Methodische Probleme der sekundären Auswertung von Versuchs- und Erhebungsdaten für die Evaluierung genetischer Ressourcen

KLAUS STEGEMANN, HALLE / SAALE
VOLKER WEISE, HALLE / SAALE

Abstract

To develop information system for evaluation genetic resources in plant breeding it is necessary to describe the nature and specific accession by statistical measurement. Solution and algoritms for analysing the existing data from trials and surveys are describt.

1 Einführung und Problemstellung

Seit Mai 1996 koordiniert das Informationszentrum für Genetische Ressourcen (IGR) der Zentralstelle für Agrardokumentation und -information (ZADI) das vom Bundesministerium für Landwirtschaft finanzierte Verbundprojekt zum Aufbau eines Informationssystems für die Evaluierung pflanzengenetischer Ressourcen in der Bundesrepublik Deutschland (EVA).
Die zu evaluierenden pflanzengenetischen Ressourcen sind Wild- und Kulturpflanzen der Genbanken. Deren Nutzbarkeit für die Pflanzenzüchtung wird wesentlich durch die Möglichkeiten bestimmt, sie zu evaluieren, ihren Wert für die Züchtung einzuschätzen.
Die hierfür nutzbaren Daten wurden zu unterschiedlichen Terminen und an verschiedenen Standorten erfaßt. Sie stammen einerseits aus Erhebungen und andererseits aus Versuchen. Damit fehlen meist die vergleichbaren Bedingungen für die Prüfung von Hypothesen. Die Datenqualifizierung erfordert deshalb datenorientierte Betrachtungs- und Vorgehensweisen. Es können keine konfirmatorischen Analysen vorgenommen werden. Hinzu kommt, daß die einzelnen Nutzer sehr unterschiedliche Fragestellungen beantwortet haben wollen.
Es wurde die Aufgabe gestellt, diese heterogenen und nicht orthogonalen Datenbestände durch die Berechnung von statistischen Maßzahlen so zu beschreiben und zu verdichten, daß dadurch eine zentrale Metadatei aufgebaut werden kann, welche komprimierte Informationen für die Evaluierung genetischer Ressourcen bereitstellt.
In dieser Metadatei könnte dann durch die Vorgabe von Schwellenwerten für die zu recherchierenden statistischen Maßzahlen, die der Züchter vorgibt, nach solchen genetischen Ressourcen recherchiert werden, die gewünschte Eigenschaften aufweisen. Sie wäre damit ein außerordentlich wichtiges Hilfsmittel zur Auswahl von Genotypen aus den vorhandenen genetischen Ressourcen , um diese für die Pflanzenzüchtung zu nutzen.
Zur Erreichung dieses Ziels war es notwendig, die für die Durchführung von datenorientierten Analysen gebräuchlichen Verfahren, z. B. Heiler und Michels (1994), hinsichtlich ihrer Eignung zur Lösung dieser Aufgabenstellung zu untersuchen und anzupassen. Die Bearbeitung der spezifischen Fragestellungen erforderte es, die Verfahren zu erweitern oder sogar erst zu entwickeln. Neue Entwicklungen wurden insbesondere dann erforderlich, wenn

bei der Auswertung von Versuchsergebnissen verwendete Algorithmen für datenorientierte Auswertungen zu modifizieren und anzupassen waren.

2 Material und Methoden

Die in die Metadatei zu integrierenden Merkmalswerte sind einerseits durch die Genbanken bei den routinemäßig durchzuführenden Arbeiten für die Erhaltung und Vermehrung der genetischen Ressourcen ermittelt worden, und sie stammen andererseits aus Sortenprüfungen, insbesondere Landessortenversuchen.
Erhebungen der Genbanken werden jeweils nur an einem Ort durchgeführt. Der Anbau der einzelnen Accessionen erfolgt mit mehrjährigen Abständen. Eine Sorte wird an mehreren Orten und in mehreren unmittelbar aufeinanderfolgenden Jahren geprüft.
Die Merkmalswerte der einzelnen Accessionen sind auf Grund der räumlichen und zeitlichen Unterschiede in der Regel nicht direkt vergleichbar Die Durchführung von Evaluierungen erfordert es aber, Vergleiche anzustellen. Um das zu ermöglichen, war es erforderlich mit einem Vergleichsmaßstab zu arbeiten.

