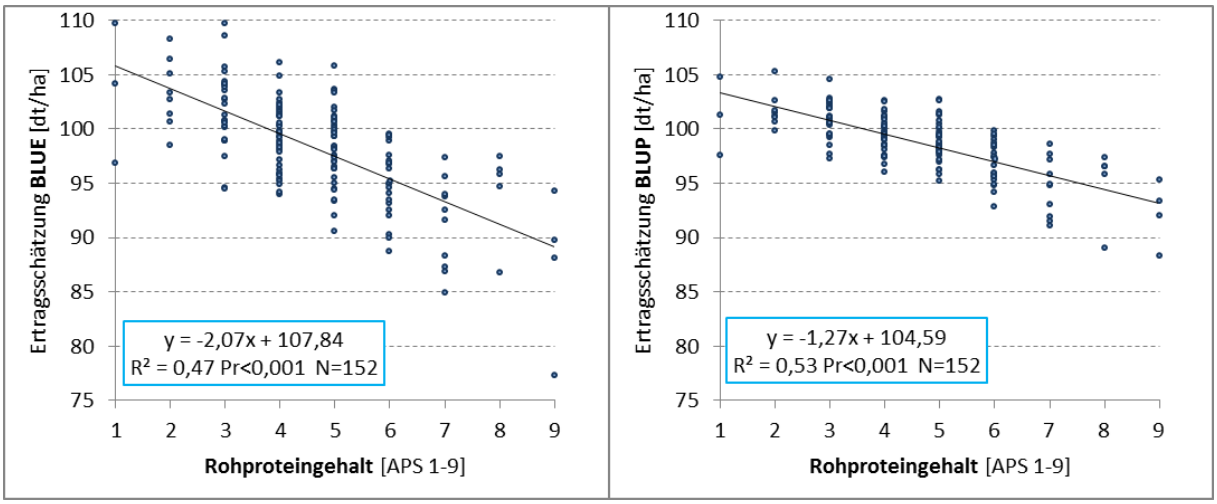
Patterson (1997) plädiert für die Sichtweise, eine Sorte in offiziellen Prüfungen als fix, als distinktes Wesen zu betrachten. Trotz der in Abb. 34 angedeuteten gleitenden Übergänge in den Prioritäten schließt sich der Autor vorliegender Arbeit für die Auswertung von Landessortenversuchen dieser Sicht an. Wie auch allgemein im deutschen Sortenprüfsystem aktuell üblich wird nicht die Gesamtheit der Sorten als Stichprobe mit gemeinsamer Verteilung, sondern jede einzelne Sorte vorrangig als distinktes (fixes) Subjekt angesehen, das jeweils auf Grund ihrer ganz spezifischen

Kombination wertbestimmender Eigenschaften zur Zulassung, zur Aufnahme in den LSV oder sogar bis zu einer regionalen Empfehlung kam und für das eine ungeschrumpfte Schätzung des Sorteneffektes in jedem Einzelmerkmal anzustreben ist.

6.4.3 Verzerrungspotenzial durch Schrumpfungsschätzung bei unterlassener Berücksichtigung von mit dem Zielmerkmal korrelierten wertbestimmenden Merkmalen

Im Sortenprüfwesen erfolgen derzeit fast ausschließlich univariate Auswertungen, die Sorten werden also zunächst in jedem Merkmal isoliert eingeschätzt. Eine Gesamtbewertung erfolgt danach über eine z.T. sehr große Anzahl wertbestimmender Eigenschaften in Form einer fachlichen Abwägung oder einer Indexberechnung mit Gewichtung der Effekte in den Einzelmerkmalen. In der aktuellen Beschreibenden Sortenliste (Bundessortenamt, 2014) werden Winterweizensorten in immerhin 30 wertbestimmenden Eigenschaften beschrieben. Zwischen diesen Merkmalen ist z.T. (positive oder negative) Korreliertheit zu beobachten. Nachfolgend soll der Frage nachgegangen werden, ob dieser Aspekt für die Entscheidung von BLUP versus BLUE Relevanz besitzt.

Für die Pflanzenart Winterweizen soll beispielhaft das Merkmal Kornertrag unter Berücksichtigung der Korreliertheit zum Proteingehalt betrachtet werden. Bereits Piepho und Möhring (2005) haben eine besondere Relevanz von Backqualitäts-Eigenschaften für die Ertragsschätzung beim Weizen ausgemacht. Die nachfolgend zugrunde gelegte univariate Auswertung für den Kornertrag erfolgte für den BLUE- und für den BLUP-Ansatz jeweils mit den der Hohenheim-Gülzower-Serienauswertung zugrunde liegenden SAS® - Makros (s. 5.4). Ein adäquater regionaler Datensatz für den Proteingehalt lag nicht vor, da Analysen bei diesem hoch heritablen Merkmal in der Regionalprüfung i.d.R. nur an einer geringen Anzahl Proben erfolgen. Daher wurde für den Proteingehalt die überregionale Einstufung der Sorten in der Beschreibenden Sortenliste (Bundessortenamt, 2009) genutzt. Diese Einstufung erfolgte beim Bundessortenamt durch proportionale Umwandlung der Schätzwerte im Originalmerkmal in eine Note (APS = Ausprägungsstufe; 1-9). Sofern man diesen Schritt als grobe Rundung des verhältnisskalierten Merkmals betrachtet (denn die Erfassung erfolgte nicht ordinalskaliert), ist die Schätzung einer funktionalen Beziehung zwischen Ertrag und Proteingehalt trotz des Informationsverlustes durch Rundung vertretbar.



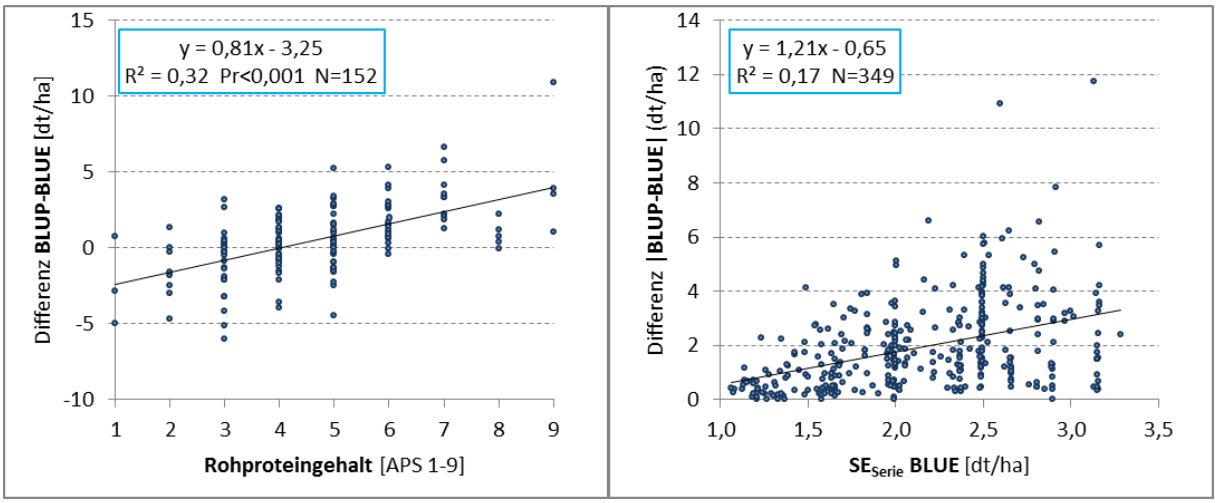
APS - Ausprägungsstufe nach Bundessortenamt (2009)

Abb. 35: Zusammenhang von Proteingehalt und BLUE- versus BLUP-Ertragsschätzungen bei Weizensorten

BLUE (Abb. 35 links) zeigt einen erheblichen negativen Zusammenhang zwischen Ertrag und Proteingehalt. Der Ertrag je Sorte ist ‚ungeschrumpft' geschätzt. Die Ertragsvarianz der Sorten innerhalb jeweils einer Protein-Note ist sehr viel geringer als die gesamte Ertragsvarianz über alle Sorten. Sorten mit hohem Proteingehalt konnten offensichtlich auch bei deutlich unterdurchschnittlichen Erträgen zugelassen werden (und umgekehrt).

Abb. 35 (rechts) zeigt auch für BLUP den Zusammenhang zwischen Ertrag und Protein, allerdings in der Steilheit deutlich abgeschwächt - also bezüglich der isolierten Betrachtung des Ertrages geschrumpft. Einerseits ist die Ertragsvarianz in jeder einzelnen Protein-APS gegenüber BLUE reduziert - dies ist der aus univariater Sicht ‚normale' Effekt der Schrumpfung. Das kritische Potenzial liegt aber darin, dass z.B. alle Sorten mit einer Protein-APS unter ‚4' *systematisch* im Ertrag nach unten zum Gesamtmittel aller Sorten korrigiert werden, während Sorten mit einer Protein-APS über ‚4' *systematisch* im Ertrag nach oben korrigiert werden. Differenzen und Rangfolgen zwischen Ertrags-Mittelwerten von Sorten unterschiedlicher Protein-APS werden im Zuge der Schrumpfung verzerrt, solange (in diesem Fall) nicht auch der Proteingehalt im Auswertungsmodell berücksichtigt wird.

Die Differenzen zwischen BLUE- und BLUP-Schätzungen sind ergänzend in Abb. 36 (links) dargestellt. Hier zeigt sich besonders anschaulich das Potenzial für eine systematische Verzerrung bei der BLUP-Schätzung, wenn das korrelierte Merkmal Proteingehalt in der Ertragsauswertung unberücksichtigt bleibt. Diese systematische Verzerrung erfolgt hier nur bei der BLUP-Schätzung in unmittelbarem Kontext der Schrumpfungsschätzung. Die Differenzen zwischen BLUP- und BLUE-Schätzungen sind umso größer, je stärker die Sorte im Proteingehalt vom Mittel aller Sorten abweicht - Sorten für spezielle Verwendungen werden also am deutlichsten betroffen sein. Auf die Möglichkeit einer multivariaten Auswertung unter Berücksichtigung mehrerer oder vieler Merkmale wird weiter unten eingegangen.



APS - Ausprägungsstufe nach Bundessortenamt (2009)

Abb. 36: Systematischer Einfluss des Proteingehaltes auf den Unterschied zwischen BLUE- und BLUP-Ertragsschätzung von Weizensorten

In Abb. 36 (rechts) sind des Weiteren die Differenzen BLUE - BLUP in Beziehung zur Schätzgenauigkeit (bei BLUE) dargestellt. Deutlich wird, dass insbesondere bei Sorten mit hohem Fehler (SE_{Serie}, s. 5.3.5), also bei geringem Datenumfang und damit also i.d.R. bei jungen Sorten die Differenzen zwischen BLUP und BLUE-Schätzungen sehr hoch sein können. Dagegen wiesen umfangreich, also i.d.R. langjährig geprüfte Sorten mit geringem Vorhersagefehler nur geringe Differenzen auf. Insofern beeinflusst die Schrumpfungsschätzung durch BLUP im unbalanzierten Fall nicht nur die Differenzen sondern auch die Rangfolgen zwischen Sortenmittelwerten nicht unerheblich.

Ein Risikopotenzial der Nutzung von BLUP besteht also, wenn relevante Korrelationen zu weiteren wertbestimmenden Eigenschaften übersehen bzw. nicht berücksichtigt werden. Am Beispiel der negativen Korrelation zwischen Ertrag und Proteingehalt wurde gezeigt, dass bei der Ertragsschätzung ohne Berücksichtigung des Proteingehaltes durch BLUP proteinreiche Sorten systematisch überschätzt und proteinstarke Sorten systematisch unterschätzt werden. Um als sogenannter E-Weizen eingetragen zu werden und damit die Chance zu haben, auf dem Markt im Höchstpreissegment gehandelt zu werden, ist im Proteingehalt die Ausprägungsstufe „6“ (APS 6) die Mindestanforderung (Bundessortenamt, 2009). Alle diese Sorten weisen hier einen Ertrag unter dem Gesamtmittel von ca. 100 dt/ha auf. Dagegen haben im anderen Extrem sogenannte C-Weizen (sonstige Weizen; i.d.R. Futterweizen) keinerlei Mindestanforderungen im Proteingehalt zu erfüllen, sie liegen ertraglich dafür meist sehr hoch. Abgesehen von der wahrscheinlich begrenzten Möglichkeit Protein-Menge zu bilden ‚durften‘ im Zulassungsverfahren Sorten mit hohem Proteingehalt geringe Erträge haben, ohne ausselektiert zu werden, Sorten mit niedrigem Proteingehalt mussten dies dagegen mit sehr hohen Erträgen kompensieren, um zugelassen zu werden. Dazwischen liegen, sowohl ertraglich als auch preislich, B-Weizen (mindestens APS 2 im Proteingehalt) und A-Weizen (mindestens APS 4 im Proteingehalt).

Dieses hier aufgezeigte Verzerrungspotenzial wird vom Autor dieser Arbeit als schwerwiegend angesehen und geht über die Beurteilung von Piepho und Möhring (2005) hinaus, welche in Teilpopulationen des Weizens hinsichtlich Backqualität

ohne Normalverteilung der Gesamtpopulation kein grundsätzliches Problem für BLUP sehen, sondern nur einen möglichen Verlust an Vorzüglichkeit von BLUP.

In diesem bivariaten Beispiel (Ertrag und Proteingehalt) wäre, sofern keine weiteren Merkmale konkurrierten (unrealistisch!), eine Schrumpfung der Erträge zum Sortenmittel der gleichen Proteineinstufung hin begründet. Dies würde eine Ertragsauswertung mit einer Kovariablen oder einen multivariaten Auswertungsansatz für die gemeinsame Auswertung von Ertrag und Proteingehalt erfordern. Speziell für die Sortenprüfung wurden für derartige Fälle geeignete Auswertungsansätze sowie auch die Umsetzung im gemischten Modell mit SAS® durch Piepho und Möhring (2011) und Piepho et al. (2014) diskutiert. Emrich et al. (2008) beschreiben Auswertungsansätze für einen vergleichbaren Fall, wobei die Selektion auf *Fusariumresistenz* in der Weise erfolgen sollte, dass die Variabilität des den Fusariumbefall beeinflussenden Sortenmerkmals ‚*Termin Ährenschieben*' erhalten bleibt. Sie zeigen, dass ein eine Kovariable enthaltendes statistisches Modell eine geeignete, leicht handhabbare Alternative zu einer bivariaten Analyse darstellt.

Allerdings wäre damit auch nur ein einzelnes weiteres Merkmal berücksichtigt. Prinzipiell sollte aber davon ausgegangen werden, dass auch zu weiteren wertbestimmenden Merkmalen konkurrierende Beziehungen zu erwarten sind. Dabei muss es sich nicht zwangsläufig um Korrelationen aufgrund einer ursächlichen Verknüpfung handeln. Bereits die Erfordernisse der Gesamtbewertung einer Sorte führen dazu, dass alle wertbestimmenden Merkmale sich - mehr oder weniger stark und in gewissen Grenzen - kompensieren können (s.o.). Alle Fragen der Gesamtbewertung (Zulassung, Weiterprüfung oder Empfehlung einer Sorte) werden multivariat entschieden. Insofern konkurrieren die wertbestimmenden Eigenschaften in der Weise miteinander, dass Schwächen in einzelnen Merkmalen durch Stärken in anderen kompensiert werden können.

In der Gesamtheit aller wertbestimmenden Merkmale könnte diese Komplexität nicht realistisch im Auswertungsmodell berücksichtigt werden. Insofern sieht der Autor dieser Arbeit für die Anwendung in der Spätphase der Sortenprüfung - in Deutschland für Landessortenversuche - ein relevantes Verzerrungspotenzial und favorisiert in Übereinstimmung mit Patterson (1997) vorerst univariate Auswertungen mit Betrachtung des Faktors Sorte als fix. Diese Sicht bezieht der Autor vorliegender Arbeit im Speziellen auf die Schätzung der Sorteneffekte. Auf die temporär abweichende Verfahrensweise im Schritt der vorgelagerten Varianzkomponentenschätzung, in welcher auch ‚Sorte' als Voraussetzung für die gewichtete Einbeziehung von Nachbargebieten zufällig gesetzt ist, wurde unter 5.3 eingegangen. Eine multivariate Auswertung mit dem Ziel, eine BLUP-Schätzung für Sorteneffekte zu begründen, wird vor dem Hintergrund der großen Anzahl im amtlichen Prüfsystem zu bewertender und dabei miteinander konkurrierender bzw. sich kompensierender Merkmale für das Gros der Pflanzenarten nicht für praktikabel erachtet.

Auch Meuwissen et a. (2001) beschreiben für die genomische Selektion in bestimmten Situationen Probleme mit BLUP infolge überhöhter Schrumpfung (s. 6.4.1). Mit diesem Hinweis und den vorgestellten Analysen soll an dieser Stelle nochmals darauf aufmerksam gemacht werden, dass bezüglich der Akzeptanz für eine Schrumpfungsschätzung die jeweilige Situation genau analysiert werden sollte und die unbestreitbare mögliche Vorteilswirkung nicht pauschalisiert werden sollte.

Differenzierte Betrachtung der Effekte ‚Sorte' und 'Sorte • Anbaugebiet' in der Modellgleichung:

Ist in der Modellgleichung der Haupteffekt *Sorte* (S) fix gesetzt, so dürfte das oben beschriebene Risiko systematischer Mittelwertverzerrungen durch unberücksichtigte konkurrierende Merkmale nicht bestehen, selbst wenn 'S • AG' in der Modellgleichung formal zufällig gesetzt ist. In diesem Fall würden nur die ‚S × AG' - Interaktionseffekte, nicht aber die ‚S' - Haupteffekte einer Schrumpfung unterzogen. U.U. ließen sich damit die MSEP und damit auch die SE_{Serie} ohne die beschriebenen systematischen Verzerrungspotenziale weiter verringern. Hierbei muss aber betont werden, dass dies ein rein formal-pragmatischer Ansatz zur Modelloptimierung wäre. ‚S • AG' in der Modellgleichung zufällig zu setzen, stünde im Grundsatz nicht in Übereinstimmung mit der Definition *fix* für die beiden Faktoren ‚S' und ‚AG'.

