

Einbindung neuer Methoden zur Routineauswertung von landwirtschaftlichen Versuchen mit Hilfe von SAS-Macros

Andrea Zenk
Volker Michel

Landesforschungsanstalt für
Landwirtschaft u. Fischerei M-V
Dorfplatz 1
Gülzow
a.zenk@lfa.mvnet.de
v.michel@lfa.mvnet.de

Jens Möhring

Institut für Pflanzenbau und Grünland,
FB Bioinformatik, Universität Hohenheim
Stuttgart
moehring@uni-hohenheim.de

Zusammenfassung

Mit **PIAF**¹ (Planung- Information- Auswertung von Feldversuchen) haben sich die Bundesländer mit Unterstützung des Bundes ein leistungsstarkes und komfortables Programm zur Erfassung und Archivierung von Daten aus landwirtschaftlichen Feldversuchen geschaffen. Über eine Schnittstelle können diese Daten an **PIAFStat**² (Programm zur Datenanalyse und statistischen Auswertung von Einzelversuchen und Versuchsserien mittels SAS) übergeben werden.

Beide Programme werden u.a. im Sortenversuchswesen aller Bundesländer eingesetzt. Die Auswertung der Landessortenversuche von Mecklenburg bis Bayern wird mit diesen Programmen durchgeführt.

Bisherige SAS-Programme für die Auswertung von Versuchen und Versuchsserien in PIAFStat wurden von Mitarbeitern einiger Länderdienstanstalten selbst entwickelt und allen Kollegen für die PIAFStat-Verfahrensbibliothek zur Verfügung gestellt.

Die derzeitige Neugestaltung im Sortenversuchssystem insbesondere für die mehrjährige Auswertung von Sortenversuchen erfordert neue SAS-Lösungen. Diese beinhalten u.a.:

- Gemischte Modelle mit den Faktoren Sorte, Jahr, Ort, Versuchstyp, Sortengruppe, Gewichtung ($1/SE^2$)
- Einbeziehung benachbarter Regionen für die regionale Sortenempfehlung mit geringerer Gewichtung als die Standorte des Zielgebietes

¹ PIAF wurde im Auftrage der Länder und des Bundes mit der ProPlant Gesellschaft für Agrar- und Umweltinformatik mbH entwickelt

² PIAFStat wurde im Auftrage der Länder mit der Firma BioMath GmbH, Gesellschaft für Angewandte Mathematische Statistik in Biologie und Medizin entwickelt.

Beide Programmsysteme werden in Gemeinschaftsarbeit von Bund und Ländern in ihrer Pflege und Weiterentwicklung begleitet.

- Auswertung unorthogonaler Serien durch Einbeziehung aller vergleichbaren Prüfungen (Wertprüfungen des BSA, EU-Prüfungen der SFG).

Diese Lösungen wurden in Zusammenarbeit mit der Universität Hohenheim mit den vorliegenden SAS-Macro's:

%Varianzkomponenten; %Auswertung; %lsm; %Gewichte und *%Boxcox* geschaffen. Die Verwendung der SAS-Macro's in Verfahren von PIAFStat und damit die Nutzbarmachung dieser Macro's für jeden Anwender, auch ohne SAS-Kenntnisse, liegt wieder in der Hand der Mitarbeiter der Länderdienststellen selbst.

Zur Ernte 2005 werden wir in der Landesforschungsanstalt Mecklenburg-Vorpommern mit diesen neuen Verfahren mehrjährige Auswertungen durchführen.

