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

...

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 $\times$ 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 Anbauggebiete 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 Sortenspezifität.

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 $S \times J \times O$ - 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 $\times$ 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 $\times$ 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 $V \times Y \times L$ - 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 $\times$ 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.