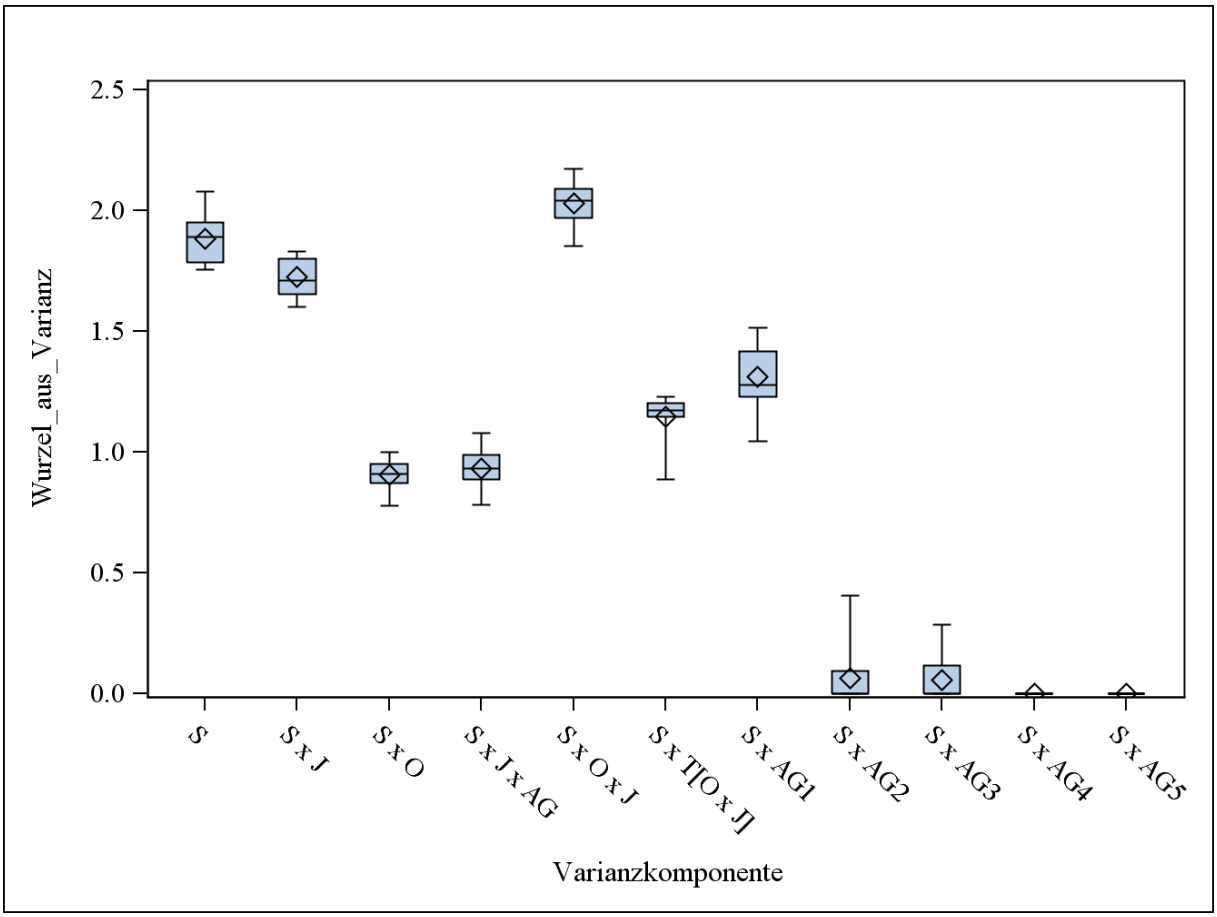
Diese Frage und Möglichkeit zur Modelloptimierung wird hier nicht weiter vertieft, wäre aber für künftige Arbeiten von Interesse.

6.5 Validierungsansätze

6.5.1 Stabilität der Varianzkomponentenschätzung

Um eine Rückinformation zur Stabilität der Varianzkomponentenschätzung bei variierenden einbezogenen Umwelten zu erhalten, wurde für den Kornertrag von Winterweizen untersucht, wie stark die - Schwankungen der geschätzten Varianzen sind, wenn jedes der insgesamt neun Jahre jeweils einmal ausgelassen wird. Der Faktor *Jahr* wurde hierfür zielbewusst gewählt, weil (1) wie unter 6.2.1 gezeigt die diesen Faktor enthaltenden Varianzkomponenten im Modell besonders groß sind und weil (2) die Aufnahme eines kompletten (neuen) bzw. auch die Weglassung eines kompletten (alten) Jahres ein typisches Vorgehen im fortlaufenden Prozess eines Sortenprüfsystems darstellt. Bei der Interpretation muss allerdings beachtet werden, dass jeweils zwei Szenarien in sieben von neun Jahren, also zu ≈ 78 %, übereinstimmen, während eine echte Kreuzvalidierung eine völlige Datensatzteilung beinhalten würde.

Aus Abb. 37 ist zu ersehen, in welchem Maße die geschätzten Varianzkomponenten infolge der Auslassung unterschiedlicher Jahre Schwankungen unterworfen sind. Um dabei alle Varianzkomponenten mit sehr unterschiedlichem Niveau der Varianz-Beträge gemeinsam visuell anschaulich darzustellen, wurden in der Abbildung nicht die Varianzen, sondern deren Quadratwurzelwerte dargestellt.



AG_1 = Anbaugebiet W AG_2 = Anbaugebiet D-Nord AG_3 = Anbaugebiet S
AG_4 = Anbaugebiet MV-Süd AG_5 = Anbaugebiet SW
Boxplots: Definition der Berechnungsformel für die Quantile in SAS® durch Option PCTLDEF=4

Abb. 37: Auswirkung der Auslassung jeweils eines von neun Jahren auf die Varianzkomponenten; Kornertrag Winterweizen

Vor dem Hintergrund der hier gezeigten begrenzten Stabilität der Varianzschätzungen soll darauf hingewiesen werden, dass die empirischen REML-Schätzungen für Varianzen und Kovarianzen quasi als gegeben in die Mixed-Model-Gleichungen eingesetzt werden. Bei den Schätzungen zufälliger Effekte handelt es sich somit um sogenannte eBLUPS.

6.5.2 Auswirkung variierender Varianzkomponentenschätzungen auf die Schätzung der Sorteneffekte

Nachfolgend wird untersucht, in welchem Maße die in 6.5.1 gefundenen Schwankungen der Varianzen zu Abweichungen in der Schätzung von Sorteneffekten im Kernsortiment führen. Datengrundlage ist ein sechsjähriger Zeitraum (hier 2004 bis 2009), wie er für die Mittelwertschätzungen verwendet wird. Die Auswertung erfolgte entsprechend dem Modell 8b mit neun unterschiedlichen Vorgaben für die Varianzkomponenten entsprechend den neun in 6.5.1 definierten Auslassungs-Szenarien.

Im Ergebnis dieser Versuchsserienauswertungen stehen für jedes der neun Szenarien für alle 1118 Sorten empirische gewichtete LS-Mittelwerte (LS Means) zur Verfügung. Die Berechnung der nachfolgend definierten und dargestellten Abwei-

chungen basiert nur auf den 44 Sorten des Kernsortimentes. Eine einzelne Abweichung$_{ij}$ ist dabei die Abweichung des Sorteneffektes einer Sorte$_i$ im Szenarium$_j$ vom mittleren Effekt dieser Sorte über alle Szenarien:

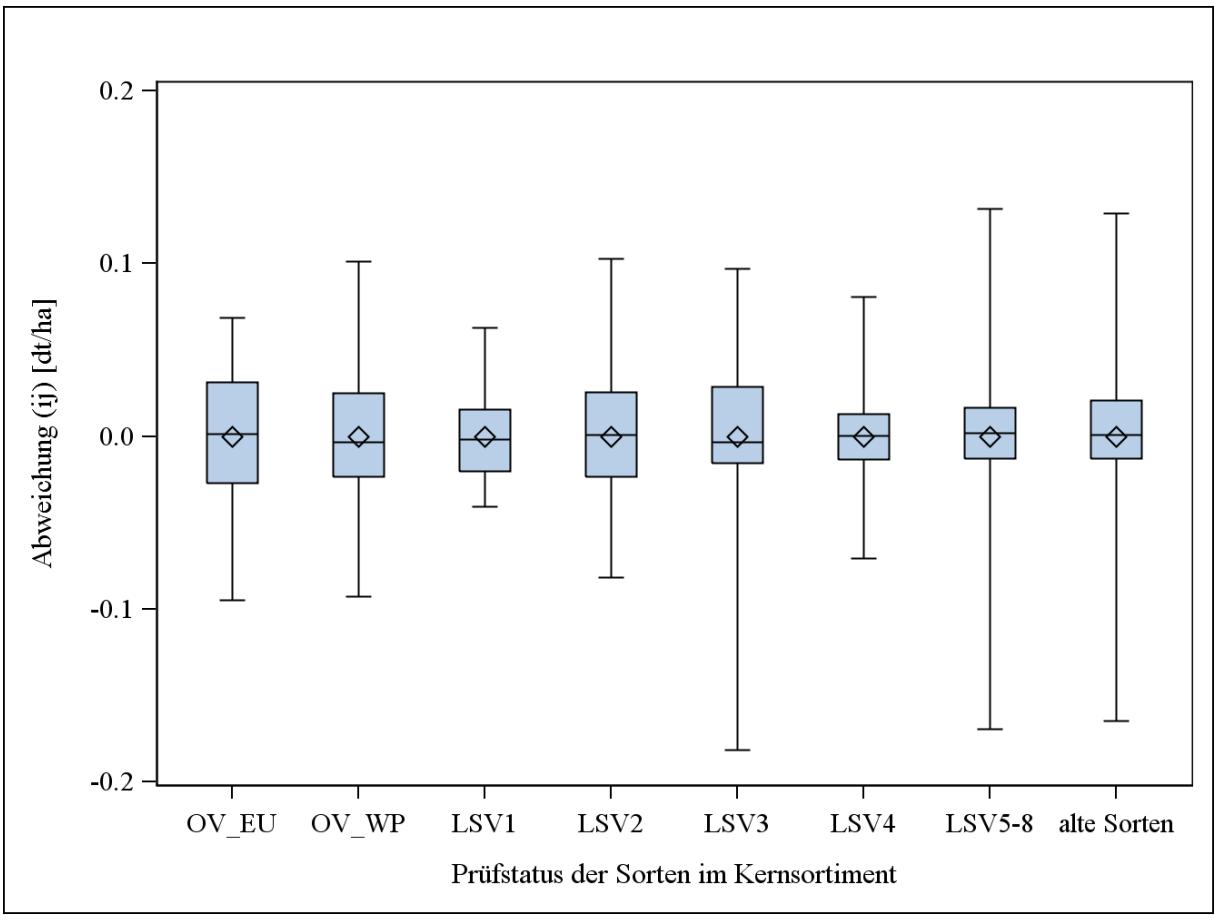
$$Abweichung_{ij} = x_{ij} - \bar{x}_{i.} - \bar{x}_{.j} + \bar{x}_{..}$$

$$x_{ij} = lsmeans(Sorte_i \; im \; Szenarium_j)$$

$$mit \; Sorte \quad i = 1 \ldots 44 \quad (Kernsortiment)$$

$$mit \; Szenarium \; j = 1 \ldots 9 \quad (Auslassung \; jeweils \; eines \; von \; neun \; Jahren)$$

Unabhängig vom Prüfstatus der einzelnen Sorten (Definition s. Tab. 8) lagen alle durch Jahresauslassungen bei der Varianzkomponentenschätzung bedingten absoluten Abweichungen unter 0,2 dt/ha (Abb. 38).



Definition des Prüfstatus‘ s. Tab. 8
für Sorte i=1 … 44 im Szenarium j=1 … 9; Definition des Abweichungen$_{ij}$ über der Abbildung

Abb. 38: Abweichungen geschätzter Sorteneffekte infolge der Auslassung jeweils eines von neun Jahren bei der Varianzkomponentenschätzung, Kornertrag Winterweizen

6.5.3 Güte der Schätzung von Sorteneffekten nach einer sechsjährigen Auswertung als Vorhersage für das Folgejahr

Unabhängig von den in 6.4 diskutierten Vorteilen bzw. Risiken einer Schätzung bzw. Vorhersage von Sorteneffekten durch BLUE versus BLUP ist das faktische Ziel einer mehrjährigen Auswertung die möglichst gute Übereinstimmung mit den im Folgejahr realisierten Sorteneffekten. Die Struktur der Daten, der Besetzung der Sorten im Datensatz bzw. spiegelbildlich die systematische Unbalanciertheit des Datensatzes gibt hierbei im Sortenprüfsystem vor, dass nicht alle beliebigen im Sortiment vorkommenden Sorten einer mehrjährigen Auswertung mit den Sortenergebnissen im Folgejahr verglichen werden sollten. Vielmehr sollte die zielgerichtete Sortimentsbildung in der Weise berücksichtigt werden, dass man sich bei der Beurteilung auf das Kernsortiment beschränkt - dieses wurde in der Versuchs- und Serienplanung a priori als von besonderem Interesse für das Folgejahr definiert.

Nachfolgend wird - beispielhaft am Kornertrag Winterweizen - analysiert, wie gut die in dieser Arbeit methodisch hergeleitete Schätzung von Sorteneffekten nach sechsjähriger Auswertung (2004-2009) mit den im folgenden Versuchsjahr (2010) realisierten Sorteneffekten in diesem Kernsortiment übereinstimmt. Vorteil dieses Ansatzes ist, dass die Daten des Testdatensatzes (Folgejahr) nicht in der Auswertung des sechsjährigen Datensatzes enthalten waren und dass somit eine unabhängige Validierung vorgenommen werden kann.

Eine völlig von der vorgestellten Methode unabhängige Gegenüberstellung mit den einjährigen Ergebnissen aus 2010 ist allerdings nicht möglich, da es kein Modellunabhängiges Ergebnis des Jahres 2010 gibt. Vielmehr erfordert die Bestimmung der Sorteneffekte in 2010 ebenfalls einen vorab zu wählenden Ansatz für das Auswertungsmodell, da außer dem Faktor Jahr und den damit assoziierten Interaktionen alle Einflussfaktoren in der einjährigen Auswertung ebenso anstehen. Gleichfalls besteht die Problematik der Unbalanciertheit, der Bestimmung der Varianzkomponenten und des möglichen Transformationserfordernisses. Zwangsläufig kommt man bei der Auswertung des einjährigen Testdatensatzes (hier 2010) zu ähnlichen Modellansätzen wie in der zu validierenden mehrjährigen Auswertung (2004-2009).

Die Sorteneffekte der mehrjährigen Auswertung entsprechen uneingeschränkt dem in dieser Arbeit verwendetem Szenarium 1 (Sz 1), also der Standardherangehensweise der Hohenheim-Gülzower-Serienauswertung. Die Schätzung der Varianzkomponenten und des Transformationsparameters für die Box-Cox-Transformation basieren auf der hier vorgestellten neunjährigen Auswertung (2001-2009) nach Modell 7a bzw. 7b. Die Schätzung der Sorteneffekte basiert auf der sechsjährigen Auswertung (2004-2009) nach Modell 8b.

Die Bestimmung der Varianzkomponenten für die einjährige Auswertung erfolgt in zwei Varianten, welche unter 5.5 als Sz 9 und Sz 10 definiert wurden. Daraus ergeben sich die für die einjährige Auswertung (2010) verwendeten Varianzkomponenten für beide Szenarien (Tab. 42).

Tab. 42: Für die Schätzung von Sorteneffekten im Folgejahr 2010 verwendete Varianzkomponenten; Kornertrag Winterweizen

Varianzkomponente	**Szenarium 1 bzw. 9 langjährige Schätzwerte**	**Szenarium 10 einjährige Schätzwerte**
S	3,76	3,97
S × J	2,87	-
S × J × AG	0,89	-
S × O	0,80	5,28
S × O × J	4,23	in $(S\times O)_{einjährig}$ enthalten
S × T (O × J)	1,24	0,59
S × AG **(D-Nord)**	0,00	-
S × AG **(MV-Süd)**	0,00	
S × AG **(Ṣ)**	0,00	
S × AG **(ṢẆ)**	0,00	
S × AG **(Ẉ)**	1,55	

Definition der Szenarien unter 5.5

Erwartungsgemäß schlägt sich die große S × O × J - Varianz der langjährigen Auswertung bei einer nur einjährigen Auswertung (Sz 10) in einer (scheinbar) hohen S × O - Varianz nieder. Bei einjähriger Auswertung überschätzt die ausgewiesene S × O - Varianz die tatsächliche S × O - Varianz um den Betrag der (dann unbekannten, aber wirksamen) S × O × J - Varianz: $(S\times O + S\times O\times J)_{mehrjährig} \cong (S\times O)_{einjährig}$

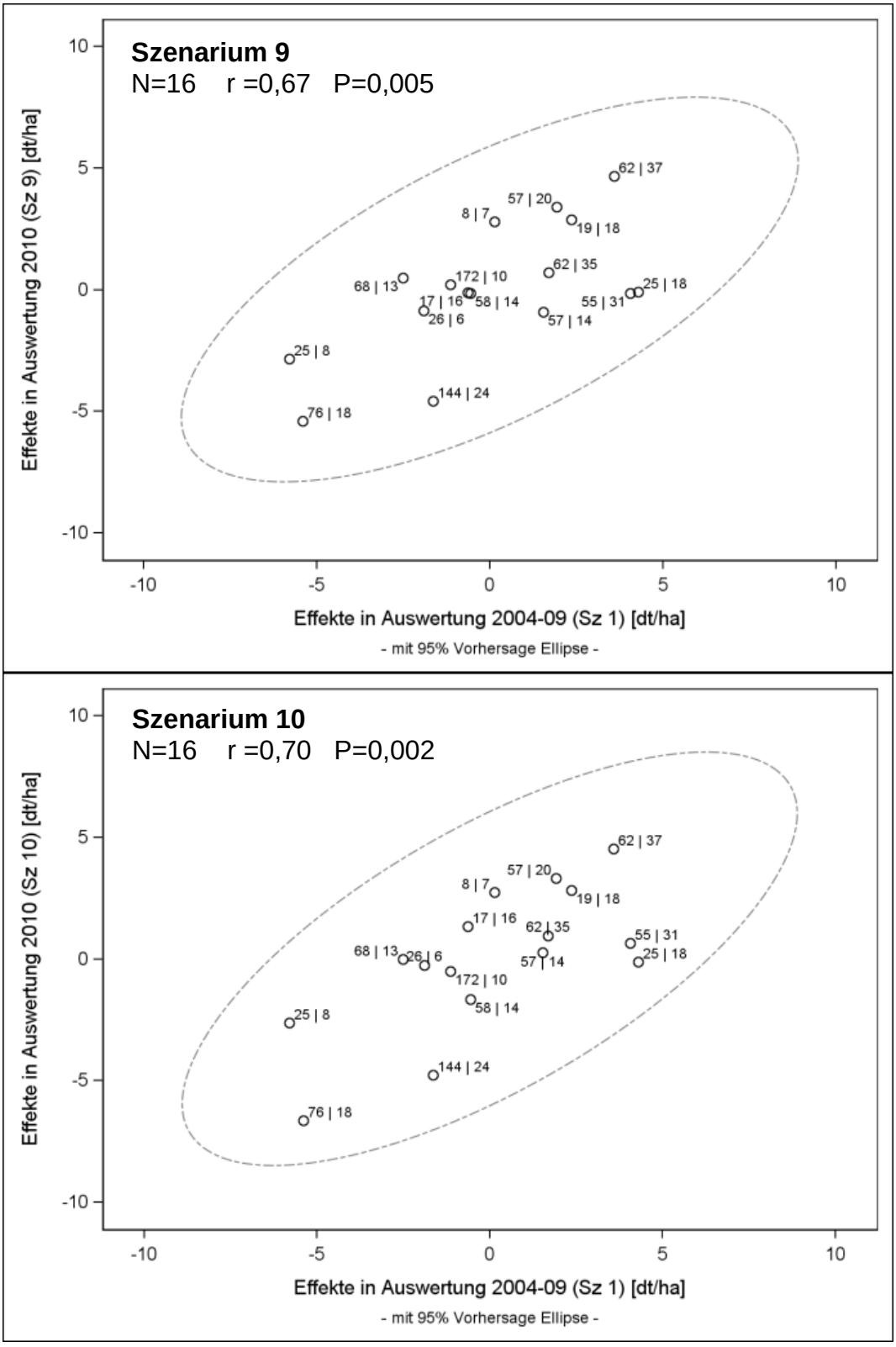
In der Auswertung an einem realen Datensatz kann sich dies nur näherungsweise zeigen, da die Daten nicht übereinstimmen. Setzt man in die o.g. Formel die Schätzwerte lt. Tab. 42 ein, so ergibt sich mit (0,80 + 4,23) ≅ 5.28 eine recht gute Annäherung. Hinzu kommt in diesem konkreten Vergleich, dass alle den Faktor ‚AG' enthaltenden Effekte, welche alle Varianzschätzungen beeinflussen, in der einjährigen Auswertung fehlen.