Schlüsselworte: PIAFStat, Proc MIXED, ‚Hohenheim-Gülzower Serienauswertung‘, Makro, SAS, Gemischte Modelle

1 Hintergrund

1.1 Modernisierung des Sortenversuchswesens in Deutschland

Zurzeit vollzieht sich in Deutschland eine Modernisierung des Sortenversuchswesens. Die Agrarministerkonferenz vom 7.10.2004 hat dabei den Auftrag für ein durchgängiges und kosteneffizientes Sortenversuchssystem bekräftigt. Die kombinierte Auswertung aller Daten aus Wertprüfungen und Landessortenversuchen auf Basis der ‚Hohenheimer Methodik‘ sichert die optimale und schnelle Nutzung der Daten für die regionale Sortenberatung. Die Effizienz der Auswertung wird durch Einbeziehung benachbarter Anbaugelände mit einer optimal abgestuften Wichtung erhöht. [1]

1.2 Entwicklung der Serienverrechnung in Mecklenburg-Vorpommern

In Mecklenburg-Vorpommern wurde im Jahr 1998 begonnen, die Serienverrechnung der Landessortenversuche mit Hilfe von SAS durchzuführen. Die Nutzung der Proc GLM ermöglichte es, unbalancierte Daten auszuwerten. So konnten in die Verrechnungen der LSV gleichartige Sortenversuche wie Wertprüfungen und EU-Sortenversuche einbezogen und die Anzahl der Versuchsjahre von 3 auf 5 bis 6 erhöht werden. [2]

Mit Umstieg auf die Proc Mixed im Jahr 2000 wurden zudem unter anderem die Sorte x Jahr – Interaktion, die Sorte x Standort- Interaktion, die Gewichtung nach Versuchspräzision und die Prüfung einzelner Sorten am gleichen Ort im gleichen Jahr optimal im Modell berücksichtigt. [3]

Eine weitere Optimierung der Verrechnung wird in diesem Jahr durch die in Zusammenarbeit der Universität Hohenheim und der Landesforschungsanstalt für Landwirtschaft und Fischerei Mecklenburg-Vorpommern entwickelten neuen Modelle erreicht. Erstmals wird eine optimal abgestufte Gewichtung von Standorten benachbarter Anbaugelände im Modell berücksichtigt. Neu ist ebenfalls die Möglichkeit der Transformation nach Bestimmung des optimalen Parameters ϕ . [4] [5] [6]

2 Voraussetzungen

In den Länderdienststellen (LDS), die die Landessortenversuche anlegen und auswerten, wird für die Planung, Erfassung und Archivierung der Versuchsdaten das Programm **PIAF** genutzt. Die Kompatibilität der Versuchsdaten, einheitliche Stammdaten und definierte Schnittstellen ermöglichen unkompliziert den Datenaustausch.

Als Standard-Software für statistische Auswertungen haben die Länder **SAS** gewählt. Mit einer vertraglich vereinbarten Gemeinschaftslizenz haben die PIAF-Nutzer sehr günstige preisliche Konditionen, ohne die der umfangreiche Einsatz von SAS in den LDS nicht finanzierbar wäre. Als Bindeglied zwischen PIAF und SAS wurde ein weiteres Programm entwickelt – **PIAFStat**. PIAFStat stellt eine anwenderfreundliche Benutzeroberfläche für SAS dar, die es PIAF-Anwendern ermöglicht, auch ohne SAS-Kenntnisse in einfacher und flexibler Weise SAS-Programme zu nutzen. [7]

PIAFStat stellt über eine Verfahrensbibliothek die Verbindung zwischen den Daten (z.B. aus PIAF) und SAS her. Die einzelnen Verfahren der Bibliothek sind SAS-Programme, erweitert um syntaktische Elemente. Diese syntaktischen Elemente von PIAFStat ermöglichen die Entwicklung von Verfahren unabhängig von den konkreten Daten. Erst durch die Zuordnung der Daten zu einem Verfahren wird durch die Bindung der Datenvariablen mit den syntaktischen Elementen ein lauffähiges SAS-Programm generiert. LIST- und LOG-Fenster von PIAFStat zeigen Output und Log-Informationen von SAS. [8]