2.1 Schaffung von Vergleichsbasen

Ohne Vergleichsbasis sind die Merkmalsdaten kaum auswertbar, weil z. B. beim Merkmal Ertrag das Ertragsniveau in den einzelnen Jahren unterschiedlich ist. Auch infolge der in den letzten Jahrzehnten sehr angestiegenen Erträge ist es zwingend erforderlich, Evaluierungen unabhängig von der absoluten Ertragshöhe vornehmen zu können.
Zur Frage der Schätzung eines Merkmalswertes für die Vergleichsbasis führte schon Wricke (1965) aus, "daß der Versuchsdurchschnitt der beste Schätzwert für die Ertragsfähigkeit der speziellen Umweltbedingungen, unter denen die Prüfglieder gestanden haben, ist. Es ist nicht möglich, die Ertragsfähigkeit eines Versuchsfeldes von vornherein abzuschätzen und diesen Schätzwert für die Berechnungen zu nutzen, so wie es manchmal gefordert wird" (S. 292).
In Sortenversuchen bildete der Mittelwert aller Sorten, die an einem Ort und in einem Jahr auf der niedrigsten Intensitätsstufe geprüft wurden, für die Auswertung der quantitativen Prüfmerkmalswerte die Vergleichsbasis. Für durch die Genbanken ermittelten Daten wurde gleichfalls der aus allen Accessionen für ein Merkmal berechnete Mittelwert als Vergleichsbasis benutzt.
Bei den qualitativen Merkmalen dienten zwei Werte als Vergleichsbasis, nämlich das Maximum und das Minimum. Die Extremwerte charakterisieren beispielsweise einen Krankheitsbefall wesentlich besser als der Mittelwert.

2.2 Statistische Maßzahlen zur Beschreibung einer Sorte und der Vergleichsbasis

Für jede Sorte, die in mehreren Versuchen in einem Jahr vorkommt, wurden je Jahr Aggregationen über die Orte vorgenommen und statistische Maßzahlen berechnet. Die Berechnungen erfolgten jeweils getrennt für die Daten der Sorte und der Vergleichsbasis. Folgende Maßzahlen wurden hierbei ermittelt:
- der arithmetische Mittelwert,
- das Maximum,
- das Minimum und
- der Variationskoeffizient.

2.3 Berechnung statistischer Maßzahlen durch gemeinsame Auswertungen der Prüfgliedmittelwerte einer Sorte und der dazugehörigen Daten der Vergleichsbasis

Durch die gemeinsame Auswertung der Daten der einzelnen Sorte und der dazugehörigen Vergleichsbasis konnten für jede Sorte, die in einem Jahr in mehreren Versuchen vorkommt, weitere statistische Maßzahlen abgeleitet werden. Sie ermöglichen Einschätzungen von:
- Differenzen zwischen Merkmalswerten einer Sorte und der Vergleichsbasis,
- Ökovalenzen und
- Zusammenhängen durch statistische Maßzahlen der Ökoregression.

Die Differenzen zwischen den Merkmalswerten der Sorte und der Vergleichsbasis wurden als absolute und relative Werte berechnet. Diese Differenzen verdeutlichen, ob sich eine Sorte von den übrigen unterscheidet. Um erkennen zu können, ob eine mittlere Differenz als bedeutsam anzusehen ist, erfolgten Prüfungen mit der Differenzmethode und die dabei ermittelte Irrtumswahrscheinlichkeit (P %) wird gespeichert.
Je Sorte wurde weiterhin die Ökovalenz (W_i) nach Wricke (1962, 1964 und 1965) berechnet. Dieser Wert beinhaltet summierte quadrierte Differenzen, die nur in ihrer Rangfolge versuchsbezogen interpretierbar sind. Sie stellen keine sortenspezifischen Maßzahlen dar. Deshalb wurde eine sortenspezifische Ökovalenz (W_i%) berechnet. Sie ergibt sich durch:

$$W_i\% = \frac{W_i}{SQ_i + SQ_{VB}} 100 \qquad (1)$$

Der W_i-Wert wird also durch die Summe des SQ-Wertes der Sorte und des dazugehörigen SQ-Wertes der Vergleichsbasis dividiert.
Bei positiven Kovarianzen liegen die W_i%-Werte zwischen 0 und 100 und bei negativen übersteigen sie den Wert 100. Je kleiner die Werte von W_i% sind, um so höher ist die Ökovalenz des Genotyps. Wird sie Null, ist die Variabiltät der Merkmalsdaten einer Sorte übereinstimmend mit der Merkmalsvariabilität der Vergleichsbasis also der Ortsmittelwerte.
Große Werte von W_i % zeigen an, daß die Merkmalsvariabilität des Genotyps und der Vergleichsbasis unterschiedlich sind. Die Reaktionen des Genotyps auf wechselnde Umweltbedingungen weichen von denen des Sortiments (Sortimentsmittelwert) ab.
Eine Ursache für das Vorhandensein einer großen Interaktion zwischen einer Sorte und den Umwelten kann darin bestehen, daß die Sorte ein anderes Ertragsverhalten als das Sortiment besitzt. In diesem Fall hat die Sorte entweder auf niedrigen Ertragsniveaus niedrigere und auf hohen höhere Erträge oder auf niedrigen Ertragsniveaus höhere und auf den hohen niedrigere Erträge. Im ersteren Fall wäre sie eine Intensiv- und im zweiten eine Extensivsorte. Um das deutlich erkennen zu können, wurden Ökoregressionen durchgeführt und die benötigten Maßzahlen zur Ableitung von Entscheidungen werden damit bereitgestellt.
Der **lineare Regressionskoeffizient** b_{1_i} ergibt sich nach Yates und Cochran (1938) und Finlay und Wilkinson (1963) durch:

$$b_{1_i} = \frac{\sum_j \left(y_{ij} \left[\overline{y}_{VBj} - \overline{y}_{VB.} \right] \right)}{\sum_j \left(y_{VBj} - y_{VB.} \right)^2} \qquad (2)$$

Zum besseren Vergleich der Sorte mit den Ortsmittelwerten (Mittelwerten der Vergleichsbasis) empfehlen Stegemann u. a. (1995) anstelle von b_{1_i} besser $b_{1_i} - 1$ zu verwenden. Ist dieser Wert wesentlich kleiner als Null kann die Sorte als Extensivtyp, ist er wesentlich größer als Null kann sie als Intensivtyp charakterisiert werden.

2.4 Berechnung von Lageparametern für Boniturnoten

Die Boniturnote einer Sorte kann mit dem Minimum des Sortiments oder mit seinem Maximum übereinstimmen oder zwischen beiden Werten liegen. Es bietet sich an, einen Lageparameter je Einzelversuch zu berechnen:

$$L_i = \frac{Bon_i - Min}{Max - Min} \qquad (3)$$

Ist die Boniturnote einer Sorte übereinstimmend mit dem Minimum beträgt der Wert Null, stimmt sie mit dem Maximum überein, ergibt sich die Zahl 1. Der Lageparameter kann also nur Werte zwischen 0 und 1 annehmen.
Für Sorten, mit Boniturnoten an mehreren Orten innerhalb eines Jahres, wurden außerdem Lageparameter berechnet, bei denen als Maximum der größte und als Minimum der kleinste Wert aller Orte verwendet wurden.

2.5 Relative Spannweite der Merkmalsdaten

Die Merkmalswerte einer Sorte variieren je Jahr ortsabhängig. Wurde eine Sorte in einem Jahr an wenigen Orten geprüft, können die Bedingungen relativ homogen sein oder sehr verschieden. Das schlägt sich in einzelnen statistischen Maßzahlen nieder. Um das erkennen zu können wurde die Spannweite aller Werte eines Merkmals einer Sorte je Jahr ermittelt und diese zur Relativierung durch das Minimum dividiert. Je größer dieser Wert ist, um so extremer waren die Bedingungen.

3. Recherchemenues

Die gespeicherten statistischen Maßzahlen können über Menues recherchiert werden Die einzelnen Maßzahlen sind hierbei inhaltlich beschrieben und vollständig oder selektiv recherchierbar. Die selektive Ausgabe wird durch die Vorgabe von Schwellenwerten erreicht.

4. Stand der Arbeiten

Die Programme zum Datenmanagement und zur Berechnung der statistischen Maßzahlen sind weitgehend fertiggestellt. Auswertungsarbeiten werden bereits durchgeführt. Die Entwicklungsarbeiten für die Recherchemenues sind noch nicht abgeschlossen. Das betrifft insbesondere Aspekte der Nutzerfreundlichkeit und der Recherchestrategie für das durch die ZADI zu realisierende Informationssystem EVA.

5. Literatur

Finlay, K. W. and Wilkinson, G. N. , 1963: The analysis of adaptation in a plant-breeding programme. Austral. J. Agric. Res. 14, S. 742 - 754.
Heiler,S. und Michels P., 1994: Deskriptive und Explorative Datenanalyse - München /Wien: R. Oldenbourg Verlag

Stegemann, K.; Dörfel, H.; Weise, V., 1995: Methodische Untersuchungen zur Sekundärauswertung von Sortenversuchen bezüglich der Ertragsstabilität von Sorten mit Hilfe der Ökoregression. In: Arch. Acker-Pf. Boden, S. 389 - 400

Wricke, G., 1962: Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. Z. Pflanzenzüchtg. 47, S. 92 - 96.

Wricke, G., 1964: Zur Berechnung der Ökovalenz bei Sommerweizen und Hafer. Z. Pflanzenzüchtg. 52, S. 127 - 138.

Wricke, G., 1965: Die Erfassung der Wechselwirkung zwischen Genotyp und Umwelt bei quantitativen Eigenschaften. Z. Pflanzenzüchtg. 53, S. 266 - 343.

Yates, F., and Cochran, W. G., 1938: The analysis of groups of experiments. J. Agric. Sci. 28, S. 566 - 580.