Dies macht das Risiko deutlich, in einjährigen Serien Effekte irrtümlich als S × O - zu interpretieren, ohne dass diese Effekte tatsächlich (stabil) an Orte gebunden wären.

Von den 44 Kernsorten, die nach Ernte 2009 zur Bewertung anstanden, wurden nur 16 Sorten im Folgejahr 2010 im Zielgebiet noch weitergeprüft. Von dieser Reduktion waren - mehr oder weniger - alle Prüfstadien betroffen. Dies verdeutlicht die bereits bei Silvey (1978) beschriebene rasante Dynamik des Sortenwechsels - hier sogar im Kernsortiment. Diese 16 Sorten werden im Weiteren nicht nach Prüfstatus (Def. s. Tab. 8) differenziert betrachtet.

In Abb. 39 ist für Sz 9 und Sz 10 der Zusammenhang zwischen der sechsjährigen Ertragsschätzung (x) und der Ertragsschätzung im Folgejahr (y) als Streudiagramm mit 95% Vorhersage-Ellipsen dargestellt. Der Unterschied zwischen beiden

Szenarien ist marginal. In beiden Fällen ist die phänotypische Korrelation hochsignifikant, der Korrelationskoeffizient beträgt 0,67 bzw. 0,70.



x | y - Anzahl Ergebnisse für eine Sorte für die zwei Datensätze: 2004-09 | 2010
N - Anzahl Wertepaare r - Pearsonscher Korrelationskoeffizient P - Prob > |r| unter Rho=0

Abb. 39: Phänotypische Korrelation zwischen der sechsjährigen Ertragsschätzung 2004-09 und der Ertragsschätzung im Folgejahr 2010; Kornertrag Winterweizen

6.5.4 Diskussion

Die verwendeten Modelle und geschätzten Parameter wurden in dieser Arbeit überwiegend anhand von Parametern zur Einschätzung der Modellanpassung (AICC oder Log-Likelihood) bzw. auf Basis von Genauigkeitsangaben wie SE_{Serie} (Definition s. 5.3.5) beurteilt. Hinsichtlich einer angemessenen Validierung ist dieses Vorgehen insofern problematisch, dass die Daten zur Parameterschätzung der Varianzen, Transformationsparameter und Sorteneffekte gleichzeitig die Basis für diese Genauigkeitsangaben und für Kriterien der Modellanpassung sind. Eine Validierung setzt aber unabhängige Datensätze voraus. Bei Nutzung eines Kreuzvalidierungsverfahrens würde der Datensatz geteilt: ein Teil für die Schätzung der Modellparameter, der zweite für die Berechnung der Modellfehler. Für die hier vorgestellte Problematik lässt die extreme Unbalanciertheit der Daten und insbesondere auch die systematische Bindung dieser Unbalanciertheit an die Struktur der Faktoren/Faktorstufen (s. Abschnitt 4.5) eine geeignete Datensatzteilung allerdings nicht zu. Die dem Sortenprüfsystem innewohnende und für eine adäquate Auswertung notwendige Struktur der Bindungen / Pfade zwischen Sorten insbesondere über Typen und Jahre, aber auch über Anbaugebiete und Orte hinweg ginge verloren. Auch eine Herangehensweise, bei der (1) jeder Beobachtungswert einmal gestrichen wird, (2) mit den übrigen Werten die Modellparameter und der fehlende Wert geschätzt werden, (3) die Differenz zum beobachteten Wert als vorhergesagtes Residuum gesetzt wird und (4) die Summe der quadrierten vorhergesagten Residuen (PRESS=Predicted Residual Sum of Squares) hinsichtlich der Übereinstimmung mit der Summe der quadrierten geschätzten Residuen beurteilt werden, ist mit derart umfangreichen Daten zeitlich-technisch nicht realisierbar. Allerdings erlauben die vorgenommenen Gesamtbetrachtungen aufgrund der Vielzahl untersuchter Pflanzenart $\times$ Merkmal - Kombinationen trotzdem Verallgemeinerungen.

Stabilität der Varianzkomponentenschätzung

Die Einbeziehung bzw. Weglassung einzelner Jahre hat keinen völlig unerheblichen Einfluss auf die geschätzten Varianzen. Gleichwohl bleiben die Relationen zwischen den Varianzkomponenten über alle Szenarien bei langjährigen Auswertungen weitgehend stabil. Eine erhöhte relative Variabilität / Instabilität zeigen die Anbaugebiet spezifischen Sorte $\times$ Anbaugebiet - Varianzen. Anbaugebietsspezifik von Sorten scheint weniger konstant über die Jahre einzutreten, sondern eher nur in einzelnen Jahr $\times$ Anbaugebiet - Konstellationen hervorzutreten. So könnte man sich vorstellen, dass z.B. auf besonders günstige Bedingungen überproportional stark ansprechende Sorten in Anbaugebieten mit guten Böden nur in denjenigen Jahren diese Eigenschaft zeigen, die zusätzlich auch witterungsbedingt im besonders günstigen Bereich liegen, in anderen Jahren aber keine vom Großraum abweichende Sortenspezifik zeigen. Dies wird hier aber nur als denkbare Erklärung postuliert und nicht weiter vertieft. Eine Verkürzung des Zeitraumes, welcher in die Varianzkomponentenschätzung einfließt, gegenüber dem in vorliegender Arbeit und in der Handhabung in Mecklenburg-Vorpommern verwendeten neunjährigen Zeitraum erscheint insofern für das Konzept der gewichteten Einbeziehung benachbarter Anbaugebiete kritisch. Vielmehr sollte der Zeitraum für diese Parameterschätzungen so lang wie fachlich vertretbar sein.

Auswirkung variierender Varianzkomponentenschätzungen auf die Schätzung der Sorteneffekte:

Die Auslassung jeweils eines von neun Jahren für die Varianzkomponentenschätzung hat nur sehr unerheblichen Einfluss auf die Schätzung der Sorteneffekte der Sorten des Kernsortimentes. Insofern kann die Schätzung von Sorteneffekten in diesem Sortenprüfsystem als sehr robust bezüglich Zeitraum bedingten Schwankungen der geschätzten Varianzkomponenten eingeschätzt werden, sofern diese Schätzung auf einem ausreichend langen Zeitraum beruht.

Güte der Schätzung von Sorteneffekten nach einer sechsjährigen Auswertung als Vorhersage für das Folgejahr:

Die mit dem Auswertungsmodell der Hohenheim-Gülzower-Serienauswertung geschätzten Sorteneffekte nach einer sechsjährigen Ertragsauswertung weisen einen signifikanten Zusammenhang zu den im Folgejahr erzielten Ergebnissen auf. Die phänotypische Korrelation betrug je nach Ansatz für die Varianzkomponentenschätzung in der einjährigen Auswertung $r = 0{,}67$ bzw. $r = 0{,}70$. Diese phänotypische Korrelation ist auch durch Varianzursachen *innerhalb* von Jahren beeinflusst, während die interessantere genetische Korrelation primär ‚nur' von der Sorte $\times$ Jahr - Interaktion beeinträchtigt sein sollte. Auf eine kalkulatorische Bestimmung der genetischen Korrelation wurde verzichtet, weil (1) nur 16 Wertepaare herangezogen werden konnten und (2) die für die Varianzschätzungen genutzten Modelle ohnehin im Zuge der Validierung nicht als ‚richtig' angesehen werden sollten. Vielmehr kann - angesichts der hohen Sorte $\times$ Jahr - Interaktion - bereits die signifikante phänotypische Korrelation als beachtlich angesehen werden und es kann von einer mehr oder weniger deutlich höheren genetischen Korrelation der mehrjährigen Ertragsschätzung zum Folgejahr ausgegangen werden. Insofern unterstützt dieser Validierungsansatz die verwendeten Auswertungsmodelle.

7 Schlussfolgerungen

Modellvoraussetzungen, Datenverhalten und Transformation

Die Box-Cox-Transformation bietet einen sehr flexiblen, situationsbezogenen Ansatz, um konkrete Daten durch ein lineares, homoskedastisches Modell bestmöglich beschreiben zu können. In der Mehrzahl der Fälle führt eine Transformation der Daten mit einem aus den Daten geschätzten Transformationsparameter φ zu einer signifikant besseren Modellanpassung als eine Auswertung der untransformierten Daten. Die Gültigkeit der Modellvoraussetzungen wird damit nicht zwangsläufig gewährleistet.

Eine schiefe Verteilung der Residuen steht häufig nicht ursächlich mit der Verbesserung der Modellanpassung durch Transformation im Zusammenhang.

Typisch sind positive, sehr große Beträge im Exzess - sowohl bei der Auswertung ohne Transformation als auch mit Transformation. Regelmäßige Ursache sind einzelne wenige, extrem auffälligen Einzelwerte. Plausibilitätsprüfungen in allen Phasen der Versuchsbetreuung und der Versuchs- und Serienauswertung haben eine große Bedeutung für die Qualität der Auswertung.

Tendenziell besteht für das Gros der Merkmale ein positiver Zusammenhang zwischen Vorhersagewerten und Residuen. Die Varianzen sind also nicht unabhängig vom Erwartungswert. Die Transformation mit dem geschätzten Transformationsparameter führt in diesem Zusammenhang zur Varianzstabilisierung. Dieser Befund kann die häufige Verbesserung der Modellanpassung durch Transformation zumindest anteilig erklären.

Besonders häufig tritt ein Datenverhalten auf, welches als *intermediär* zwischen einem additiven und einem multiplikativen Erklärungsansatz ausgelegt werden kann - der Transformationsparameter liegt bei der Box-Cox-Transformation dann im Bereich $0 < \varphi < 1$.

Auch ein Datenverhalten, bei dem die Varianz mit steigendem Vorhersagewert abnimmt, ist in Einzelfällen - überwiegend bei Prozentmerkmalen - relevant ($\varphi > 1$). Dies ist i.d.R. dann der Fall und plausibel, wenn die Wertespanne sich dem oberen Grenzbereich des Parameterraums nähert.

Als Ursache für die häufige Verbesserung der Modellanpassung durch Transformation kommen darüber hinaus auch Probleme mit Varianzheterogenitäten in Betracht. Die Detektion ist aufgrund der Komplexität des Auswertungsmodells schwer. Bislang wird für zwei zufällige Effekte in der Modellgleichung, für die Restvarianz, und für die Sorte $\times$ Anbaugebiet - Varianz, Varianzheterogenität berücksichtigt. Künftig sollten Fragen der Varianzinhomogenität für unterschiedliche zufällige Effekte des Modells bei der Fortentwicklung der *Hohenheim-Gülzower-Serienauswertung* Berücksichtigung finden.

Eine Überlegenheit der ‚*folded exponential transformation*' gegenüber der *Box-Cox-Transformation* zeigt sich bei den analysierten Prozentmerkmalen nicht. Bei Merkmalsdaten, welche einen breiten Skalenabschnitt des Parameterraums und dabei insbesondere den Wendepunkt 0,5 (50%) überspannen oder häufig die Grenzwerte ‚0' bzw. ‚1' einschließen, ist eine Überlegenheit der ‚*folded exponential transformation*' eher zu erwarten.

Für die Weiterentwicklung der Hohenheim-Gülzower-Serienauswertung wird vorgeschlagen, ein möglicherweise differenziertes Datenverhalten hinsichtlich der unterschiedlichen zufälligen und fixen Effekte der Modellgleichung in Betracht zu ziehen

und die Untersuchungen zur Einhaltung der Modellvoraussetzungen nicht auf die Residuen zu beschränken. Ein Schwerpunkt hierbei sollte die Problematik der möglichen Varianzheterogenität hinsichtlich einzelner Effekte im Modell sein.

In diesem Zusammenhang sollten auch Fragen der Umweltstabilität von Sorten in die Fortentwicklung der *Hohenheim-Gülzower-Serienauswertung* integriert werden.

Die Regel, die Erfüllung der Modellvoraussetzungen für untransformierte Daten als Ausgangshypothese (H0) anzusehen, erscheint zumindest für große Versuchsserien nicht weiter gerechtfertigt. Vielmehr sollte in jedem Einzelfall die zur besten Modellanpassung führende Transformation gesucht werden, ohne dass eine zu widerlegende Ausgangshypothese für das Datenverhalten zugrunde gelegt wird.

Die Entscheidung zur Transformation sollte aber nicht ausschließlich davon abhängig gemacht werden, ob eine Auswertung transformierter Daten zu einer signifikanten Verbesserung der Modellanpassung gegenüber der Auswertung ohne Transformation ($\varphi = 1$) führt. Vielmehr sollten zusätzlich Diagnosegrafiken für die Entscheidung zur Transformation in den Vordergrund gerückt werden. Führt die Transformation dabei zur Verschlechterung der Übereinstimmung mit einzelnen Modellvoraussetzungen, so sollte die Transformation kritisch hinterfragt und ggf. eher unterlassen werden.

Der Datensatz zur Schätzung des Transformationsparameters und zur Diagnose auf Übereinstimmung mit Modellvoraussetzungen sollte nicht mit dem Datensatz zur Schätzung der Sorteneffekte identisch sein. Hierzu bietet es sich an, einen zeitlichen Versatz beider Datensätze einzuplanen und beim Diagnosedatensatz einen längeren Zeitraum einzubeziehen, als für die Mittelwertschätzung. Eine gewisse Datensatzüberschneidung wird im realen Sortenprüfwesen unvermeidlich sein.

Künftig sollte untersucht werden, ob die Anwendung von GLMMs (generalized linear mixed models) eine geeignete Alternative zur Vorabtransformation der Daten ist. Dies kommt insbesondere bei Merkmalen in Betracht, die Probleme hinsichtlich der Normalverteilung der Residuen oder anderer Zufallseffekte anzeigen.

Varianzkomponenten

Der Gesamtkomplex der Sorte × Umwelt - Varianz hat besonders bei Ertragsmerkmalen eine herausragende Bedeutung.

Die hohe Sorte × Umwelt - Varianz begrenzt die Reproduzierbarkeit von Versuchsergebnissen für den Einzelfall erheblich. Für den Landwirt als Ergebnisnutzer ist allerdings auch eine im *Mittel* von Jahren und Schlägen gute Sortenwahl ausgesprochen vorteilhaft, auch wenn nicht jede *Einzelentscheidung* rückblickend optimal erscheint. Eine gute Reproduzierbarkeit unter dieser mehrjährigen und mehrortigen Sicht erfordert eine Serienauswertung mit entsprechend vielen Versuchen. Die Dimensionierung des Prüfsystems sollte dabei besonders die Präzisionsanforderung an die Ertragseinschätzung zugrunde legen.

Bereits in unmittelbarer Nachbarschaft zeitgleich angelegte, vom selben Versuchsansteller betreute separate Versuche weisen z.T. signifikant unterschiedliche Sorteneffekte auf. Dies ist besonders beim Raps zu beobachten. Dies unterstreicht die sehr eingeschränkte Verallgemeinerungsfähigkeit eines Einzelversuchsergebnisses bzw. von Beobachtungen auf einzelnen Schlägen.

Eine hohe Versuchsanzahl kann nicht durch Intensivierung der Arbeit in Einzelversuchen bzw. an Einzelstandorten (z.B. Erhöhung der Wiederholungsanzahl) ersetzt werden. Viele Versuche sind wichtiger als viele Wiederholungen je Einzelversuch.

Ein Mindestmaß an Wiederholungen bleibt allerdings bedeutsam zur Kontrolle und Bewertung der Plausibilität und Qualität der Einzelversuchs-Daten wie auch ggf. zur Nutzung der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen.

Die Sorte × Jahr × Ort - Varianz macht im Mittel aller Merkmale etwa die Hälfte der gesamten Sorte × *Umwelt* - Varianz aus und ist in den meisten Fällen die größte Varianzkomponente. Die Sorten reagieren an einzelnen Ort × Jahr - Kombinationen also spezifisch auf Umwelt-Ereignisse, die weder für den Ort (in mehrjähriger Betrachtung) noch für das Jahr (in mehrortiger Betrachtung) als typisch angesehen werden können. Dieser Umstand erschwert die beratungsrelevante Interpretation der Sorten-Reaktionen auf Umwelteinflüsse.

Die Sorte × Jahr - Varianz ist überwiegend die zweitgrößte Komponente der Sorte × *Umwelt* - Varianz. Das jahresspezifische Witterungsgeschehen beeinflusst die Sorteneffekte eher großräumig als Anbaugebiet-spezifisch. Die große Interaktion von Sorten mit Jahren zeigt, wie wichtig die Einbeziehung möglichst vieler Jahre für die Schätzgenauigkeit ist. Gerade für junge, beratungsrelevante Sorten ist also die Einbeziehung aller Sortenversuche des integrierten Prüfsystems, speziell die der den LSV zeitlich vorgelagerten WP und EUSV, von hoher Bedeutung. Außer Gründen der Datenbesetzung sollten nur grundlegende Veränderungen im Anbausystem etc. die zeitliche Spanne begrenzen.

Die Sorte × Ort - Varianz *innerhalb* der vordefinierten Anbaugebiete hat - im Gegensatz zu dieser Komponente in großen Räumen - überwiegend geringere Relevanz. Dieses Ergebnis bestätigt den zugrunde gelegten Regionalisierungsansatz, wonach die Homogenität eines Anbaugebietes so hinreichend sein sollte, dass es nur ausnahmsweise bzw. nur geringe Ort-spezifische Sortenreaktionen innerhalb von Anbaugebieten gibt. Zudem bestätigt sich die Notwendigkeit einer Regionalprüfung in Anbaugebieten im Anschluss an die nationale Zulassung. Über den idealen räumlichen Zuschnitt und die Größe von Anbaugebieten im Einzelnen ist damit allerdings noch keine Aussage getroffen.

Dessen ungeachtet behält eine ausreichende Ortsanzahl je Anbaugebiet große Bedeutung, da besonders bei jungen Sorten zu wenige Jahre zur Verfügung stehen und damit nur über eine ausreichende Versuchsanzahl insgesamt ein hinreichend schätzgenaues regionales Ergebnis entstehen kann.

Effekte der ‚Alterung' von Sorten und der Selektion sollten für die Weiterentwicklung der Modelle der Hohenheim-Gülzower-Serienauswertung in Betracht gezogen werden.

Langjährige Auswertung und Einbeziehung vorgelagerter Versuche

Unbalanciertheit ist dem Sortenprüfsystem immanent. So waren z.B. bei der Ertragsauswertung Winterweizen lediglich 2,4 % der orthogonal-vollständigen Sorte × Jahr × Ort - Struktur besetzt. Die Auswertungsmethoden sind darauf einzustellen. Der traditionelle Ansatz, orthogonalen Strukturen künstlich zuzuschneiden, verschenkt zu viel Information.

Der zentrale Nutzen der zugrunde gelegten gemischten Modelle für das Sortenprüfwesen liegt in der Möglichkeit der Auswertung der hochgradig unbalancierten Daten und damit in der Ausschöpfung des vollständigen Datenpools.

Dies ermöglicht die zusammenfassende Auswertung langjähriger Ergebnisse des gesamten integrierten Sortenprüfsystems. Dabei können in der regionalbezogenen

Auswertung der Landessortenversuche auch vorgelagerte Versuche wie Wertprüfungen und EU-Sortenversuche einbezogen werden. Die mehrjährige Datenbasis erhöht sich für alle Sorten dadurch erheblich, insbesondere aber für die jungen besonders beratungsrelevanten Sorten.

Die Einbeziehung vorgelagerter Versuche wie WP und EUSV wirkt sich erheblich auf die Schätzgenauigkeit aus, obgleich die Anzahl Versuche je Jahr und Anbaugebiet bei diesen Versuchstypen noch deutlich niedriger liegt, als nach der Aufnahme in den LSV. Für junge Sorten erhöht sich die Schätzgenauigkeit besonders beträchtlich.