3 PIAFStat-Verfahren

In PIAFStat sind von einigen LDS zahlreiche Verfahren entwickelt worden. So entstanden unter anderem in Brandenburg und Mecklenburg-Vorpommern Verfahren zur varianzanalytischen Auswertung von Einzelversuchen und Serien, in Bayern Verfahren zur Plausibilitätsprüfung und in Baden-Württemberg Verfahren zur grafi-

schen Darstellung von Versuchsergebnissen im Lageplan. Eine Auswahl dieser Verfahren ist im Internet allen PIAFStat-Anwendern zur Verfügung gestellt worden. Zu diesen Verfahren kommt nun eine Verfahrensgruppe hinzu, die die ‚Hohenheim-Gülzower Serienauswertung‘ beinhaltet. Dazu wurden die von Jens Möhring geschriebenen SAS-Macro's (Hohenheimer Methode) in PIAFStat-Verfahren eingebunden. [9]

Tabelle 1: PIAFStat-Verfahren zur ‚Hohenheim-Gülzower Serienauswertung‘

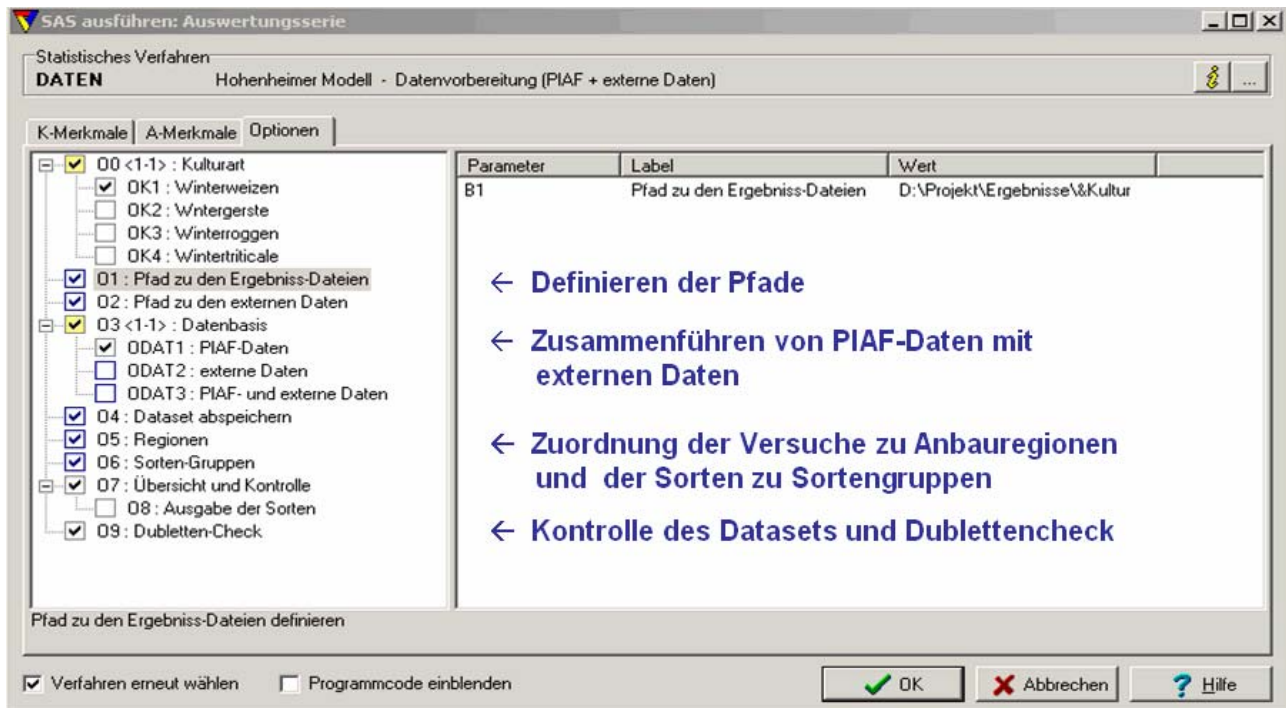
| PIAFStat-Verfahren | | enthaltene SAS-Macro's |
|--------------------|--|------------------------|
| Name | Label | |
| DATEN | Datenvorbereitung (PIAF + externe Daten) | %zuordnung_typ |
| VK | Bestimmung der Varianzkomponenten | %varianzkomponenten |
| MW | Bestimmung der Mittelwerte | %auswertung und %lsm |
| PHI | optimaler Transformationsparameter | %boxcox |
| GEW | Bestimmung der Gewichte | %gewichte |

Insbesondere auf die Verfahren DATEN, VK und MW soll kurz eingegangen werden.