Auch ältere noch interessierende Sorten, die nicht mehr im aktuellen LSV stehen, werden noch mit geringem Schätzfehler bewertet, da durch die langjährige Auswertung viele zurückliegende Versuchsergebnisse einbezogen werden. Letzteres bestätigt die Vorgehensweise in Mecklenburg-Vorpommern, langjährig durchgeprüfte, aber noch interessierende ‚*alte Sorten*' nicht mehr weiter zu prüfen, deren adjustierte Mittelwerte aber weiterhin zur Sortencharakterisierung darzustellen.

Die in Mecklenburg-Vorpommern eingeführte Vorgehensweise der Kombination von langjähriger Auswertung einerseits und Einbeziehung vorgelagerter Versuche anderseits erbringt für Sorten unterschiedlichsten Prüfstatus' erhebliche Synergien.

Bei gleichmäßiger Verteilung der Versuchsergebnisse über Jahre und Orte können folgende pauschalisierte ‚*Faustwerte*' als Orientierung gegeben werden: Weniger als 10 Versuchsergebnisse je Sorte ergeben bestenfalls eine grobe Voreinschätzung von Sorten; bis zu etwa 20 Versuchsergebnissen bringt jedes zusätzliche Ergebnis einen erheblichen Gewinn an Genauigkeit; im Bereich von etwa 20 bis 40 Versuchsergebnissen flacht der Genauigkeitsgewinn rapide ab, ab ca. 40 Versuchsergebnissen bringen weitere Ergebnisse keinen nennenswerten Zugewinn an Genauigkeit mehr - es eröffnen sich in diesem Bereich aber Möglichkeiten zur Bewertung der Ertragsstabilität einzelner Sorten.

Im Vergleich aller untersuchten Szenarien kann festgestellt werden, dass die Kombination von Einbeziehung langjähriger Ergebnisse und den LSV vorgelagerter Versuche die wirksamste der im Zuge der Einführung der *Hohenheim-Gülzower-Serienauswertung* eingeleiteten Maßnahmen ist, um die Genauigkeit zu erhöhen. Insbesondere trifft dies für die Früheinschätzung junger Sorten zu. Gegenüber einer traditionellen dreijährigen reinen LSV-Auswertung werden allein durch diese Komponente der *Hohenheim-Gülzower-Serienauswertung* ein bis zwei Jahre bis zur Sortenempfehlung gewonnen.

Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen

Die Gewichtung der Sortenmittelwerte mittels der Kehrwerte ihrer Fehlervarianzen kann sich im Einzelfall auf die Genauigkeit der Mittelwertschätzung für einzelne Sorten mit sehr schwacher Datenbasis auswirken. Bei den Kernsorten des LSV sind die Effekte aber nahezu irrelevant für die Sortenbewertung.

Aus prinzipiellen Überlegungen heraus erscheint es trotzdem sinnvoll, die Genauigkeit der Mittelwerte im Einzelversuch zu berücksichtigen. Die Nutzung der Gewichtung erleichtert die Entscheidung, auch Ergebnisse aus Versuchen mit Fehlwerten bzw. mit erhöhtem Versuchsfehler (insofern der Versuch ansonsten als fachlich relevant, repräsentativ und ‚*auswertbar*' angesehen wird) einzubeziehen. Für eine routinemäßige Auswertung wird eine Gewichtung vorgeschlagen, um ggf. sehr ungleiche Fehlervarianzen bestmöglich zu berücksichtigen.

Die in den PIAFStat-Verfahren für die Hohenheim-Gülzower-Serienauswertung implementierte Gewichtung sollte künftig methodisch vervollkommnet werden, um eine adäquate Einschrittanalyse möglichst vollständig zu reproduzieren.

Sorte × Anbaugebiet - Spezifik und Gewichtung von Nachbargebieten

Bei den meisten ausgewerteten Merkmalen liegt für mindestens ein einbezogenes Anbaugebiet eine signifikante Sortenspezifik vor. Die standortkundlich vordefinierten Anbaugebiete erweisen sich also in ihren räumlichen Dimensionen im Grundansatz als sinnvoll, wenngleich sie im Einzelfall zu hinterfragen sind.

Nachbargebiete, die zu Zielgebieten in Mecklenburg-Vorpommern westlich ausgerichtet sind, zeigten häufiger Sortenspezifik, also geringere genetische Korrelationen zum Zielgebiet. Entsprechend wird die Gewichtung dieser Gebiete in der Auswertung gedämpft.

Für südliche Nachbargebiete ist mit den aktuellen Modellen weniger Spezifik nachzuweisen. Um die zu erwartende Spezifik auf solchen Standorten mit häufiger variierenden saisonalen Mangelbedingungen besser durch das Modell zu greifen, wird vorgeschlagen, mögliche Varianzheterogenität z.B. für den Effekt Sorte × Jahr in Anbaugebieten im Modell zu untersuchen und ggf. zu berücksichtigen.

In Mecklenburg-Vorpommern ist eine Anbaugebietsunterteilung in zwei räumlich relativ kleine Zielgebiete *D-Nord* und *MV-Süd* beim Raps sinnvoll. Dagegen konnte dies beim Winterweizen nicht gezeigt werden - es sollte also geprüft werden, ob diese Unterteilung zu Gunsten eines größeren Zielgebietes aufgegeben werden kann. Insbesondere scheint hier eine eigenständige Ertragsauswertung für das Zielgebiet *MV-Süd* nicht angemessen zu sein.

In keinem Fall sind die genetischen Korrelationen derart niedrig, dass die Einbeziehung eines standortkundlich relevanten Nachbargebietes prinzipiell ungeeignet wäre. Die Begrenzung des einzubeziehenden Raumes ist demnach eher standortkundlich zu entscheiden als über ggf. sehr geringe genetische Korrelationen.

Bei einigen diskutierten Merkmalen sollte entschieden werden, ob aufgrund der nicht nachgewiesenen Sorte × Anbaugebiet - Interaktion auf die Unterteilung in Anbaugebiete innerhalb des einbezogenen Raumes künftig verzichtet werden kann.

Im Verlauf des Prüfprozesses erfolgt eine zunehmende Eigengewichtung des Zielgebietes. Dies entspricht einer Herangehensweise, bei der die regionale Bewertung der Sorten mit fortschreitender Prüfdauer - beginnend mit einer großräumigen Sicht - zunehmend räumlich feiner skaliert wird. Jede Sorte erfährt vor dem Hintergrund ihrer individuellen Prüfsituation eine Gewichtungs-Konstellation, die zur Minimierung des Vorhersagefehlers für diese Sorte führt. Diese Optimierung berücksichtigt sowohl die regionale Verteilung der Datenbesetzung als auch die regionale Korreliertheit.

Die räumliche Erweiterung der Datenbasis durch die Einbeziehung von Nachbargebieten mit korrelierten Sorteneffekten führt zu einer deutlichen Erhöhung der Genauigkeit der Mittelwertschätzungen.

Die neue Strategie der abgestuft gewichteten Einbeziehung von Nachbargebieten ist auch einer Großraumstrategie in allen Phasen der Prüfung überlegen. Voraussetzung ist allerdings, dass jedes einzelne Anbaugebiet wenigstens so gut besetzt ist, dass die Varianzkomponenten und Korrelationen hinreichend gut schätzbar sind.

Die Ergebnisse geben Hinweise für die Optimierung des Zuschnittes der Ziel- und Nachbargebiete. Die konservative Strategie, das Zielgebiet im Zweifelsfall eher über- als unterzugewichten, wird eher durch ausreichend große, ggf. zusammengelegte

Nachbargebiete erzielt, als durch Verkleinerung und Trennung. Wenn es die genetischen Korrelationen und standortkundlichen Betrachtungen sinnvoll erscheinen lassen, sollten einzelne sich berührende Nachbargebiete eher zusammengefasst werden, so dass der Prüfumfang im Vergleich zum Zielgebiet nicht kleiner, im Zweifelsfall eher etwas größer ausfällt. In diesem Sinne sollten im Ergebnis dieser Arbeit im Nachgang auch die hier zunächst vordefinierten Regionalisierungen punktuell überdacht und angepasst werden.

Für die Weiterentwicklung der Regionalisierung im Rahmen der Hohenheim-Gülzower-Serienauswertung sollten künftig neben rein standortkundlichen Ansätzen auch Versuchsdaten basierte Analysen hinzugezogen werden. Die Hohenheim-Gülzower-Serienauswertung bietet hierfür eine geeignete methodische Grundlage.

Vor diesem Hintergrund soll an dieser Stelle auch davor gewarnt werden, das Versuchsstellennetz weiter zu reduzieren. Der hier diskutierte Methodenkomplex hängt unmittelbar von der Möglichkeit und Güte der Parameterschätzungen ab, somit insbesondere neben der Langjährigkeit auch von einem repräsentativen und ausreichend großen Versuchsstellennetz in allen Anbaugebieten.

Von allen eingeleiteten Neuerungen hat die abgestuft gewichtete Einbeziehung von korrelierten Nachbargebieten nach dem Kombinationseffekt ‚langjährig + Einbeziehung vorgelagerter Versuche' die zweitgrößte Bedeutung für die Schätzgenauigkeit. Der Zeitgewinn bis zur Sortenempfehlung beträgt allein durch diese Maßnahme mindestens ein Jahr.

Der Prüfumfang je Zielgebiet ist in Mecklenburg-Vorpommern derzeit bereits zu gering, um ‚*aus eigener Kraft*', also nur mit Daten des Zielgebietes, sehr frühzeitig genaue Sorteneinschätzungen zu erzielen - ein Strategiewechsel zur Effizienzsteigerung stand insofern an.

Faktor ‚Sorte' fix oder zufällig?

Besonders in der Phase der züchterischen Selektion auf ein einzelnes Merkmal können Vorzüge von BLUP (Sorte zufällig) gegenüber BLUE (Sorte fix) genutzt werden.

Im Verlaufe des Prozesses von züchterischer Selektion bis zur amtlichen Regionalprüfung von zugelassenen Sorten mit landeskulturellem Wert verschieben sich die Aufgaben und damit die Erwartungen an die Auswertung. In keiner Phase dieses Prozesses ist die Definition des Faktors ‚Sorte' allerdings völlig eindeutig. In der Phase der amtlichen Sortenprüfung zugelassener Sorten wird die Definition ‚zufällig' für den Faktor Sorte allerdings zunehmend problematischer.

Die jeweilige Situation sollte genau analysiert und die unbestreitbar mögliche Vorteilswirkung von BLUP nicht pauschalisiert werden.

Im amtlichen Sortenprüfsystem sollte der Faktor ‚Sorte' nach Auffassung des Autors dieser Arbeit in univariaten Serienauswertungen als fix betrachtet werden.

Wenn relevante Korrelationen eines Auswertungsmerkmals zu weiteren wertbestimmenden Eigenschaften übersehen und nicht im Auswertungsmodell berücksichtigt werden, besteht ein Verzerrungspotenzial durch BLUP. Kann sichergestellt werden, dass korrelierte wertbestimmende Merkmale vollständig und adäquat im Auswertungsmodell berücksichtigt wurden, so können multivariate Auswertungsansätze die Definition ‚zufällig' für den Faktor Sorte ermöglichen. Allerdings wird eine multivariate Auswertung unter Einbeziehung von vielen Merkmalen für das amtliche Sortenprüfsystem als überwiegend nicht praktikabel erachtet.

Würden die Haupteffekte ‚Sorte' ungeschrumpft geschätzt (Sorte fix), so bestünden gegenüber einer Schrumpfungsschätzung der Sorte × Anbaugebiet - Interaktionseffekte keine formalen Bedenken. Von Interesse wären weiterführende Untersuchungen, ob dieser Ansatz zur Minimierung des MSEP beitragen könnte, ohne dass systematische Verzerrungen der Sorteneffekte riskiert werden.

Validierung

Bei langjährigen Auswertungen sind die Relationen zwischen den Varianzkomponenten über alle Szenarien weitgehend stabil und robust hinsichtlich der Hinzunahme bzw. Weglassung einzelner Jahre, wie dies bei einer jährlichen oder anderweitig periodischen Aktualisierung der Datensätze zur Parameterschätzung zu erwarten ist.

Der einbezogene Zeitraum für die Schätzung eines Transformationsparameters und der Varianzkomponenten sollte so lang wie fachlich vertretbar sein. Eine nicht zu große Überlappung mit dem Datensatz zur Mittelwertschätzung ist anzustreben.

Die Mittelwertschätzung aus sechsjährigen Daten ist sehr robust gegenüber einer Variation des Zeitraums für die Varianzschätzungen.

Die Schätzung von Sorteneffekten mit dem Methodenkomplex der Hohenheim-Gülzower-Serienauswertung auf Basis eines sechsjährigen Datensatzes steht in hochsignifikantem Zusammenhang zu den Sorteneffekten in einem unabhängigen Folgejahr. Hohe Sorte × Jahr - Interaktionen beeinträchtigen allerdings die Reproduzierbarkeit für ein Einzeljahr in relevantem Maße. Von einer hohen Präzision der Schätzung langjähriger Mittelwerte profitieren die Landwirte in vollem Maße also erst in mehrjähriger Betrachtung.

Zusammenfassung

Im Bundesland Mecklenburg-Vorpommern wurde die sogenannte *Hohenheim - Gülzower - Serienauswertung* (**HGS**) als methodisches Instrument im Sortenversuchswesen eingeführt. Eines der Kernziele ist die frühe und präzise Schätzung von Sorteneffekten als eine Grundlage für Empfehlungen für Landwirte. Die vorliegende Dissertation untersucht erstmals die praktische Anwendung der HGS, erarbeitet Verallgemeinerungen und bewertet die Wirksamkeit der wesentlichsten Optionen dieser Methode.

Alle Analysen mit der HGS erfolgen unter Nutzung der MIXED-Prozedur des SAS-Software-Systems mit dem REML-Algorithmus. Die gewählten Modellgleichungen sind auf die Besonderheiten eines Sortenprüfsystems ausgerichtet. Außer dem zentralen Faktor *Sorte* (S) werden insbesondere die Faktoren *Jahr* (J), *Ort* (O), *Anbaugebiet* (AG) und *Versuchstyp* (T) berücksichtigt. Der Faktor O ist in AG stratifiziert, während T innerhalb von $J \times O$ - Kombinationen stratifiziert ist und eine gemeinsame Auswertung von unterschiedlichen Stadien des Prüfsystems ermöglicht, insbesondere die kombinierte Auswertung von regionalisierten Prüfungen nach der Zulassung (*Landessortenversuche*) mit den deutschen *Wertprüfungen* im Sortenzulassungsverfahren in Zuständigkeit des Bundessortenamtes. In die Auswertung wurden insgesamt 36 stetige Merkmale über 13 landwirtschaftliche Pflanzenarten einbezogen. Die Untersuchungen ergaben folgende Ergebnisse.

In der Mehrzahl der Fälle führte eine *Datentransformation* zu einer signifikant besseren Modellanpassung als eine Auswertung untransformierter Daten. Die Gültigkeit eines additiven Modells für untransformierte Daten ist für Versuchsserien offensichtlich nicht der Regelfall. Ohne Transformation zeigten sich bei vielen Merkmalen positive Korrelationen zwischen Vorhersagewerten und Residuen. In Übereinstimmung zu dieser Auffälligkeit zeigte sich häufig ein Datenverhalten, welches als intermediär zwischen einem additiven und einem multiplikativen Verhalten interpretiert werden könnte. Für den *Exzess* wurden häufig sehr große positive Werte gefunden, wenn keine Transformation erfolgt. Dieses Problem konnte durch Transformation nicht substantiell abgebaut werden. Mit der Anwendung einer Datentransformation sollte sorgsam umgegangen werden. Außer Kriterien für die Güte der Modellanpassung sollten auch Diagnosegrafiken genutzt werden, um den Einfluss einer Transformation auf die Einhaltung der Modellvoraussetzungen zu beurteilen. Bei Abweichungen von der Normalverteilung sollte künftig auch untersucht werden, ob die Anwendung von GLMMs (generalized linear mixed models) eine geeignete Alternative zur Transformation ist.

Als indirekte Quelle für eine unbefriedigende Modellanpassung kommt auch *Varianzheterogenität* in Betracht. Die Detektion der von Varianzheterogenität betroffenen Effekte der Modellgleichung ist aufgrund der Komplexität des Modells schwierig. Derzeit wird im Modell Varianzheterogenität für den Interaktionseffekt $S \times AG$ und für die Fehlervarianzen der Mittelwerte aus Einzelversuchen zugelassen. Es sollte hinterfragt werden, ob Varianzheterogenität bezüglich weiterer Effekte der Modellgleichung zu berücksichtigen ist. Zu diesem grundsätzlichen Problem besteht weiterer Untersuchungsbedarf an Daten aus dem Sortenprüfsystem.

Der Komplex der Sorte × Umwelt - Variabilität in Nordostdeutschland wurde mittels *Varianzkomponentenschätzung* bewertet. Dies ergab, unter anderem, in einer stabilen Weise folgende Rangfolge mit abnehmenden Varianzen:

$Varianz(S \times J \times O) > Varianz(S \times J) > Varianz(S \times O)$. Diese Relationen unterstreichen die Bedeutung langjähriger Auswertungen unter Einbeziehung aller Versuchstypen

und zeigen, dass die vordefinierten Anbaugebiete in sich hinreichend homogen sind. Die genetischen Korrelationen der Sorteneffekte zwischen nordostdeutschen Anbaugebieten sind überwiegend sehr groß. Trotzdem zeigte sich bei vielen Merkmalen für einzelne Anbaugebiete eine signifikante Sortenspezifik.

Die Kombination von Langzeitanalyse und Zusammenführung von Wertprüfungen und Landessortenversuchen zeigte sich als die wirksamste Option der HGS. Voraussetzung für die Nutzung dieser Synergie ist die Berücksichtigung der damit verbundenen *Unbalanciertheit* durch die Modellwahl (beim Merkmal ‚Kornertrag von Winterweizen' waren lediglich 2,4 % der Zellen in der orthogonalen $S \times J \times 0$ - Struktur besetzt). Ohne diesen Ansatz wäre eine hinreichend gute Sorteneinschätzung aufgrund geringer Versuchskapazitäten in Mecklenburg-Vorpommern kaum noch möglich.

Auch die *Einbeziehung von Nachbargebieten* zur Schätzung von Sorteneffekten in einem Zielgebiet führt zu einer deutlichen Erhöhung der Genauigkeit der Schätzungen. Dabei bekommen die Gebiete in Abhängigkeit von den genetischen Korrelationen sowie vom Prüfumfang in den Gebieten ein differenziertes Gewicht. Diese Strategie der abgestuften Gewichte ist auch einer Bildung von größeren Räumen ohne Wichtungsabstufung überlegen. Der Zeitgewinn bis zu einer fundierten Sortenempfehlung beträgt durch diese Maßnahme mindestens ein Jahr.

Eine Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen ist eine weitere optionale Möglichkeit der HGS. Dies kann im Einzelfall nützlich sein. Der Einfluss auf die Schätzung der Sorteneffekte und deren Genauigkeit war überwiegend aber eher marginal. Die Gewichtung wird vom Autor vorliegender Dissertation für Routineanwendungen trotzdem empfohlen.

Mit dem Methodenkomplex der HGS auf Basis eines sechsjährigen Datensatzes geschätzte Sorteneffekte waren phänotypisch signifikant korreliert zu den Sorteneffekten eines unabhängigen Folgejahres. Diese Situation entspricht der typischen Herausforderung an die Sortenberatung für Landwirte.

Die Annahme ‚fix' versus ‚zufällig' für den Faktor Sorte (S) wird für die Pflanzenzüchtung und für die amtliche Sortenprüfung differenziert diskutiert. Die Nutzung von BLUP (S ist als zufällig gesetzt) gegenüber BLUE (S ist als fix gesetzt) hat potenzielle Vorteile, z.B. bei Betrachtung des Selektionsgewinnes in einem Merkmal. Es werden in dieser Dissertation aber auch Probleme und Risiken bei der Nutzung von BLUP in der Sortenprüfung aufgezeigt.

Vorzüge der HGS hängen maßgeblich von der hinreichend präzisen Schätzung der Sorte × Umwelt - Varianzen ab. Neben der Einbeziehung vieler Jahre ist somit auch die Erhaltung eines repräsentativen Netzes von Versuchsstandorten in allen Anbaugebieten essenziell.

Summary

The so-called '*Hohenheim - Gülzower - Serienauswertung'* (**HGS**) was introduced to the variety testing system of the German federal state 'Mecklenburg-Vorpommern' as a methodical instrument to analyse series of trials. One of its central aims is the early and precise estimation of variety effects as a basis for recommendations to the farmers. This thesis investigates - for the first time - the practical use of HGS, draws up some generalizations, and evaluates the efficiency of key options of this method.

All analyses based on the HGS are done using the MIXED procedure of the SAS-Software-System with the REML-algorithm. The chosen model equations are aligned to the special features of a variety trial system. Apart from the main factor *Variety* (V), the factors *Year* (Y), *Location* (L), *Growing-Region* (R), and *Type of Trial* (T) are included. The factor L is nested within R, whereas T is nested within $Y \times L$ - combinations and allows a combined analysis of different stages of the testing system, especially to combine regional *post-registration trials* in various federal states of Germany with *pre-registration* national *trials* under the responsibility of the German federal authority 'Bundessortenamt'. The analysis comprises 36 continuously scaled traits from 13 agricultural crops. The investigations revealed the following results.

In most cases a *transformation* of data resulted in a significantly better model fit, compared with the analysis of non-transformed data. The validity of an additive model for untransformed data is for series of trials, therefore, not the norm. Without a transformation there were positive correlations between predicted values and residuals for many traits. In accordance with this phenomenon I frequently found a data behaviour that might be interpreted as intermediate between an additive and a multiplicative behaviour. Values of *kurtosis* were frequently very high and positive, if no transformation was carried out. This problem couldn't be substantially reduced by transformation. A transformation of the data should be applied with the greatest care. In addition to criteria for the quality of model fitting, diagnostic graphics should be used as well in order to assess the influence of a transformation with respect to compliance with the model assumptions. In case of deviations from the normal distribution, it should be investigated in the future whether the application of GLMMs (generalized linear mixed models) is a suitable alternative to the transformation.

As an indirect source of an unsatisfactory fit of models *variance heterogeneity* comes into account. The detection of the effects of the model equation that are affected by variance heterogeneity is difficult because of the complexity of the model. Currently heterogeneity of variance is permitted for the interaction effect $V \times R$ and for the error variances of the mean values of varieties of individual experiments in the model. It should be questioned whether variance heterogeneity with respect to other effects of the model equation must be considered. Concerning this fundamental problem further studies on variety testing data are necessary.

The complex of variety × environmental variability within northeastern Germany was evaluated by *estimating the components of variance*. In a robust manner this resulted, for example, in the following order of decreasing variances: $Variance(V \times Y \times L) > Variance(V \times Y) > Variance(V \times L)$. These relations emphasize the importance of long-term analyses involving all types of trials (T) and point out that the pre-defined regions are sufficiently homogeneous within themselves. Genetic correlations of the variety effects between regions (R) in northeastern Germany are predominantly considerable. Nevertheless, many traits showed a significant specificity of variety effects in distinct regions.

The combination of long-term analysis and merging of data from pre- and post-registration variety trials appeared to be the most effective of all options of the HGS. A precondition for using this synergy is the consideration of the associated *unbalancedness* due to the choice of model (the trait ‘grain yield of winter-wheat’ showed that only 2.4% of the cells in the orthogonal $S \times J \times O$ - structure were occupied). Without this approach, a sufficiently appropriate variety evaluation would hardly be possible due to low test capacities in ‘Mecklenburg-Vorpommern’.

Furthermore, the *inclusion of neighboring regions* to estimate variety effects in a target region leads to a considerable increase of accuracy of the estimates. Depending on the genetic correlations and on the number of trials per region, each region receives its own specific weight. This strategy of graded weighting is also superior to the formation of larger areas without weighting gradation. Using this method, the gain of time until a viable recommendation of varieties can be given amounts at least one year.

A weighting of the mean values of varieties of individual experiments using the reciprocals of their error variances is a further optional possibility of the HGS. This option may be useful in individual cases. The effect on the estimation of the performance of varieties and their accuracy, however, was mostly rather marginal. Nevertheless, the author of the present thesis recommends this weighting for routine applications.

The estimation of variety effects with HGS, based on a six-year dataset, was significantly correlated phenotypically to the effects of an independent subsequent year. This situation corresponds to the typical challenge of advising the farmers regarding the choice of varieties.

The assumption ‘fixed' versus ‘random' regarding the factor variety (V) is discussed in a differentiated way for plant breeding on the one hand and for the official variety testing system on the other. The use of BLUP (the effect V is fitted as random) versus BLUE (the effect V is fitted as fixed) has potential benefits, e.g. when considering the selection gain in a trait. On the other hand, problems and risks are identified in this thesis when using BLUP in variety testing.

Advantages of the HGS crucially depend on a sufficiently precise estimation of the variety × environmental variances. Therefore, apart from the inclusion of many years, the preservation of a representative network of experimental sites in all growing regions is essential.

Abkürzungsverzeichnis

AG	Anbaugebiet (s. ZG oder NG)
AICC	Akaike's Corrected Information Criterion; (Kriterium für die Güte der Modellanpassung)
ANOVA	analysis of variance
APS	Ausprägungsstufe [1 - 9]
b	geschätzter Regressionskoeffizient (Anstieg) in einfacher linearer Regression
B%	relatives Bestimmtheitsmaß (0% bis 100%-Skala), hier in einfacher linearer Regression; hier ausschließlich als adjustiertes (Freiheitsgrad bezogenes, korrigiertes) Bestimmtheitsmaß ausgewiesen
BKR	Boden-Klima-Raum
BLUE	best linear unbiased estimation
BLUP	best linear unbiased prediction
bzw.	beziehungsweise
Def.	Definition
eBLUP	empirical best linear unbiased predictor
EUSV	EU-Sortenversuch
ggf.	gegebenenfalls
Gr	Faktor *Sortengruppe*
GLMM	generalized linear mixed model
i.d.R.	in der Regel
IT	Informationstechnologie
J	Faktor *Jahr*
L	Land (Bundesland)
LDS	Länderdienststelle, für das Sortenwesen in einem Bundesland zuständig
LFA	Landesforschungsanstalt für Landwirtschaft und Fischerei MV
LS Means	least squares means (kleinste Quadrat Mittelwerte)
LSV	Landessortenversuch
lt.	laut (entsprechend)
MAR	missing at random
m.E.	meines Erachtens
ML	maximum likelihood

MQ	mittlere Abweichungsquadrate
MSEP	mean squared error of prediction (mittlere Fehlervarianz einer Vorhersage, hier einer Differenz von Sortenmittelwerten)
MV	Mecklenburg-Vorpommern
NG	Nachbargebiet
O	Faktor *Ort*
o.ä.	oder ähnlich
o.g.	(weiter) oben genannt
OV	Orientierungsversuch
PIAF	Planungs-, Informations- und Auswertungssystem für das Feldversuchswesen
PIAFStat	Auswertungssoftware im PIAF-System
r	Pearsonscher Korrelationskoeffizient
r_g	genetische Korrelation (hier die der Sorteneffekte zwischen Anbaugebieten)
REML	restricted (residual) maximum likelihood
S	Faktor *Sorte*
SE	standard error (hier Standardfehler eines *Mittelwertes*)
SE_{Serie}	Standardfehler eines Sorten-Mittelwertes aus der Versuchsserienauswertung als Funktion des MSEP (Definition s. 5.3.5)
$SE_{Versuch}$	Standardfehler eines Sortenmittelwertes aus einem Einzelversuch für die Gewichtung in der Versuchsserienauswertung
$SE_{Versuch}$%	Variationskoeffizient eines Mittelwertes (!) aus einem Einzelversuch
s.	siehe
s.o.	siehe (weiter) oben
s.u.	siehe (weiter) unten
Ṣ	Symbol für südlich zu MV ausgerichtete AG
ṢẆ	Symbol für südwestlich zu MV ausgerichtete AG
Sz	Szenarium
T	Faktor *Versuchstyp*
u.a.	unter Anderem
u.ä.	und ähnliche etc.
u.U.	unter Umständen
u.v.m.	und vieles mehr
vgl.	vergleiche

VGL	Vergleichssorte
VRS	Verrechnungssorte
WP	Wertprüfung
W̥	Symbol für westlich zu MV ausgerichtete AG
z.B.	zum Beispiel
ZG	Zielgebiet
z.T.	zum Teil

Literaturverzeichnis

Anmerkung: Englischsprachige Artikel wurden bezüglich Groß-/ Kleinschreibung von Substantiven und Eigennamen entsprechend der Originalveröffentlichung aufgenommen - die Handhabung ist daher nicht einheitlich.

Atlin, G.N., R.J. Baker, K.B. McRae & X. Lu, 2000: Selection response in subdivided target regions. Crop Science 40: 7-13.

Autorenkollektiv, 1987: Einführung in die Methodik des Feldversuchs. VEB Deutscher Landwirtschaftsverlag, Berlin.

Autorenkollektiv, 1987b: Biometrisches Wörterbuch. VEB Deutscher Landwirtschaftsverlag, Berlin.

Agrarministerkonferenz, 7.10.2004: Protokoll TOP 4.13: Neuordnung des Sortenversuchswesens. Burg Warberg.

Bätz, G., 1984: Empfehlungen zur erweiterten Auswertung von Versuchsserien, insbesondere unter Berücksichtigung der Prüfglied/Umwelt-Wechselwirkung. Feldversuchswesen, Berlin 1: 20-72.

Bätz, G., R. Richter & J. Held, 1985: Zur Auswertung von Feldversuchen mit unzureichender Präzision durch kritische Wertung der Zufallseffekte. Archiv für Züchtungsforschung 15, Berlin.

Bartlett, M.S., 1947: The use of transformations. Biometrics 3: 39-52.

Basford, K.E., W.T. Federer & I.H. DeLacy, 2004: Mixed model formulations for multi-environment trials. Agron. Journal 96: 143-147.

Buhtz, E. & Bätz, G., 1984: Zur Repräsentanz von Feldversuchen. Feldversuchswesen, Berlin 1: 98-106.

Becker, H.C., 1981: Correlations among sme statistical measures of phenotypic stability. Quantitative genetics and breeding methods. INRA, Versailles; France: 307-314.

Box, G.E. P. & C.R. Cox, 1964: An analysis of transformations (with discussion). J. R. Statist. Soc. B. 26: 211-246.

Bundessortenamt, 2009: Beschreibende Sortenliste 2009 Getreide, Mais, Ölfrüchte, Leguminosen (großkörnig), Hackfrüchte (außer Kartoffeln). Hannover.

Bundessortenamt, 2014: Beschreibende Sortenliste 2014 Getreide, Mais, Öl- und Faserpflanzen, Leguminosen, Rüben, Zwischenfrüchte. Hannover.

Burnham, K.P. & D.R. Anderson, 1998: Model selection and inference. A practical information-theoretic approach. Springer, New York.

Cochran, W.E. & G.M. Cox, 1957: Experimental designs. Wiley. New York.

Cullis, B.R., A. Smith, C. Hunt & A. Gilmour, 2000: An examination of the efficiency of Australian crop variety evaluation programmes. Journal of Agricultural Science 135: 213-222.

Cullis, B.R., P. Jefferson, R. Thompson & A.S. Smith, 2014: Factor analytic and reduced animal models for the investigation of additive genotype-by-environment interaction in outcrossing plant species with application to a *Pinus radiata* breeding programme. Theor. Appl. Genet. 127: 2193-2210.

Curnow, R.N., 1988: The use of correlated information on treatment effects when selecting the best treatment. Biometrika 75: 287-293.

Denis, J., H.P. Piepho & F.A. van Eeuwijk, 1997: Modelling expectation and variance for genotype by environment data. Heredity 79: 162-171.

Emrich, K., F. Wilde, T. Miedaner & H.P. Piepho, 2008: REML approach for adjusting the fusarium head blight rating to a phenological date in inoculated selection experiments of wheat. Theor. Appl. Genet. 117: 65-73.

Fernando, R.L. & M. Grossmann, 1989: Marker-assisted selection using best linear unbiased prediction. Genet. Sel. Evol. 21: 246-477.

Fisher, R.A. & W.A. Mackenzie, 1923: Studies in crop variation: II. The manuarial response of different potato varieties. Journal of Agricultural Science 13: 311-320.

Forkman, J., 2013: The use of a reference variety for comparisons in incomplete series of crop variety trials. Journal of Applied Statistics 40 (12): 2681-2698.

Franko, J., H.-J. Möbius, G. Bätz & K. Stegemann, 1982: Einsatzmöglichkeiten des EDV-Projektes „Datenspeicher Versuchsergebnisse-Pflanzenproduktion ESER“ (DAVEP-ESER) für die Stammes- und Sortenprüfung. Archiv Züchtungsforschung 12: 43-55.

Frensham, A., B. Cullis & A. Verbyla, 1997: Genotype by environment variance heterogeneity in a two-stage analysis. Biometrics 53: 1373–1383.

Gilmour, A.R., R. Thompson & B.R. Cullis, 1995: Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51: 1440-1450.

Gilmour, A.R., B.R. Cullis, A.B. Frensham & R. Thompson, 1998: (Co)variance structures for linear models in the analysis of plant improvement trials. COMSTAT 98 Proceedings in Computational Statistics. Physica Verlag, Heidelberg.

Graf, R., V. Michel, D. Roßberg & R. Neukampf, 2009: Definition pflanzenartspezifischer Anbaugebiete für ein regionalisiertes Versuchswesen im Pflanzenbau. Journal für Kulturpflanzen 61(7): 247-253.

Gurka, M.J., L.J. Edwards, K.E. Muller & L. L. Kupper, 2006: Extending the box-cox transformation to the linear mixed model. J. R. Statist. Soc. A 169: 273-288.

Hamblin, J, H.M. Fisher & H. J. Ridings, 1980: The choice of locality for plant breeding when selecting for high yield and general adaption. Euphytica 29: 161-168.

Heffner, E.L., M.E. Sorrels, J. & Jannink, 2009: Genomic selection for crop improvement. Crop Science 49: 1-12.

Henderson, C.R., 1984: Applications of linear models in animal breeding. University of Guelph, Guelph, Ontario, Canada.

Hu, X. & J. Spilke, 2011: Variance-covariance structure and its influence on variety assessment in regional crop trials. Field Crops Research 120: 1-8.

Isaaks, H. & R.M. Srivastava, 1989: Applied geostatistics. Oxford University Press, New York, Oxford.

John, J.A. & E.R. Williams, 1995: Cyclic and computer generated designs. Chapman and Hall, London.

Johnson, N.L., S. Kotz, & A.W. Kemp, 1993: Univariate discrete distributions. Wiley, New York.

Khattree, R. & N.N. Dayanand, 2000: Multivariate data reduction and discrimination with SAS® software. Cary, NC: SAS Institute Inc. 2000: 347-442.

Kelly, A.M., A.B. Smith, J.A. Eccleston & B.R. Cullis, 2007: The accuracy of varietal selection using factor analytic models for multi-environment plant breeding trials. Crop Science 47: 1063-1070

Kempton, R.A., 1984: The use of biplots in interpreting variety by environment interaction. Journal of Agricultural Science 103: 123-135.

Kienzl, H., 1974: Biometrische Charakterisierung von Standorteinflüssen auf das Sortenversuchswesen; dargestellt an den Bayerischen Landessortenversuchen bei Winterweizen und Sommergerste in den Jahren 1963-1972. TU München-Weihenstephan, Dissertation.

Kleinknecht, K., F. Laidig, H.P. Piepho & J. Möhring, 2011: Best linear unbiased prediction (BLUP): is it beneficial in official variety performance trials? Biuletyn Oceny Odmian 33: 21-33.

Kleinknecht, K., J. Möhring, K.P. Singh, P.H. Zaidi, G.N. Atlin & H.P. Piepho, 2013: Comparison of the performance of best linear unbiased estimation and best linear unbiased prediction of genotype effects from zoned Indian maize data. Crop Science 53: 1384-1391.

Kubinger, K.D., D. Rasch & K. Moder, 2009: Zur Legende der Voraussetzungen des t-Tersts für unabhängige Stichproben. Psychologische Rundschau 60(1): 26-27.

Kuckuck, H. & A. Mudra, 1950: Lehrbuch der allgemeinen Pflanzenzüchtung. S. Hirzel Verlag, Stuttgart

Laidig, F., T. Drobek & U. Meyer, 2008: Genotypic and environmental variability of yield for cultivars from 30 different crops in German official variety trials. Plant Breeding 127: 541-547.

Laidig, F., 2013: Anmerkungen zu FITCON. Mitteilung per Email, Hannover.

Laidig, F., H.P. Piepho, T. Drobek & U. Meyer, 2014: Genetic and non-genetic long-term trends of 12 different crops in German official variety performance trials and on-farm yield trends. Theor. Appl. Genet. 127: 2599-2617.

Limpert, E. & W.A. Stahel, 2006: Die Normalverteilung als bisheriger Standard der Variation von Versuchsdaten: Wie normal ist sie eigentlich? Zusammenfassung. Tagungsband 55. Deutsche Pflanzenschutztagung. Göttingen.

Little, R.J. A. & D.B. Rubin, 2002: Statistical analysis with incomplete data. 2nd ed. John Wiley & Sons, New York.

Meuwissen, T.H.E., B.J. Hayes & M.E. Goddard, 2001: Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829

Michel, V. & G. Pienz, 2007: Sortenwahl in der Praxis - eine komplexe Abwägung. Welchen Anteil hat die Resistenz?. Mitteilungen der Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern 37: 37-40.

Michel, V., R. Graf, J. Möhring, A. Büchse & H.P. Piepho, 2007a: The „Hohenheim-Gülzow-method" for analysis of series of trials as basic procedure for PIAFStat and SAS in a regionalized field trials system. Proceedings of the international symposium 'Agricultural Field Trials – Today and Tomorrow' October 8th to 10th, 2007, Stuttgart, Germany: 136-141.

Michel, V., G. Zink, J. Schmidtke & A. Anderl, 2007b: Product presentation ‚PIAF and PIAFStat'. proceedings of the international symposium 'Agricultural Field Trials – Today and Tomorrow' October 8th to 10th, 2007, Stuttgart, Germany: 278-279.

Michel, V. & A. Zenk, 2009: http://www.landwirtschaft-mv.de/cms2/LFA_prod/LFA/content/de/Fachinformationen/Sorten/Anbaugebiete/index.jsp. Gülzow.

Michel, V. & A. Zenk, 2010: Eignung von Winterweizensorten unter speziellen Anbaubedingungen und Einführung neuer Parameter zur Bewertung von Sorten unter besonderer Berücksichtigung klimatischer Veränderungen. Forschungsbericht 6/10 der Landdesforschungsanstalt für Landw. u. Fischerei MV. http://www.landwirtschaft-mv.de/cms2/LFA_prod/LFA/content/de/ Fachinformationen/Sorten/Beitraege/index.jsp?&artikel=2425.