3.1 DATEN – Datenvorbereitung (PIAF + externe Daten)

Ziel des Verfahrens DATEN ist die Vorbereitung des Datasets. Dabei kann der Anwender interaktiv Klassifikations-Merkmale und Analyse-Merkmale zuordnen und verschiedene Funktionen des Verfahrens aktivieren bzw. deaktivieren. Folgende Abbildung zeigt die optionalen Möglichkeiten des Verfahrens auf.

Abbildung 1: PIAFStat-Verfahren Daten



Im Verfahren ist es dem Anwender möglich, seine Pfade zu den Ergebnis-Dateien und den externen Daten zu definieren. Er kann die Datenbasis für die nachfolgenden Berechnungen bestimmen. Erstmals ist es dabei möglich, PIAF-Daten und externe Daten, die in einer definierten Struktur in Excel vorliegen, zusammenzufügen. Bei den Optionen Regionen und Sortengruppen werden dem Dataset externe Stammdaten zu den Anbaugebieten und den Sortengruppen zugefügt. Werden diese Optionen deaktiviert, erfolgt automatisch die Einrichtung eines Anbaugebietes ‚100‘ und einer Sortengruppe ‚alle‘ für alle Datensätze des Datasets. Nach diesen Schritten werden zur Plausibilitätskontrolle Tabellen (Tab. 2 und 3) erzeugt, die das Überprüfen des Datasets vereinfachen sollen. Insbesondere erfolgt ein Dubletten-Check. Mit diesem Check können Versuche erkannt werden, die im Dataset z.B. durch das Zusammenführen von PIAF- und externen Daten doppelt vorhanden sind.

Tabelle 2: Anzahl Versuche je Jahr und Ort (Auszug)

| | | | Anzahl Versuche | | | | | |
|-----|------|-------------|-----------------|------|------|------|------|------|
| | | | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
| AG | Land | | | | | | | |
| 4 | 11 | Biestow | 1 | 1 | 1 | 1 | 1 | 1 |
| | | Gülzow | 2 | 2 | 2 | 2 | 1 | 1 |
| | | Köchelstorf | 1 | . | 1 | 1 | 1 | 1 |
| | | Tützpatz | 3 | 3 | 4 | 4 | 3 | 3 |
| | 12 | Prenzlau | 3 | 3 | 3 | 3 | 3 | . |
| 5 | | Vipperow | 1 | . | 1 | 1 | 1 | 1 |
| | 12 | Badingen | . | 4 | 2 | 4 | 3 | 1 |
| | | Berge | . | . | . | . | 1 | 1 |
| ... | | | | | | | | |