Möhring, J., A. Büchse & H.P. Piepho, 2004a: Optimierung des Prüfsystems bei Winterweizen. Studie im Auftrag von Bundessortenamt, Bundesverband Deutscher Pflanzenzüchter und Verband der Landwirtschaftskammern. 19.03.2004, Stuttgart.

Möhring, J., A. Büchse & H.P. Piepho, 2004b: Bereitstellung von Werkzeugen für die ‚Auswertung von überlappenden Anbaugebieten' am Beispiel Winterweizen. Projekt vom Bundesverband Deutscher Pflanzenzüchter und dem Verband der Landwirtschaftskammern. 24.08.2004, Stuttgart.

Möhring, J, A. Büchse, H.P. Piepho, V. Michel, J. Rath & F. Laidig, 2004c: Gesundsparen ohne Nachteile. DLG-Mitteilungen 6.

Möhring & H.P. Piepho, 2009: Comparison of weighting in two-stage analysis of plant breeding trials. Crop Science 49: 1977-1988.

Mudra, A., 1949: Anleitung zur Durchführung und Auswertung von Feldversuchen nach neueren Methoden. S. Hirzel Verlag, Leipzig.

Mudra, A., 1952: Einführung in die Methodik der Feldversuche. S. Hirzel Verlag, Leipzig.

Mühleisen, J., H.P. Piepho, H.P. Maurer, C.F.H. Longin & J.C. Reif, 2014a: Yield stability of hybrids versus lines in wheat, barley, and triticale. Theor. Appl. Genet. 127: 309-316.

Mühleisen, J., H.P. Piepho, H.P. Maurer, Y. Zhao & J.C. Reif, 2014b: Exploitation of yield stability in barley. Theor. Appl. Genet. 127: 1949-1962.

Nelder, J.A., 1986: Statistics, science and technology: the address of the President. Journal of the Royal Statistical Society Series A 149: 109-121.

Ouk, M., J. Basnayake, M. Tsubo, S. Fukai, K.S. Fischer, S. Kang, S. Men, V. Thun & M. Cooper, 2006: Genotype-by environment interactions for grain yield associated with water availability in rainfed lowland rice. Field Crops Research 101: 145-154.

Patterson, H.D. & Thompson, R., 1971: Recovery of inter-block information when block sizes are unequal. Biometrika, 58: 545-554.

Patterson, H.D., 1978: Routine least squares estimation of variety means in incomplete tables. J. Natn. Inst. Agric. Bot. 14: 401-412.

Patterson, H.D., 1997: Analysis of series of variety trials, in: Statistical methods for Plant Variety Evaluation (ed. Kempton, R.A. & P.N. Fox). Chapman & Hall, London: 139-161.

Piepho, H.P., 1997: Analyzing genotype-environment data by mixed models with multiplicative effects. Biometrics 53: 761-766.

Piepho, H.P., 1998: Auswertung von Bonituren des Typs “Prozent Befall” mit SAS-Prozeduren für generalisierte Lineare Modelle. Zeitschrift für Agrarinformatik 2/98: 26-37.

Piepho, H.P., 1999: Fitting a regression model for genotype-by-environment data by methods for nonlinear mixed models. Biometrics 55: 1120-1128.

Piepho, H.P. & F.A. van Eeuwijk, 1999: Stability analyses in crop performance evaluation. In: Kang, M. (ed.), Crop improvement for the 21st century. Haworth Press, New York.

Piepho, H.P. & V. Michel, 2001: Überlegungen zur regionalen Auswertung von Landessorten-versuchen. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 31: 123-136.

Piepho, H.P., 2003: The folded exponential transformation for proportions. The Statistician 52: 575-589.

Piepho, H.P., A. Büchse & K. Emrich, 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. & J.O. Ogutu, 2003: Inference for the break point in segmented regression with application to longitudinal data. Biometrical Journal 45: 591-601.

Piepho, H.P., 2005: Erfassung der phänotypischen Stabilität in züchterischen Versuchen. In: Bericht über die 56. Tagung der Vereinigung der Pflanzenzüchter und Saatgutkaufleute Österreichs. Gumpenstein 22. - 24. 11.05: 93-98.

Piepho, H.P. & J. Möhring, 2005: Best linear unbiased prediction for subdivided target regions. Crop Science 45: 1151-1159.

Piepho, H.P. & J. Möhring, 2006: Selection in cultivar trials – is it ignorable?. Crop Science 46: 193-202.

Piepho, H.P. & J. Möhring, 2007: On weighting in two-stage analysis of series of experiments. Biuletyn Oceny Odmian 32: 109-121.

Piepho, H.P., J. Möhring, A., E. Melchinger & A. Büchse, 2008: BLUP for phenotypic selection in plant breeding and variety testing. Euphytica 161: 209-228.

Piepho, H.P., 2009: Data transformation in statistical analysis of field trials with changing treatment variance. Agron. Journal 101: 865-869.

Piepho, H.P. & J. Möhring, 2011: On Estimation of Genotypic Correlations and Their Standard Errors by Multivariate REML using the MIXED Procedure of the SAS System. Crop Science 51: 2449-2554.

Piepho, H.P., J. Möhring, A, T. Schulz-Streeck & J.O. Ogutu, 2012: A stage-wise approach for the analysis of multi-environment trials. Biometrical Journal 54(6): 844-860.

Piepho, H.P., B.U. Müller & C. Jansen, 2014: Analysis of a complex trait with missing data on the component traits. Communications in Biometry and Crop Science 9(1): 26-10.

Przystalski, M., A. Osman, E.M. Thiemt, B. Rolland, L. Ericson, H. Østergård, L. Kevy, M. Wolfe, A. Büchse, H.P. Piepho & P. Krajewski, 2008: Comparing the performance of cereal varieties in organic and non-organic cropping systems in different European countries. Euphytica 163: 417-433.

Ramburan, S., M. Zhou & M.T. Labuschagne, 2012: Investigating test site similarity, trait relations and causes of genotype × environment interactions of sugarcane in the Midlands region of South Africa. Field Crops Research 129: 71-80.

Rasch, D., 1987: Biometrisches Wörterbuch. VEB Deutscher Landwirtschaftsverlag, Berlin.

Rasch, D., G. Herrendörfer, J Bock, N. Victor & V. Guiard, 1998: Verfahrensbibliothek. Versuchsplanung und -auswertung. Band II. Oldenbourg, München.

Richter, Ch., V. Guiard & F. Krüger, 1999: Auswertung von Versuchsserien mit zwei Prüffaktoren in Anlagen mit vollständigen Blocks. Zeitschrift für Agrarinformatik 7: 10-22.

Richter, Ch., 2004: Grundbegriffe und Datenanalyse. Einführung in die Biometrie 1. Senat der Bundesforschungsanstalten des Bundesministeriums für Verbraucherschutz, Ernährung und Landwirtschaft.

Richter, Ch., B. Kroschewski, V. Michel & A. Zenk, 2007: Analysis of standard designs using spatial models. Proceedings of the international symposium ‘Agricultural Field Trials – Today and Tomorrow’ October 8th to 10th, 2007, Stuttgart, Germany: 176-181.

Robinson, D.L., 1987: Estimation and use of variance components. The Statistician 36: 3-14.

Rochon, J. & M. Kieser, 2011: A closer look at the effect of preliminary goodness-of-fit testing for normality for the one-sample t-test. Br. J. Math. Stat. Psychol. 64(3): 410-426.

Roemer, Th., 1930: Der Feldversuch - eine kritische Studie. Arbeiten der DLG, Berlin.

Roßberg, D., V. Michel, R. Graf & R. Neukampf, 2007: Definition von Boden-Klima-Räumen für die Bundesrepublik Deutschland. Nachrichtenblatt Deut. Pflanzenschutz. 59(7): 155-161.

Rutz, H.W., 2006: Sorten- und Saatgut-Recht. 11. Aufl. Agrimedia, Bergen: 61-62.

SAS Institute, Inc., 1999: SAS/STAT User's Guide, version 8, Cary, NC,USA.

Schucany, W.R. & H.K. Tony Ng, 2006: Preliminary Goodness-of-Fit Tests for Normality do not validate the One-Sample Student t. Communications in Statistics - Theory and Methods 35(12): 2275-2286.

Searle, S.R., G. Casella & C.E. McCulloch, 1992: Variance components. J.Wiley&Sons, New York.

Silvey, V., 1978: Methods of analysing NIAB variety trials over many sites and several seasons. Journal of the National Institute of Agricultural Botany, 14: 385-400.

Smith, A.B., B.R. Cullis & R. Thompson, 2001a: Analyzing variety by environment data using multiplicative mixed models and adjustments for spatial field trend. Biometrics 57: 1138-1147.

Smith, A.B., B.R. Cullis & A.R. Gilmour, 2001b: The analysis of crop variety evaluation data in Australia. Australian and New Zealand Journal of Statistics 43: 129-145.

Smith, A.B., B.R. Cullis & R. Thompson, 2005: The analysis of crop cultivar breeding and evaluation trials: an overview of current mixed model approaches. Journal of Agricultural Science 143: 449-462.

Smith, A.B., A. Ganesalingam, H. Kuchel & B.R. Cullis, 2015: Factor analytic mixed models for the provision of grower information from national crop variety testing programs. Theor. Appl. Genet. 128: 55-72.

Stroup, W., 2014: Rethinking the analysis of non-normal data in plant and soil science. Agron. Journal. 106: 1-17.

Student, 1923: On testing varieties of cereals. Biometrika 13: 271-295.

Thomas, E., 2006: Feldversuchswesen. Eugen Ulmer KG, Stuttgart.

Thöni, H., 2007: Anmerkungen zu: Die Hohenheim-Gülzower Serienauswertung als bundesweites Basisverfahren im regionalisierten Sortenwesen. Mitteilung per Email, Stuttgart.

Utz, H.F., 1972: Die Zerlegung der Genotyp x Umwelt-Interaktionen. EDV in Medizin und Biologie. Stuttgart 2(1972).

van Eeuwijk, F., 2007: Statistical methods for the analysis of series of plant breeding experiments. Proceedings of the international symposium 'Agricultural Field Trials – Today and Tomorrow' October 8th to 10th, 2007, Stuttgart, Germany: 242-249.

Weber, V.S., S. Wagener, C. Magorokosho, D. Makumbi, B. Vivek, H.P. Piepho, A.E. Melchinger & G.N. Atlin, 2012: Strategies to subdivide a target population of environments: Result from the CIMMYT-Led Maize Hybrid Testing Programs in Africa. Crop Science 52: 1-10.

Welham, S., B.J. Gogel, A.B. Smith & B.R. Cullis, 2010: A comparison of analysis methods for late-stage evaluation trials. Australian and New Zealand Journal of Statistics 52: 125-149.

Williams, E., 2007: Developments in experimental design for agricultural field trials - spatial gain and randomization loss. Proceedings of the international symposium 'Agricultural Field Trials – Today and Tomorrow' October 8th to 10th, 2007, Stuttgart, Germany: 264-269.

Wricke, G., 1967: Die Erfassung der Wechselwirkung zwischen Genotyp und Umwelt bei quantitativen Eigenschaften. Rundschreiben 1/67 der Arbeitsgemeinschaft für Biometrie bei der DLG-Pflanzenzuchtabteilung, Frankfurt.

Yates, F., 1933: The principles of orthogonality and confounding in replicated experiments. Journal of Agricultural Science 23: 108-145.

Yates, F. & W.G. Cochran, 1938: The analysis of groups of experiments. Journal of Agricultural Science 28: 556-580.

Zimmermann, D.W., 1996: Some Properties of Preliminary Tests of Equality of Variances in the Two-Sample Location Problem. Journal of General Psychology 123(3): 217-231.

Zimmermann, D.W., 2004: A note on preliminary tests of equality of variances. Br. J. Math. Stat. Psychol.57(1): 173-181.

Tabellenverzeichnis

Tab. 1: Aussagebereich von Feldversuchen nach Buhtz und Bätz (1984) 14
Tab. 2: In die Auswertungen einbezogene Pflanzenarten 24
Tab. 3: Einbezogene Merkmale; Teil 1 Kategorien ‚Ertrag' und ‚Minderertrag' 27
Tab. 4: Einbezogene Merkmale; Teil 2 Kategorien *Qualität [%]* und *sonstige Qualität* 28
Tab. 5: Faktoren, Klassifizierungsgrößen und verwendete Kürzel 34
Tab. 6: Anzahl Faktorstufen und Werte in den Datasets der Auswertungsserien 35
Tab. 7: Verwendete Label für Stufen der Faktoren und Klassifizierungsgrößen am Beispiel Winterweizen; Datensatz für Auswertung des Kornertrages für Zielgebiete in Mecklenburg-Vorpommern 36
Tab. 8: Gruppierung der Winterweizensorten nach Prüfstatus* 38
Tab. 9: Kernsortiment Winterweizen - Statusinformationen im Anbaujahr 2008/09 40
Tab. 10: Kernsortiment Winterweizen - Anzahl Versuche je Sorte im Auswertungszeitraum 2001-2009 (N=9) bzw. 2004-2009 (N=6) 41
Tab. 11: Kernsortiment Winterweizen - Versuchstyp je Sorte und Jahr 42
Tab. 12: Räumliche und zeitliche Datensatzstruktur; Kornertrag Winterweizen 46
Tab. 13: Anzahl Mittelwerte je Sorte im Kernsortiment, gruppiert nach Prüfstatus 51
Tab. 14: Durchschnittliche Anzahl Mittelwerte je Sorte im Kernsortiment nach Jahren, gruppiert nach Prüfstatus 52
Tab. 15: Regionalisierungskonstellationen für Zielgebiete in Mecklenburg-Vorpommern und einbezogene Nachbargebiete 62
Tab. 16: Auswertung für ein Anbaugebiet; schrittweise Herleitung des Auswertungsmodells 72
Tab. 17: Modellerweiterungen zur Auswertung für ein Zielgebiet bei gewichteter Einbeziehung von Nachbargebieten 73
Tab. 18: Modellwahl für die genetische Kovarianzmatrix in unterschiedlich besetzten Datensätzen 79
Tab. 19: Bestimmung der Gewichte für Ziel- und Nachbargebiete in Abhängigkeit vom Regionalisierungskonzept 82
Tab. 20: Übersicht über die SAS® - Makros und deren Kurzbeschreibung 91
Tab. 21: PIAFStat-Verfahren der *Hohenheim-Gülzower-Serienauswertung* 93
Tab. 22: Geschätzte Transformationsparameter φ ; *Box-Cox-Transformation* 100
Tab. 23: Klassifizierung und Interpretation des Datenverhaltens für den Transformationsparameter φ (*Box-Cox-Transformation*) 101
Tab. 24: Schiefe und Exzess der Verteilung der Residuen; ohne versus mit Transformation 104
Tab. 25: Lineare Abhängigkeit der absoluten Residuen von den Vorhersagewerten 109
Tab. 26: Vergleich der Transformationsansätze bei Prozentmerkmalen (Qualität [%]) 111
Tab. 27: Relevanz einer Transformation für die Mittelwertschätzung im Kernsortiment; ‚Minderertrag geringe Intensität' bei Winterweizen 116
Tab. 28: Sorte × Umwelt - Interaktionsvarianzen im Vergleich zu Fehlervarianzen der Sortenmittelwerte aus den Einzelversuchen 126
Tab. 29: Relativer Anteil der Komponenten der Sorte × Umwelt - Varianz an dieser Umweltvarianz 127
Tab. 30: Sortenvarianz und ihre Anbaugebietsspezifik 130
Tab. 31: Definition von Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen 131

Tab. 32: Auswirkung von Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen 132

Tab. 33: Relevanz von Begrenzungen der Datenbasis für die Mittelwertschätzung im Kernsortiment; Kornertrag Winterweizen 137

Tab. 34: Relevanz einer Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte ihrer Fehlervarianzen für die Mittelwertschätzung in der Serienauswertung; Kornertrag Winterweizen 140

Tab. 35: Genetisch Korrelation der Sorteneffekte von Nachbargebieten zum jeweiligen Zielgebiet 147

Tab. 36: Verteilung der Gewichte bei der Ertragsschätzung für ein Zielgebiet auf einbezogene Ziel- und Nachbargebiete; Kornertrag Winterweizen, Zielgebiet D-Nord 150

Tab. 37: Häufigkeitsverteilung für die Gewichtung der S × AG - Effekte (W_{AG}) in den Anbaugebieten; Kornertrag Winterweizen-Kernsortiment, Zielgebiet D-Nord, ... 151

Tab. 38: Definition von Szenarien für Regionalisierungsstrategien (s. 5.5) 151

Tab. 39: Auswirkung von Szenarien für die Regionalisierung auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen 152

Tab. 40: Auswirkung der Regionalisierungsstrategien auf die Mittelwerte; Kornertrag Winterweizen 154

Tab. 41: Unterschiede zwischen BLUE- und BLUP- Ansatz für Sorteneffekte in der Hohenheim-Gülzower-Serienauswertung 162

Tab. 42: Für die Schätzung von Sorteneffekten im Folgejahr 2010 verwendete Varianzkomponenten; Kornertrag Winterweizen 173

Abbildungsverzeichnis

Abb. 1: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Ausgangsdaten; Merkmale der Kategorien ‚Ertrag' und ‚Minderertrag' 29

Abb. 2: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Variationskoeffizienten der Mittelwerte in Einzelversuchen ($SE_{Versuch}$%); Merkmale der Kategorien ‚Ertrag' und ‚Minderertrag' 30

Abb. 3: Deskriptive Darstellung (Box-Whisker-Plots) der relativen Variabilität und Symmetrie der Ausgangsdaten; Merkmale der Kategorien ‚Qualität [%]' und ‚sonstige Qualität' 31

Abb. 4: Regelprüfablauf im integrierten Sortenprüfsystem am Beispiel Winter-getreide; Stand 2009, MV 48

Abb. 5: Anzahl Mittelwerte je Sorte aus Einzelversuchen und deren Häufigkeiten im Sortiment Winterweizen (log-skaliert) 53

Abb. 6: Boden-Klima-Räume für alle Pflanzenarten in Deutschland (nach Roßberg et al., 2007) 58

Abb. 7: Anbaugebiete Winterweizen in Deutschland (Graf et al., 2009) 60

Abb. 8: Anbaugebiete Sommerweizen in Deutschland (Graf et al., 2009) 60

Abb. 9: Regionalisierung Winterweizen und Raps (*Ertrag*); Zielgebiete (a) D-Nord bzw. (b) MV-Süd und jeweils einbezogene Nachbargebiete 63

Abb. 10: Schätzung des Transformationsparameter phi (φ) (Beispiel) 65

Abb. 11: Stauchung und Streckung bei Transformationsfamilien ohne versus mit Wendepunkt 68

Abb. 12: PIAFStat als Verbindung zwischen PIAF und SAS® nach Schmidtke in Michel et al. (2007b) 90

Abb. 13: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *PHI* 94

Abb. 14: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *VK* 95

Abb. 15: Menüfenster ‚*Optionen*' des PIAFStat-Verfahrens *MW* 95

Abb. 16: Vertrauensintervall für den Transformationsparameter bei großem und bei kleinem Beispiel-Datensatz 99

Abb. 17: Häufigkeit von Größenklassen für φ nach Box-Cox-Transformation; gruppiert nach Merkmalskategorien 101

Abb. 18: Plot der Kennziffern Schiefe und Exzess über alle Merkmale gegen die geschätzten Transformationsparameter; ohne und mit Transformation 105

Abb. 19: Anpassung studentisierter Residuen an die Normalverteilung bei besonders auffälligen Merkmalen (Q-Q-Plots) 107

Abb. 20: Plot der linearen Anstiege (absolute Residuen als Funktion der Vorhersagewerte) über alle Merkmale gegen die Transformationsparameter φ 110

Abb. 21: Häufigkeitsverteilungen; Merkmal ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste; Original-Daten und zwei Transformationsansätze 112

Abb. 22: Häufigkeitsverteilung der Residuen, ohne und mit der *‚folded exponential transformation'* ; ‚*Sortierung > 2,5 mm (%)*' bei Sommerbraugerste; 113

Abb. 23: Plot der Residuen gegen die geschätzten Vorhersagewerte, ohne und mit der *‚folded exponential transformation'*; ‚*Sortierung > 2,5 mm (%)*'/100 bei Sommerbraugerste 113

Abb. 24: Transformationsparameter in Beziehung zur Lage des Wertebereiches bei Merkmalen der Kategorie ‚*Qualität [%]* ' 114

Abb. 25: Einfluss der Transformation auf die Mittelwertschätzung; ‚Minderertrag geringe Intensität' bei Winterweizen 115

Abb. 26: Einfluss der Szenarien zur Einbeziehung von Jahren und vorgelagerten Versuchen auf die Genauigkeit im Kernsortiment; Kornertrag Winterweizen 133

Abb. 27: Beziehung zwischen Datenumfang und Genauigkeit bei Begrenzung des Auswertungszeitraumes auf sechs Jahre; Kornertrag Winterweizen 135

Abb. 28: Einfluss von Begrenzungen der Datenbasis auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen 136

Abb. 29: Häufigkeitsverteilung der Standardfehler der Sortenmittelwerte in Einzelversuchen ($SE_{Versuch}$); Kornertrag Winterweizen..... 138

Abb. 30: Einfluss der Gewichtung der Sortenmittelwerte aus Einzelversuchen mittels der Kehrwerte der Fehlervarianzen auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen..... 139

Abb. 31: Einfluss von Regionalisierungsstrategien auf die Genauigkeit; Kornertrag Winterweizen..... 153

Abb. 32: Einfluss der gewichteten Einbeziehung von Nachbargebieten auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen 155

Abb. 33: Einfluss der Gewichtungsabstufung von Anbaugebieten gegenüber einer Großraumauswertung auf Genauigkeit und Mittelwertschätzung; Kornertrag Winterweizen..... 155

Abb. 34: Prioritätenwandel bei der Einordnung des Faktors ‚Sorte' im Verlauf des Prüfprozesses - schematisch..... 164

Abb. 35: Zusammenhang von Proteingehalt und BLUE- versus BLUP-Ertragsschätzungen bei Weizensorten 166

Abb. 36: Systematischer Einfluss des Proteingehaltes auf den Unterschied zwischen BLUE- und BLUP-Ertragsschätzung von Weizensorten 167

Abb. 37: Auswirkung der Auslassung jeweils eines von neun Jahren auf die Varianzkomponenten; Kornertrag Winterweizen 170

Abb. 38: Abweichungen geschätzter Sorteneffekte infolge der Auslassung jeweils eines von neun Jahren bei der Varianzkomponentenschätzung, Kornertrag Winterweizen..... 171

Abb. 39: Phänotypische Korrelation zwischen der sechsjährigen Ertragsschätzung 2004-09 und der Ertragsschätzung im Folgejahr 2010; Kornertrag Winterweizen 174

Anlagenverzeichnis

Anlage 1: Datensatz für die Versuchsserienauswertung mit PIAFStat; Kornertrag Winterweizen 2001-2009 (Auszug) 201

Anlage 2: Auszug aus dem SAS®-Makro *%boxcox* *; Definitionen, Variablen, Grobstruktur des Makros und Makroaufruf 202

Anlage 3: Output des PIAFStat-Verfahrens ‚PHI' (zentrales SAS®-Makro: *%boxcox*); Datensatz Kornertrag Winterweizen; (Auszug) 203

Anlage 4: Auszug aus dem SAS®-Makro *%varianzkomponenten* *; Teil 1: Definitionen, Variablen und Grobstruktur des Makros 205

Anlage 5: Auszug aus dem SAS®-Makro *%varianzkomponenten* *; Teil 2: Makroaufruf und dynamisch erzeugte Kernprozedur an Beispielen 206

Anlage 6: Output des PIAFStat-Verfahrens ‚VK' (zentrales SAS®-Makro: *%varianzkomponenten*); Datensatz Kornertrag Winterweizen; (Auszug) 207

Anlage 7: Auszug aus dem SAS®-Makro *%lsm* *; Teil 1: Definitionen, Variablen und Grobstruktur des Makros 210

Anlage 8: Auszug aus dem SAS®-Makro *%lsm*; Teil 2: Makroaufruf und dynamisch erzeugte Kernprozedur an Beispielen 211

Anlage 9: Output des PIAFStat-Verfahrens ‚MW' (zentrales SAS®-Makro: *%lsm*); Kornertrag sechsjährig Winterweizen 2004-2009 für das Zielgebiet D-Nord*; (Auszug) 212

Anlagen

Anlage 1: Datensatz für die Versuchsserienauswertung mit PIAFStat; Kornertrag Winterweizen 2001-2009 (Auszug)

jahr	ORTBEZ	ort	VTYP	typ	r_BEZ	r	s_Name	s	gr	ERTR86DT	SE	w
2001	Badingen	1223	LSV	1	MV-Süd	45	Bussard	WW 01641	E	54,5	2,98	0,11
2001	Badingen	1223	LSV	1	MV-Süd	45	Ritmo	WW 01889	B	60,3	2,98	0,11
2001	Badingen	1223	LSV	1	MV-Süd	45	Batis	WW 01968	A	65,0	2,98	0,11
2001	Gülzow	118	LSV	1	D-Nord	4	Bussard	WW 01641	E	97,9	1,97	0,26
2001	Gülzow	118	LSV	1	D-Nord	4	Toronto	WW 01649	A	108,0	1,97	0,26
2001	Gülzow	118	LSV	1	D-Nord	4	Ritmo	WW 01889	B	121,6	1,97	0,26
2001	Gülzow	118	OV	2	D-Nord	4	Ritmo	WW 01889	B	115,3	1,64	0,37
2001	Gülzow	118	OV	2	D-Nord	4	Batis	WW 01968	A	106,1	1,64	0,37
2001	Gülzow	118	OV	2	D-Nord	4	Drifter	WW 02528	B	113,7	1,64	0,37
2002	Tützpatz	113	WP3	1	D-Nord	4	Bussard	WW 01641	E	91,8	2,19	0,21
2002	Tützpatz	113	WP3	1	D-Nord	4	Ritmo	WW 01889	B	99,3	2,19	0,21
2002	Tützpatz	113	WP3	1	D-Nord	4	Batis	WW 01968	A	103,5	2,19	0,21
2002	Tützpatz	113	LSV	2	D-Nord	4	Bussard	WW 01641	E	90,6	1,52	0,44
2002	Tützpatz	113	LSV	2	D-Nord	4	Toronto	WW 01649	A	103,0	1,52	0,44
2002	Tützpatz	113	LSV	2	D-Nord	4	Ritmo	WW 01889	B	97,2	1,52	0,44
2002	Tützpatz	113	EU	3	D-Nord	4	Batis	WW 01968	A	106,4	1,44	0,48
2002	Tützpatz	113	EU	3	D-Nord	4	Aspirant	WW 02475	A	102,2	1,44	0,48
2002	Tützpatz	113	EU	3	D-Nord	4	Drifter	WW 02528	B	112,5	1,44	0,48
2002	Tützpatz	113	OV	4	D-Nord	4	Batis	WW 01968	A	102,5	1,56	0,41
2002	Tützpatz	113	OV	4	D-Nord	4	Aspirant	WW 02475	A	101,2	1,56	0,41
2002	Tützpatz	113	OV	4	D-Nord	4	Drifter	WW 02528	B	106,3	1,56	0,41
2009	Vipperow	116	LSV	2	MV-Süd	45	Cubus	WW 02787	A	97,8	2,13	0,22
2009	Vipperow	116	LSV	2	MV-Süd	45	Tommi	WW 02880	A	97,5	2,13	0,22
2009	Vipperow	116	LSV	2	MV-Süd	45	Türkis	WW 02991	A	104,1	2,13	0,22

••• Anzahl Werte (Zeilen) insgesamt: 9993

r = AG (Anbaugebiet)

Legende:

Klassifizierungsgrößen in Kernprozeduren der SAS®-Makros

Label von Klassifizierungsgrößen

ausgewählte Analysevariable

Gewichtungsgröße (optional)

mit $w = 1/SE^2$

zeigt Datensatzunterbrechung an: Auslassung von Sorten, Typen, Orten, Jahren

Beispiel für mehrere Versuchstypen an einer Ort × Jahr - Kombination; aufsteigende Nummerierung in ‚typ' durch Makro %zuordnung_typ

Anlage 2: Auszug aus dem SAS®-Makro *%boxcox* *; Definitionen, Variablen, Grobstruktur des Makros und Makroaufruf

```
*************************************************************************;
*  Bestimmung den Transformationsparameter mit bester Modellanpassung   ;
*************************************************************************;

*data      Name des Datensatzes                                         ;
*phimin    unterer Startwert                                            ;
*phimax    oberer Startwert                                             ;
*response= Zielmerkmal (optional)                                       ;
*class=    Klassifizierungsvariablen (optional)                         ;
*model=    feste Effekte (optional)                                     ;
*ddfm=     Methode zur Berechnung der Freiheitsgrade (optional) sowie   ;
*mix_statem=sonstige mixed-Anweisungen wie z.B. random (optional)       ;
*Angabe durch : mix_statem=%str("Anweisungen")                          ;
*ohne optionale Angabe wird das unten ausgewiesene Modell angepasst     ;
*Der Datensatz sollte folgende Spalten enthalten:                       ;
*s, jahr, ort, r, typ, gr sowie event. w (für Sorte, Jahr, Ort,         ;
*Region=AG=Anbaugebiet, Versuchstyp, Gruppe und Gewicht).               ;
*************************************************************************;

%macro boxcox(data,phimin=&PMIN,phimax=&PMAX,response=y,
class=%str(gr s jahr r ort typ;),
model= jahr*ort*typ gr*r /* aus performance-Gründen optional
auskommentierbar, da in Kombinationse-Effekten enthalten: gr jahr r ort
jahr*r jahr*ort gr*jahr gr*ort gr*jahr*r gr*jahr*ort */
,ddfm=%str(ddfm=residual),
mix_statem=%str(random int jahr ort jahr*r jahr*ort /subject=s;random
r/type=UN(1) subject=s;),

*Bestimme Varianzkomponenten als Startwerte für die Iteration;    …
*Bestimme zweiten Punkt;                                          …
*Bestimme dritten Punkt;                                          …
*Bestimme neues phi_n;                                            …
*Falls phi_n=. dann goldener Schnitt;                             …

%mend boxcox; *************Macro Ende***********************************;

**************************Macro ausführen*******************************;
%boxcox(dataset);
*************************************************************************;
```

… Programmcode ausgeblendet (nicht frei verfügbar), daher nicht selbsterklärend

* auf Grundlage von Möhring et al. (2004b)

Anlage 3: Output des PIAFStat-Verfahrens ‚PHI' (zentrales SAS®-Makro: *%boxcox*); Datensatz Kornertrag Winterweizen; (Auszug)

SAS-List

Verfahrensname: PHI - Hohenheim-Gülzower Serienauswertg.- optimale Datentransformation
Konzeption: Michel, Piepho
Umsetzung in PIAFStat: Zenk
Biomertische Grundlagen: Piepho, Michel, Möhring, Büchse
SAS-Macros: Möhring

Auswertungsserie Datum
LSV WW 2009 Montag, 14. Dezember 2009

9993 Datensätze wurden eingelesen!

Übersicht: Versuchanzahl je Jahr und Ort - Zuordnung zu den Anbaugebieten (AG)

Bitte sorgen Sie dafür, dass jeder Ort im Sinne eines Standortes nur einmal vorkommt!
Alle Orte müssen zudem einem Anbaugebiet ungleich 0 zugeordnet sein!

				Anzahl Versuche								
				2001	2002	2003	2004	2005	2006	2007	2008	2009
AG	BKR	Land										
4	101	11	Boldebuck	.	1	1	1	1	.	1	1	1
			Groß Kiesow	1	1	1	1	1	1	1	1	1
			Gülzow	2	2	1	.	2	2	2	2	2
			Tützpatz	4	4	3	3	3	3	3	3	3
		12	Prenzlau	3	3	3	3	3	3	2	2	3
	158	11	Biestow	1	1	1	1	1	1	1	1	1
			Granskevitz	2	2	2	1	.	1	1	1	1
			Köchelstorf	1	1	1	1	1	1	1	1	1
5	104	12	Dürrenhofe	1	1	1	1	1	.	.	1	.
			Güterfelde	1	1	1	1	1	1	1	1	1

(Continued) **ab hier weitere Zeilen dieser Übersicht ausgeblendet**

Übersicht zu den Class-Merkmalen im Gesamt-Datensatz

Überprüfen Sie bitte insbesondere, ob kein Anbaugebiet mit null klassifiziert ist!

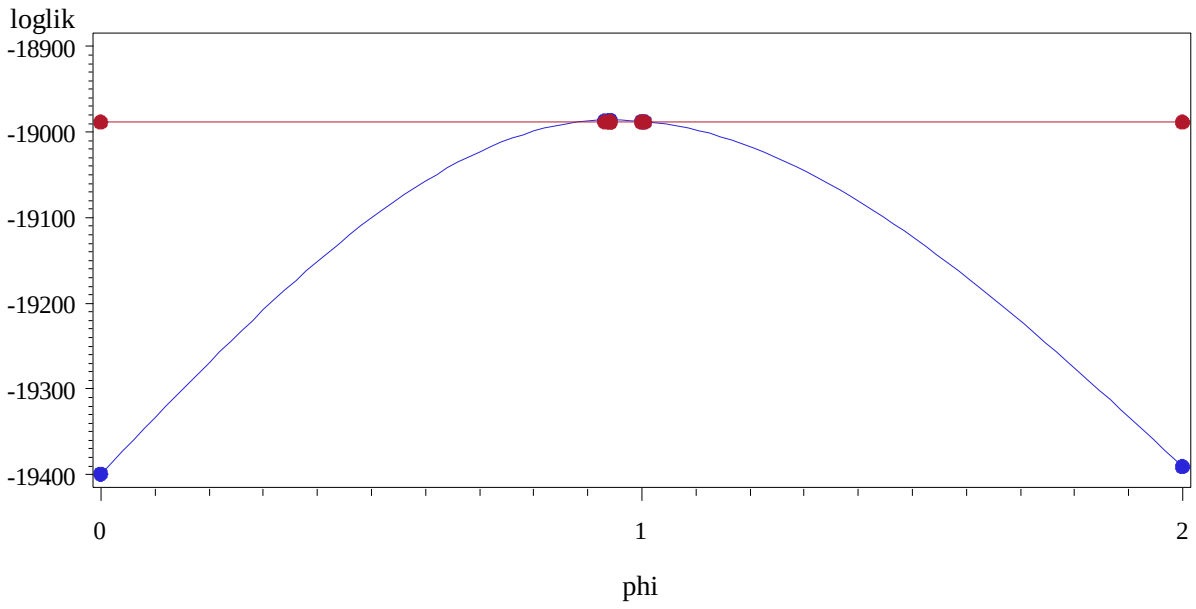
Class-Merkmal		Anzahl Klassen
JAHR	2001 2002 2003 2004 2005 2006 2007 2008 2009	9
Land	1 3 11 12 13	5
Orte	...	39
gr	A B BH C CH CK E	7
r = AG	3 4 5 45 82	5
s	...	1118

gewählte Startwerte: Minimum=0 und Maximum=1

Optimaler Wert von Phi

phi	loglik
0.94136	-18986.18

NOTE: The SAS System used:
real time 46:45:28.48
cpu time 20:55:25.57



© BioMath GmbH - PIAFStat Version 6.08 vom 01.09.2009

Anlage 4: Auszug aus dem SAS®-Makro *%varianzkomponenten* *; Teil 1: Definitionen, Variablen und Grobstruktur des Makros

```
*************************************************************************************;
*             Schätzen der Varianzkomponenten aus langjährigen Daten                *;
*************************************************************************************;

*data=         Name des Datensatzes;
*regionen=     in den Daten vorkommende Anbaugebiete                                  ;
*              Angabe durch : (AG1,..., AGn), wobei AGi als Zahl codiert sein muss    ;
*class=        Klassifizierungsvariablen                                              ;
*model=        feste Effekte                                                          ;
*response=     Zielmerkmal                                                            ;
*ddfm=         Methode zur Berechnung der Freiheitsgrade sowie weitere Optionen zur;
*              model-Anweisung(z.B. notest) Angabe durch :  ddfm=%str("Optionen")     ;
*mixstatem=    sonstige mixed-Anweisungen wie z.B. random,                            ;
*              Angabe durch : mix_statem=%str("Anweisungen")                          ;
*parms=        optionale Angabe, dass eine Datei mit Varianzschätzwerten folgt        ;
*              ,Angabe durch :parms=%str(parms/parmsdata=)                            ;
*parmsdata=    Datei mit Startwerten für die Varianzkomponenten                       ;
*phi=          Transformationsparameter (Box-Cox-Transformation)                      ;
*alpha=        Irrtumswahrscheinlichkeit für das Vertrauensintervall der VK           ;
*out_varcomp= Ausgabedatei mit den Varianzschätzwerten                                ;
*out_corr=Ausgabedatei mit den Korrelationen                                         ;
*ohne optionale Angabe wird das unten ausgewiesene Modell angepasst                  ;
*Der Datensatz sollte folgende Spalten enthalten: s, jahr, ort, r, typ, gr sowie     ;
*evtl. w für Sorte, Jahr, Ort, Region=AG=Anbaugebiet, Versuchstyp, Gruppe,Gewicht.;
*Ausgabedatei ist die in out_varcomp angegebene Datei. Sie enthält die Var.kompon.;
*************************************************************************************;

%Macro varianzkomponenten(data=,regionen=,
class=,model=,response=,ddfm=,
mixstatem=,parms=,parmsdata=,out_varcomp=,out_corr=,
alpha=,phi=);
*Erzeuge Arbeitsdatendatei;                                                          …
*Transformiere die Daten und die SE für Gewichtung;                                  …
*Bestimme Varianzkomponenten;                                                        …
*Definiere Variablen;                                                                …
*Anzahl der Regionen (anzahl_r), größte Nummer einer Region (max_nr_r);              …
*Bestimme genetische Korrelation;                                                    …
*************************************************************************************;
%Mend varianzkomponenten; **** Ende der Varianzkomponenten-Berechnung ****;
*************************************************************************************;
```

… Programmcode ausgeblendet (nicht frei verfügbar), daher nicht selbsterklärend

* auf Grundlage von Möhring et al. (2004b)

Anlage 5: Auszug aus dem SAS®-Makro *%varianzkomponenten* *; Teil 2: Makroaufruf und dynamisch erzeugte Kernprozedur an Beispielen

```
**************************Macro ausführen*******************************;

*Kernprozedur dynamisch erzeugt: bei 3 Anbaugebieten, phi=1, ungewichtet,
VC-Struktur;

%varianzkomponenten(data=dataset1,regionen=(3,20,21),
class= gr s jahr r ort typ,response=y,ddfm=%str(ddfm=residual notest;),
model=gr*r typ(jahr*ort) /*jahr*ort standardmäßig auskommentiert, da in
typ(jahr*ort) enthalten*/,mixstatem=%str(
random int jahr ort jahr*r jahr*ort /subject=s;
random r/subject=s type=VC;),
out_varcomp=varianzk3,out_corr=corr,alpha=0.05,phi=1);

************************************************************************;

*Kernprozedur dynamisch erzeugt: bei 5 Anbaugebieten, phi=0.5 gewichtet,
UN(1)-Struktur;

%varianzkomponenten(data=datenk,regionen=(2,3,4,5),
class= gr s jahr r ort typ,response=y,ddfm=%str(ddfm=residual notest;),
model=gr*r typ(jahr*ort) /* jahr*ort standardmäßig auskommentiert, da in
typ(jahr*ort) enthalten*/,mixstatem=%str(
random int jahr ort jahr*r jahr*ort typ(jahr*ort)/subject=s;
random r/subject=s type=UN(1);repeated; weight w;
parms(1)(1)(1)(1)(1)(1)(1)(0)(1)(0)(0)(1)(0)(0)(0)(1)(1)/eqcons=17;),
out_varcomp=varianzk5,out_corr=corr,alpha=0.05,phi=0.5);

************************************************************************;
```

* auf Grundlage von Möhring et al. (2004b)

Anlage 6: Output des PIAFStat-Verfahrens ‚VK' (zentrales SAS®-Makro: *%varianzkomponenten*); Datensatz Kornertrag Winterweizen; (Auszug)

SAS-List

Verfahrensname: VK - Hohenheim-Gülzower Serienauswertung - Varianzkomponenten
Konzeption: Michel, Piepho
Umsetzung in PIAFStat: Zenk
Biomertische Grundlagen: Piepho, Michel, Möhring, Büchse
SAS-Macros: Möhring

Auswertungsserie Datum
LSV WW 2009 Mittwoch, 16. Dezember 2009

9993 Datensätze wurden eingelesen!

Übersicht: Versuchanzahl je Jahr und Ort - Zuordnung zu den Anbaugebieten (AG)

Bitte sorgen Sie dafür, dass jeder Ort im Sinne eines Standortes nur einmal vorkommt!
Alle Orte müssen zudem einem Anbaugebiet ungleich 0 zugeordnet sein!