Tabelle 3: Übersicht zu den Class- Merkmalen

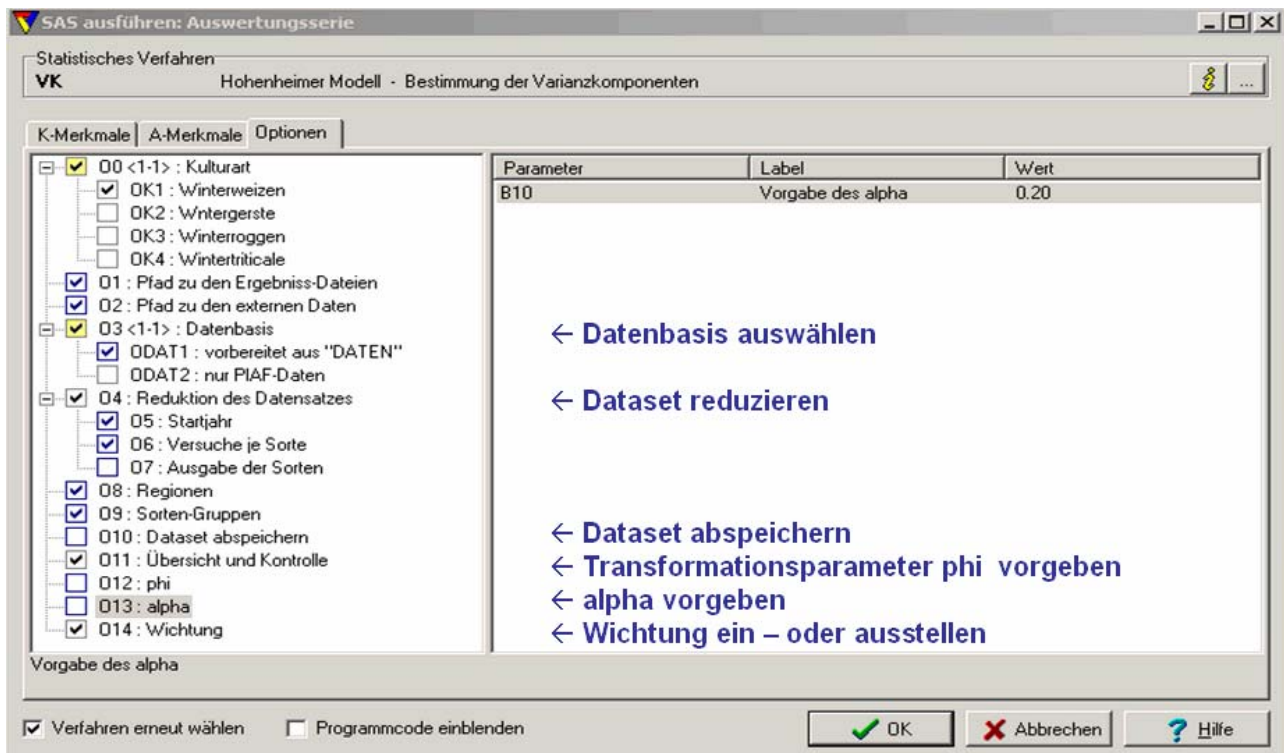
| Class-Merkmal | | Anzahl |
|---------------|-----------------------------|--------|
| Jahr | 99 2000 2001 2002 2003 2004 | 6 |
| Land | ... | 4 |
| Ort | ... | 26 |
| gr | A B BH C CH E K | 7 |
| r | 2 3 4 5 | 4 |
| s | ... | 127 |

3.2 VK – Bestimmung der Varianzkomponenten

Mit dem Verfahren VK werden die zur Auswertung (Mittelwert-Bestimmung) benötigten Varianzkomponenten geschätzt. Die gesonderte Ermittlung der Varianzkomponenten ermöglicht es dem Anwender, diese Schätzung zu einem Zeitpunkt durchzuführen, der der von Zeitdruck geprägten Ernte- und Auswertungs-Phase vorausgeht. Als vorteilhaft wird zudem angesehen, wenn das für diese Schätzung verwendete Dataset eine Vielzahl von Jahren (mehr als 6) und Sorten beinhaltet. Trotzdem hat der Anwender die Möglichkeit, die Datenbasis hinsichtlich Anzahl einzubeziehender Jahre und Sorten zu reduzieren. Das empfiehlt sich z.B. bei Testläufen, um die Rechenzeit zu reduzieren.

Im Verfahren VK können der Transformationsparameter phi (default=1), alpha für das Vertrauensintervall der Varianzkomponenten (default=0,20) und Wichtung nach Versuchspräzision (default=ON) vom Anwender vorgegeben werden.

Abbildung 2: PIAFStat-Verfahren VK



Im Ergebnis des Verfahrens wird im Output die Tabelle mit den Varianzkomponenten und deren Vertrauensintervallen (Tab. 4) gezeigt. Die mit UN(x,x) bezeichneten Parameter stellen die Anbauggebiete dar. Zusätzlich sind in diese Tabelle die genetischen Korrelationen zwischen den Anbaugebieten in den mit UN(x,y) bezeichneten Zeilen ausgewiesen. Die Tabelle wird als SAS-Datei und in Excel abgespeichert. In unserem Beispiel zeigt sich z.B. eine sehr hohe genetische Korrelation zwischen Anbaugebiet 4 (Lehmige Ackerbauggebiete im Ostseeküsten- und Übergangsklima) und Anbaugebiet 5 (Sandige Ackerbauggebiete des nordostdeutschen Binnentieflandes).

Tabelle 4: Output der Varianzkomponenten

| CovParm | Subject | Estimate | Alpha | Lower | Upper | r_gen |
|-----------|---------|----------|-------|---------|--------|--------|
| Intercept | s | 2.6873 | 0.2 | 1.6338 | 5.922 | |
| Jahr | s | 5.3259 | 0.2 | 4.2362 | 7.016 | |
| ort | s | 0.1819 | 0.2 | 0.0602 | 6740.8 | |
| Jahr*r | s | 1.4473 | 0.2 | 0.8582 | 3.404 | |
| Jahr*ort | s | 0 | | . | | |
| UN(2,2) | s | 6.5182 | 0.2 | 2.4507 | 305.3 | |
| UN(3,2) | s | 0 | | . | | 0.3785 |
| UN(3,3) | s | 2.7900 | 0.2 | 1.2043 | 27.741 | |
| UN(4,2) | s | 0 | | . | | 0.4979 |
| UN(4,3) | s | 0 | | . | | 0.6455 |
| UN(4,4) | s | 0.4772 | 0.2 | 0.1581 | 4095.5 | |
| UN(5,2) | s | 0 | | . | | 0.5403 |
| UN(5,3) | s | 0 | | . | | 0.7005 |
| UN(5,4) | s | 0 | | . | | 0.9215 |
| UN(5,5) | s | 1.44E-16 | | . | | |
| Residual | | 21.0998 | 0.2 | 20.0936 | 22.197 | |

3.3 MW – Bestimmung der Mittelwerte

Mit dem Verfahren MW wird das eigentliche Ziel aller Berechnungen der ‚Hohenheim-Gülzower Serienauswertung‘ erreicht. Es werden für das Zielgebiet die mehrjährigen Sorten-Mittelwerte errechnet. Welches Anbaugelände das Zielgebiet der jeweiligen Verrechnung sein soll, kann im Verfahren vorgegeben werden. Als Ergebnis-Tabellen (Tab. 5) werden die mehrjährigen Sorten-Mittelwerte in verschiedenen Sortier-Reihenfolgen angeboten:

1. Sortierung innerhalb der Sortengruppen alphabetisch nach Sorten-Namen
2. Sortierung generell nach Bedeutung (nach kleinstem Standardfehler se).

Als Zusatzinformation wird eine Übersichtstabelle zu den Mittelwerten der Sortengruppen in den Anbaugeländen (Tab. 6) ausgegeben.

Alle Tabellen können in Excel abgespeichert werden.