				Anzahl Versuche								
				2001	2002	2003	2004	2005	2006	2007	2008	2009
AG	BKR	Land										
4	101	11	Boldebuck	.	1	1	1	1	.	1	1	1
			Groß Kiesow	1	1	1	1	1	1	1	1	1
			Gülzow	2	2	1	.	2	2	2	2	2
			Tützpatz	4	4	3	3	3	3	3	3	3
		12	Prenzlau	3	3	3	3	3	3	2	2	3
	158	11	Biestow	1	1	1	1	1	1	1	1	1
			Granskevitz	2	2	2	1	.	1	1	1	1
			Köchelstorf	1	1	1	1	1	1	1	1	1
5	104	12	Dürrenhofe	1	1	1	1	1	.	.	1	.
			Güterfelde	1	1	1	1	1	1	1	1	1

(Continued) **ab hier weitere Zeilen dieser Übersicht ausgeblendet**

Übersicht zu den Class-Merkmalen im Gesamt-Datensatz

Überprüfen Sie bitte insbesondere, ob kein Anbaugebiet mit null klassifiziert ist!

Class-Merkmal		Anzahl Klassen
JAHR	2001 2002 2003 2004 2005 2006 2007 2008 2009	9
Land	1 3 11 12 13	5
Orte	...	39
gr	A B BH C CH CK E	7
r = AG	3 4 5 45 82	5
s	...	1118

Festlegungen

Gewichtung nach SE (=Versuchspräzision): ja

Transformation: Phi = '0.94'

gewähltes Kovarianzstruktur für S*AG: UN(1)

Ausgabe der Varianzkomponenten

Sorten ohne Gruppenzuordnung und Orte mit AG=0 wurden nicht einbezogen!

azobs	CovParm	Subject	Estimate	Alpha	Lower	Upper	r=AG
1	Intercept	s	3.8146	0.2	3.1410	4.7865	.
2	JAHR	s	2.8581	0.2	2.5223	3.2800	.
3	ort	s	0.7946	0.2	0.6227	1.0697	.
4	JAHR*r	s	0.9244	0.2	0.7280	1.2355	.
5	JAHR*ort	s	4.2163	0.2	3.7775	4.7519	.
6	typ(JAHR*ort)	s	1.2247	0.2	0.9176	1.7680	.
7	UN(1,1)	s	1.5496	0.2	1.0518	2.6695	3
8	UN(2,1)	s	0	.	.	.	.
9	UN(2,2)	s	3.66E-34	.	.	.	4
10	UN(3,1)	s	0	.	.	.	.
11	UN(3,2)	s	0	.	.	.	.
12	UN(3,3)	s	4.37E-19	.	.	.	5
13	UN(4,1)	s	0	.	.	.	.
14	UN(4,2)	s	0	.	.	.	.
15	UN(4,3)	s	0	.	.	.	.
16	UN(4,4)	s	0	.	.	.	45
17	UN(5,1)	s	0	.	.	.	.
18	UN(5,2)	s	0	.	.	.	.
19	UN(5,3)	s	0	.	.	.	.
20	UN(5,4)	s	0	.	.	.	.
21	UN(5,5)	s	1.3E-17	.	.	.	82
22	Residual		1.0000	.	.	.	.

Ausgabe Mittelwert aller se² (für residual)

se²	se²(transf.)
7.34550	5.86935

Ausgabe der genetischen Korrelationen

Obs	label	r
1	Korrelation zw. AG3 und AG4	0.84328
2	Korrelation zw. AG3 und AG5	0.84328
3	Korrelation zw. AG3 und AG45	0.84328
4	Korrelation zw. AG3 und AG82	0.84328
5	Korrelation zw. AG4 und AG5	1.00000
6	Korrelation zw. AG4 und AG45	1.00000
7	Korrelation zw. AG4 und AG82	1.00000
8	Korrelation zw. AG5 und AG45	1.00000
9	Korrelation zw. AG5 und AG82	1.00000
10	Korrelation zw. AG45 und AG82	1.00000

© BioMath GmbH - PIAFStat Version 6.08 vom 01.09.2009

Anlage 7: Auszug aus dem SAS®-Makro *%lsm* *; Teil 1: Definitionen, Variablen und Grobstruktur des Makros

```
*************************************************************************;
*Berechnung von gewichteten Mittelwerten (lsmeans)im Zielgebiet          ;
*unter Einbeziehung von Nachbargebieten - ohne Schrumpfungsschätzung     ;
*************************************************************************;
*data=       Name des Datensatzes                                        ;
*parmsdata=  Name des Datensatzes mit den Varianzkomponenten             ;
*out=        Name der Ausgabedatei für die Gewichte                      ;
*out_lsm=    Name der Ausgabedatei für die Sortenschätzwerte             ;
*zielag=     Nummer/ Bezeichnung des Zielgebietes                    ;
*class=      Klassifizierungsvariablen                                   ;
*model=      feste Effekte                                               ;
*response=   Zielmerkmal                                                 ;
*ddfm=       Methode zur Berechnung der Freiheitsgrade sowie             ;
*weitere Optionen zur model-Anweisung (z.B. notest)                      ;
*Angabe durch:  ddfm=%str("Optionen")                                    ;
*mixstatem=  sonstige mixed-Anweisungen wie z.B. random (optional),      ;
*Angabe durch: mix_statem=%str("Anweisungen")                            ;
*parms=      Angabe, dass eine Datei mit Varianzschätzwerten folgt       ;
*Angabe durch: parms=%str(parms/parmsdata=)                              ;
*phi=optionaler Transformationsparameter                                 ;
*Datensatz sollte folgende Spalten enthalten: s, jahr, ort, r, typ, gr sowie evtl.;
*w (für Sorte, Jahr, Ort, Region=AG=Anbaugebiet,Versuchstyp, Gruppe, Gewicht).    ;
*In Modellzeile sollte Effekt s(gr*r) bzw. identisch s*r enthalten sein.  ;
*Ausgabedateien sind die in out und out_lsm angegebene Dateien.          ;
*Sie enthalten die Gewichte und die Sortenschätzwerte für Sorten.        ;
*************************************************************************;
%macro estimate_LSM(anzahl_r);                                           …
%macro lsm(data=,parmsdata=,out=,out_lsm=,zielag=,
class=,model=,response=,ddfm=,
mixstatem=,parms=,phi=);                                                 …
*Transformation von &response und Gewichten;                             …
*Bestimme die umweltbedinge Varianz von Sortenvergleichen;               …
*Estimate-Statements &anzahl_r Anbaugebiete;                             …
*Bestimme Gewichte und MSEP für alle Sorten; Proc iml;                   …
*************************************************************************;
* Schleife über alle Sorten;
*************************************************************************;
  *Einlesen der Varianzen und Kovarianzen (var und cov);                 …
  *Einlesen des geprüften Regionen (Vektor reg);                         …
  *Bestimme umweltbedingte Varianz-Kovarianz-Matrix;                     …
  *Bestimme genetische Varianz-Kovarianz-Matrix;                         …
  *Betrachte Differenz-->doppelte genetische Varianz;                    …
  *Bestimme Summe aus sume und sumd;
  ************************************************************************;
  *Gewicht ohne Summenrestriktion (w=u*sumd*inv(x)                       ;
  ************************************************************************;
  *Gewichte mit Summenrestriktion, La Grange-Multiplikator               …
  ************************************************************************;
  *Berechne Varianz des Prüfsystems, Bias und MSEP                       …
*************************************************************************;
* Ende der Schleife;
*************************************************************************;
*Bestimme lsmeans gr*r und s(gr*r)=s*r;                                  …
*Berechne Differenz zwischen gr*r und s(gr*r);                           …
*Berechne gewichtetes Mittel pro Sorte;                                  …
*Addiere nun Gruppe*Regionen-Mittel hinzu;                               …
*Rücktransformation;                                                     …
*************************************************************************;
%Mend lsm; ********** Ende der Lsmeans-Berechnung **********************;
*************************************************************************;
```

… Programmcode ausgeblendet (nicht frei verfügbar), daher nicht selbsterklärend

* auf Grundlage von Möhring et al. (2004b)

Anlage 8: Auszug aus dem SAS®-Makro *%lsm*; Teil 2: Makroaufruf und dynamisch erzeugte Kernprozedur an Beispielen

```
**************************Macro ausführen*******************************;

* Kernprozedur zur Schätzung der Sorte*Anbaugebiet-Mittelwerte als Basis
für über Anbaugebiete kombinierte gewichtete Mittelwerte je Sorte;

************************************************************************;

*(a)ohne Gewichtung der Sortenmittelwerte aus Einzelversuchen mit den
Kehrwerten ihrer Fehlervarianzen bei Vorgabe eines Zielgebietes (4) und
eines Transformationsparaketers (phi=0.7);

%lsm(data=dataset,out=ausgabe10a,out_lsm=lsm_ges10a,zielag=4,class= r gr s
jahr ort typ,
/*model=%str(gr*r s(gr*r) jahr*ort(r) typ(jahr*ort*r)),response=y,
dieses Ursprungsmodell ist wegen eindeutiger Schachtelung von s in gr und
ort in r (AG) identisch zu nachfolgendem reduzierten Modell: */
model=%str(gr*r s*r /* jahr*ort standardmäßig auskommentiert, da in
typ(jahr*ort) enthalten*/ typ(jahr*ort*r)),response=y,
ddfm=%str(/ddfm=residual noint notest),
mixstatem=%str(random jahr ort jahr*r jahr*ort /subject=s;),
parmsdata=e.varianzk,phi=0.7);

************************************************************************;

*(b)mit Gewichtung der Sortenmittelwerte aus Einzelversuchen mit den
Kehrwerten ihrer Fehlervarianzen bei Vorgabe eines Zielgebietes (4) und
eines Transformationsparaketers (phi=0.2);

%lsm(data=dataset,out=ausgabe,out_lsm=lsm_ges,zielag=4,class= r gr s jahr
ort typ,
/*model=%str(gr*r s(gr*r) jahr*ort(r) typ(jahr*ort*r)),response=y,
dieses Ursprungsmodell ist wegen eindeutiger Schachtelung von s in gr und
ort in r (AG) identisch zu nachfolgendem reduzierten Modell: */
model=%str(gr*r s*r /* jahr*ort standardmäßig auskommentiert, da in
typ(jahr*ort) enthalten*/ typ(jahr*ort*r)),response=y,
ddfm=%str(/ddfm=residual noint notest),
mixstatem=%str(random jahr ort jahr*r jahr*ort typ(jahr*ort)/subject=s;
repeated;weight w;),
parmsdata=varianzk,phi=0.2);

************************************************************************;
```

* auf Grundlage von Möhring et al. (2004b)

Anlage 9: Output des PIAFStat-Verfahrens ‚MW' (zentrales SAS®-Makro: *%lsm*); Kornertrag sechsjährig Winterweizen 2004-2009 für das Zielgebiet D-Nord*; (Auszug)

Nr	Name	Prüfstatus	Ertrag	SE_{Serie}	W_Ökov.	Ökoregr.	N_zg	N_ges
1	Akratos	alte Sorten	99,8	1,2	4,7		16	56
2	Boomer	alte Sorten	99,6	1,2	2,7		20	57
3	Impression	alte Sorten	99,3	1,3	3,2		19	50
4	Magister	alte Sorten	92,6	1,4	3,8		17	22
5	Schamane	alte Sorten	99,2	1,3	5,0		21	47
6	Skagen	alte Sorten	94,9	1,2	5,1		20	37
7	Skalmeje	alte Sorten	101,4	1,2	3,9		16	48
8	Cubus	LSV5-8	99,8	1,1	4,0		23	69
9	Tommi	LSV5-8	98,9	1,1	5,1	1,1	69	172
10	Dekan	LSV5-8	98,5	1,2	4,0		30	78
11	Akteur	LSV5-8	94,7	1,1	6,1	0,9	38	76
12	Toras	LSV5-8	97,0	1,1	5,5	0,9	25	70
13	Türkis	LSV5-8	98,4	1,1	4,7	0,9	61	144
14	Brilliant	LSV5-8	97,6	1,2	4,5	0,9	30	68
15	Hermann	LSV5-8	102,3	1,1	4,5		35	134
16	Mulan	LSV4	100,4	1,2	4,3		27	73
17	Potenzial	LSV4	102,0	1,1	4,3	0,9	25	57
18	Discus	LSV3	99,5	1,2	4,0	0,9	21	58
19	Esket	LSV3	98,4	1,2	6,2	0,9	28	69
20	Jenga	LSV3	101,6	1,2	3,7		21	57
21	Adler	LSV2	94,3	1,3	5,4		15	25
22	JB Asano	LSV2	103,7	1,3	4,2		22	62
23	Julius	LSV2	101,8	1,4	4,6		22	62
24	Pamier	LSV2	98,2	1,3	4,6		12	26
25	Sophytra	LSV2	100,4	1,5	5,1		11	23
26	Tabasco	LSV2	104,1	1,4	5,9	1,1	19	55
27	Event	LSV1	99,4	1,6			9	17
28	Global	LSV1	104,4	1,7	4,5		9	25
29	Kredo	LSV1	102,4	1,6			9	19
30	Primus	LSV1	104,0	1,7			8	15
31	Agil	OV_WP	93,7	2,0			5	10
32	Elbar	OV_WP	100,1	2,0			5	10
33	Expo	OV_WP	99,6	2,0			5	10
34	KWS Adlon	OV_WP	103,3	2,0			5	10
35	Leo	OV_WP	94,1	2,0			5	10
36	Nathan	OV_WP	101,6	2,0			5	10
37	Ohlin	OV_WP	104,2	2,0			5	10
38	Quintus	OV_WP	100,3	2,0			5	10
39	Rufus	OV_WP	102,8	2,0			5	10
40	Smaragd	OV_WP	103,3	1,8			5	16
41	Tarkus	OV_WP	100,8	1,7			6	14
42	Zappa	OV_WP	102,7	1,8			5	16
43	Mercato	OV_EU	100,2	2,0			4	8
44	Philipp	OV_EU	88,1	1,8			5	9
•••	Anzahl Sorten (Zeilen) insgesamt: 812 *							
	KWS Klausus	sonstige Sorten	98,3	3,2			2	4

Nr. lfd. Nummer, nur für Sorten des definierten Kernsortimentes, nicht für sonstige Sorten

W_Ökov. $= \sqrt{Ökovalenz}$; nur für Sorten mit $N_ges \geq 20$ ausgewiesen

Ökoregr. Ökoregression, nur für Sorten mit signifikanter Abweichung von ‚1' ausgewiesen ($\alpha \leq 5\%$); (methodische Darlegung dieser beiden Parameter war nicht Gegenstand vorliegender Arbeit)

N_zg Anzahl Versuche je Sorte im Zielgebiet

N_ges Anzahl Versuche je Sorte in Ziel- und einbezogenen Nachbargebieten insgesamt

* durch Beschränkung auf Zielgebiet D-Nord (ohne MV-Süd) ergibt sich unter ‚N_ges' eine gegenüber Tab. 10 verringerte Anzahl Versuche je Sorte, zudem verringerte Anzahl Sorten

Erklärung

Ich erkläre hiermit,

- dass keine Zusammenarbeit mit gewerblichen Promotionsberatern stattfand,
- dass ich die dem angestrebten Verfahren zugrunde liegende Promotionsordnung (Amtliches Mitteilungsblatt der Humboldt-Universität zu Berlin Nr. 12/2014 vom 31. März 2014) zur Kenntnis genommen habe,
- dass die Dissertation nicht bereits bei einer anderen wissenschaftlichen Einrichtung eingereicht, angenommen oder abgelehnt wurde,
- dass die Dissertation auf der Grundlage der angegebenen Hilfsmittel und Hilfen selbstständig angefertigt worden ist und
- dass die Grundsätze der Humboldt-Universität zu Berlin zur Sicherung guter wissenschaftlicher Praxis eingehalten wurden.

…… ……

Unterschrift Promovend Ort, Datum

Danksagung

Mein besonderer Dank gilt Frau Prof. Richter für die Betreuung der Dissertation - ich habe sehr viel dabei gelernt.

Herrn Prof. Piepho danke ich für den methodischen Gedankenaustausch und für die vielen wertvollen Anregungen.

Ich danke Herrn Dr. Laidig für die Begutachtung der Dissertationsschrift, insbesondere für die vertiefte Betrachtung der Spezifik des Sortenversuchswesens.

Mein Dank gilt auch allen weiteren Mitgliedern der Promotionskommission - Frau Dr. Kroschewski sowie den Herren Prof. Ellmer, Dr. Meyer, Dr. Jentsch und Dr. Graf. Herrn Dr. Graf danke ich insbesondere auch für die konzeptionelle Zusammenarbeit und die gemeinsame Umsetzung von Innovationen im Bereich der Regionalisierung und der Versuchsserienauswertung, welche mit dieser Dissertation in Zusammenhang stehen.

Meinen Direktoren, zunächst Herrn Dr. Gienapp und nun Herrn Dr. Heilmann, danke ich für den erwiesenen Vertrauensvorschuss und für die Freiräume zur Ideenentwicklung - ich bin überzeugt, dass es sich für die Landesforschungsanstalt Mecklenburg-Vorpommern, für das Sortenversuchswesen und für die Landwirtschaft auszahlt.

Meinen Berufskollegen in den Fachrichtungen Biometrie, Sorten- und Versuchswesen und in der PIAF-Entwicklung - in der Landesforschungsanstalt, in allen Bundesländern und in verschiedensten Organisationen - danke ich für die umfangreiche fachliche Zusammenarbeit sowie für die Datengewinnung und den Datenaustausch, ohne die diese Dissertation und die vorgestellte Methodik nicht möglich wäre. Ganz besonders danke ich Frau Dr. Zenk für die Unterstützung in der Datenbankarbeit und SAS-Programmierung.

Nicht zuletzt denke ich an meine Eltern (†) und danke meinen Kindern, Brüdern und Freunden für die freudige Anteilnahme.