Abbildung 3: PIAFStat-Verfahren Daten

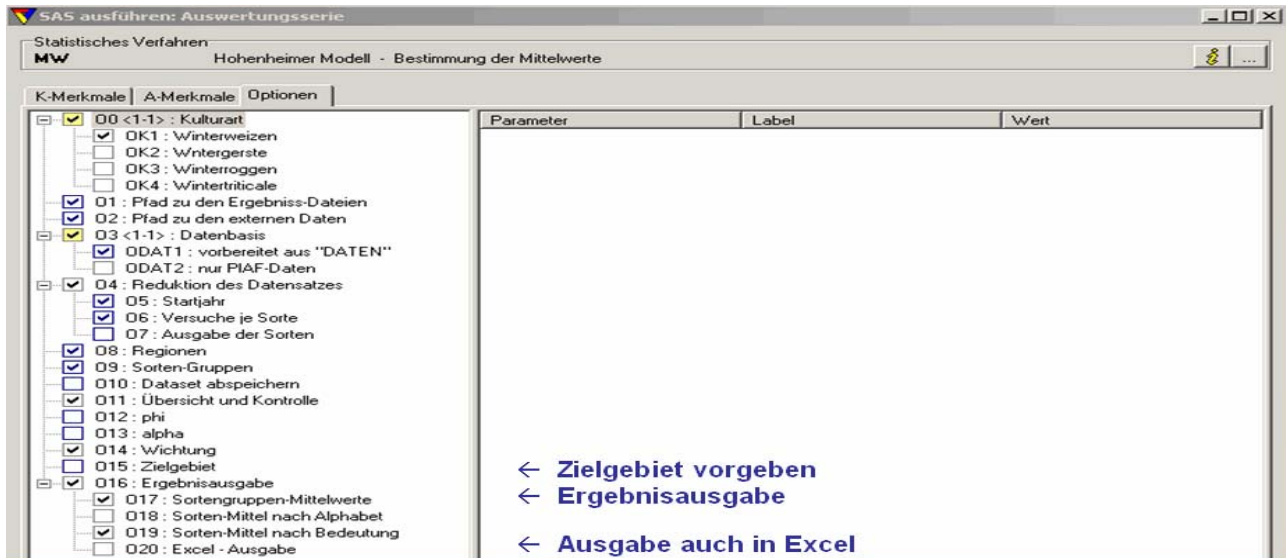


Tabelle 5: Mittelwerte der Sorten im Zielgebiet 4 (Auszug)

| Sorten | MW | Effekt | se | Versuche | |
|------------|----------|--------|------|----------|----|
| A Akratos | WW 03046 | 100.5 | -2.2 | 3.3 | 15 |
| Atoll | WW 02774 | 100.0 | -2.7 | 2.0 | 43 |
| Batis | WW 01968 | 96.4 | -6.3 | 1.6 | 84 |
| Türkis | WW 02991 | 100.8 | -1.9 | 3.3 | 17 |
| B Buteo | WW 03069 | 103.0 | 1.5 | 3.1 | 19 |
| Versailles | WW 02239 | 103.3 | 1.8 | 3.1 | 12 |
| ... | | | | | |

Tabelle 6: Mittelwerte der Sortengruppen in Anbaugeländen

| Sortengruppen | Anbaugelände | | | |
|---------------|--------------|-------|-------|------|
| | 2 | 3 | 4 | 5 |
| A | 82.1 | 101.0 | 97.1 | 67.7 |
| B | 84.6 | 101.4 | 99.9 | 69.9 |
| BH | 87.2 | 103.7 | 105.0 | 75.8 |
| C | 85.0 | 104.6 | 102.4 | 72.6 |
| CH | 89.6 | 98.4 | 107.4 | 74.5 |
| E | . | 89.8 | 90.4 | 63.0 |

3.4 Lösungsbeispiele in SAS

Es soll kurz auf einige wenige Lösungsansätze in SAS bei der Verfahrens- Erstellung in PIAFStat eingegangen werden.

globale Macrovariablen:

Als sehr wichtig hat sich die Arbeit mit globalen Macrovariablen erwiesen. Beispiel: Definition der Pfade für die externen Daten:

```
%global PEXT; %let PEXT=[BEXTERN]; libname d "&PEXT";
```

[BEXTERN] stellt dabei ein Element aus PIAFStat dar, das den Pfad (z.B. D:\temp) verschlüsselt.

Sortengruppen-Definition:

Nicht jede Kulturart wird in Sortengruppen unterteilt (Hafer, Erbsen). Deshalb wird dem Dataset generell eine Sortengruppe (=alle) zugefügt:

```
Data dataset; set dataset; gr='alle'; run;
```

Gibt es aber externe Stammdaten mit Sortengruppen, wird die Sortengruppe ,alle' wieder aus dem Dataset entfernt:

```
Data dataset; set dataset; drop gr; run;
```

Dann erfolgt der Import der Stammdatei:

```
proc import datafile="&PEXT\[BREG]"  
out= gruppen  
dbms=excel2000  
replace; sheet="Tabelle1"; getnames=yes; run;
```

und das Zusammenfügen von Stammdatei und Dataset:

```
data neu; merge dataset gruppen; by s; run;
```

Tabellierung der Class-Merkmale:

Um die Class-Merkmale übersichtlich tabellieren zu können, wurde als erstes eine einfache Proc GLM mit Ausgabe der SAS-Datei ClassLevels durchgeführt:

```
proc glm data=Dataset ;  
class s gr jahr land ortbez R ;  
model y=;  
ods output ClassLevels=classlev; run;
```

Als nächstes wurde die Variable values z.B. bei s (Sorten) mit ,...' überschrieben:

```
data classlev; set classlev;  
if class='s' then values='... '; run;
```

Im nächsten Schritt konnte in der Proc Tabulate auf die SAS-Datei ClassLevels zugegriffen und eine übersichtliche Darstellung der Class-Merkmale erreicht werden.

```
Proc Tabulate data=classlev; class class values; var levels ;  
table class=' '* values=' ',levels='Anzahl'*max=' '*F=25.0  
/RTS=50 BOX='Class-Merkmal'; run;
```

Dubletten-Check

Für die Kontrolle, ob nicht zufällig ein Versuch im Dataset doppelt vorkommt, wurde das Macro *%MULTKEY* auf die vorliegende Datenstruktur angepasst. Dieses Macro ist im SAS-Anwenderhandbuch im Internet zu finden und wurde 1999 von Heinrich Stürzl entwickelt.

4 Ausblick

Mit der Erstellung der vorgestellten PIAFStat-Verfahren zur ‚Hohenheim-Gülzower Serienauswertung‘ sind die Arbeiten nicht abgeschlossen. In Zukunft wird es vorrangig darum gehen, den Output der Verfahren zu verbessern und zu erweitern. So ist es geplant, Diagramm-Erstellung, mehrjährige und einjährige Tabellierung für die Berichterstellung in SAS zu realisieren. Zudem wird das Verfahren DATEN um die Möglichkeit erweitert werden, mehrere externe Datensätze miteinander zu verknüpfen.

In Mecklenburg-Vorpommern werden zur Ernte 2005 bei ausgewählten Kulturarten (Winterweizen, Silomais, Wirtschaftskartoffeln) diese PIAFStat-Verfahren in praxi getestet und eingeführt.

Dies alles sehen wir als Voraussetzung an, um das von der Agrarministerkonferenz vorgegebene Ziel, die Einführung dieser Methode bundesweit zur Ernte 2006 zu vollziehen, erreichen zu können.

Literatur

- [1] Möhring, J., A. BÜchse, H. Piepho, V. Michel, J. Rath, F. Laidig.. 2004: Gesundsparen ohne Nachteile. DLG-Mitteilungen 6 (2004), 22-23
- [2] Michel, V., Neuster methodischer Stand bei der Versuchsauswertung in Mecklenburg-Vorpommern; VII. Rapskolloquium, Schleswig-Holstein/Mecklenburg-Vorpommern, am 4. und 5.12.2003 in BÜdelsdorf
- [3] Piepho, H.-P., V. Michel, 2001: Überlegungen zur regionalen Auswertung von Landessortenversuchen. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 31/4, 123-139
- [4] Möhring, J., A. BÜchse, H. Piepho., 2003: Bericht Prüfsystem Winterweizen 24.12.2003
- [5] Möhring, J., A. BÜchse, H. Piepho., 2004: Prüfsystem Winterweizen 19.03.2004
- [6] Möhring, J., A. BÜchse, H. Piepho., 2004: Prüfsystem Winterweizen 24.08.2004
- [7] Schmidtke, J., Voit, K., 2001: PIAFStat – Datenanalyse für landwirtschaftliche Feldversuche. Proceedings der 5. Konferenz der SAS-Anwender in Forschung und Entwicklung, 317-321
- [8] BioMath GmbH: 2004: PIAFStat – Handbuch, Version 4.01, 2004
- [9] Michel, V., 2004: Anmerkungen zum Prüfsystem Winterweizen 24.08.2004