

**Institut für Angewandte Mathematik und Statistik
der Universität Hohenheim**

**OPTIMAL PLANNING OF
UNREPLICATED FIELD TRIALS
IN
RECURRENT SELECTION**

**Dissertation
zur Erlangung des Grades eines Doktors
der Naturwissenschaften**

**der Fakultät I
Allgemeine und Angewandte Naturwissenschaften
der Universität Hohenheim**

**vorgelegt von
Subhash Chandra
aus
Indien**

1991

S U M M A R Y

Optimal planning of unreplicated field trials in a given cycle of a recurrent selection scheme requires making decisions as to the

- optimal number of test units (T^*) and test locations (P^*), and
- optimal experimental design and adjustment method [the two together termed here as a local control (LC) strategy] for effective local control of soil heterogeneity in field trials.

These decisions need to be made in a manner that the expected gain from selection of a predetermined number s of superior selection units is maximized subject to a fixed total number N of plots available for test in the trial year of the cycle in question. To facilitate this task, the concept of "LC-BASED EXPECTED GAIN FROM SELECTION", and the "(A, δP)-CRITERION" have been introduced in this study.

In the LC-based expected gain from selection, the expected gain has been defined with reference to a specified LC strategy, and the choice of the number of test units and test locations is made conditional on the specified LC strategy. Moreover, it takes collectively into account the genotype-environment interactions and the trial fields' soil heterogeneity. The advantage accruing from these features of the LC-based expected gain is that, by optimizing it with respect to its arguments, answers as to the optimal form of the specified LC strategy as well as T^* and P^* can be objectively obtained, unlike hitherto, in a coherent and unified manner. Having obtained these answers for a number of LC strategies, the (A, δP)-criterion can be used to select the better one of any two optimized LC strategies. The quantity A is the per cent ratio of the optimal expected

gain in one LC strategy to that in a second one; it measures the per cent relative efficiency of the first as compared to the second LC strategy. The quantity δP is the algebraic difference between the optimal number of test locations for the first and the second LC strategy; it specifies the number of test locations that might be additionally required to get an efficiency of A per cent by adopting the first LC strategy.

Taking the linear variance (LV) soil heterogeneity model as an example, the derivation of the LC-based expected gain has been illustrated on a few one-dimensional LC strategies with reference to the following four design situations

- d_0 no grouping of plots, no checks,
- d_1 grouping of plots into fixed incomplete blocks,
- d_2 systematic check plot design, and
- d_3 checks-in-incomplete-block design.

For each LC strategy, analytical expressions for the first and the second derivatives of the LC-based expected gain have been provided to facilitate solution of the attendant optimization problem.

With no adjustment in d_0 , the results show that P^* , hence T^* for fixed $N-TP$, is a function of the relative testing capacity $N_r = (N/s)$ and the variance ratio $V = (1+w_q) + (w_p + w_{pq} + w)$ where $w_p = \sigma_{tp}^2 / \sigma_t^2$, $w_q = \sigma_{tq}^2 / \sigma_t^2$, $w_{pq} = \sigma_{tpq}^2 / \sigma_t^2$, $w = \sigma_e^2 / \sigma_t^2$ with σ_t^2 , σ_{tp}^2 , σ_{tq}^2 , σ_{tpq}^2 , σ_e^2 being respectively the genetic, genotype-location-interaction-, genotype-year-interaction-, genotype-location-year-interaction-, and plot-error-variance. It is found that, for fixed V , P^* increases, hence T^* decreases, as N_r is increased. On the contrary, for fixed N_r , P^* decreases, hence T^* increases, as V is increased. The P^* values are graphically presented as a function of N_r and V for practical use.

For adjustment by moving average in d_0 , and for adjustment by block mean not containing the observation to be adjusted in d_1 , a general solution of optimal block size (ζ^*) has been derived on the assumption of the LV model. It has been illustrated how the ζ^* value can be used to determine P^* , hence T^* for fixed $N=TP$, directly from the graphical results for the case of no adjustment in d_0 .

For the two design situations d_2 and d_3 , optimization techniques, likely to be useful to obtain solutions, have been given. These include the analytical approach ignoring the inequality restrictions, and the interior penalty function technique.

With the help of published data, it has been illustrated with reference to d_0 and d_1 how the $(A, \delta P)$ -criterion can be used.

ZUSAMMENFASSUNG

Um in einem gegebenen Zyklus eines rekurrenten Selektionsschemas die nichtwiederholten Feldversuche optimal planen zu können, fragt der Pflanzenzüchter nach der

- optimalen Anzahl für die Testeinheiten (T^*) und die Prüfforte (P^*); und der
- optimalen Versuchsanlage sowie Adjustierungsmethode - die beiden zusammen werden hier als "local control" (LC) Strategie bezeichnet - zur effektiven "local control" der Bodenheterogenität der Versuchsfelder.

Diese Fragen sollten in solcher Weise beantwortet werden, daß der bei festgelegter Anzahl s bester Selektionseinheiten zu erwartende Selektionsgewinn - unter Einhaltung der Restriktion einer fixen Gesamtzahl N von im Prüffjahr des gegebenen Zyklus verfügbaren Parzellen - maximiert wird. Um dieses Problem rational lösen zu können, werden in der vorliegenden Arbeit das Konzept des LC-ABHÄNGIGEN ZU ERWARTENDEN SELEKTIONSGEWINNES und das $(A, \delta P)$ -KRITERIUM eingeführt.

Der LC-abhängige zu erwartende Selektionsgewinn stellt denjenigen Selektionsgewinn dar, der einer spezifischen LC Strategie gegenüber definiert wird, wobei die Wahl der Testeinheits- und Prüffortsanzahl durch die betrachtete LC Strategie bedingt wird. Daneben werden auch die Bodenheterogenität der Versuchsfelder und die Genotyp-Umwelt-Interaktionen mitberücksichtigt. Diese Aspekte des LC-abhängigen zu erwartenden Selektionsgewinnes - durch seine Maximierung bezüglich seiner Argumente - ermöglichen es, objektiven Antworten auf die Fragen der optimalen Form der in Betracht gezogenen LC Strategie sowie T^* und P^* in kohärenter und geeinigter Weise zu bekommen. Nachdem solche Antworten bezüglich einer gewissen Zahl von LC Strategien erhalten worden sind, dient das $(A, \delta P)$ -Kriterium zu dem Zwecke, die

bessere von zwei beliebig optimierten LC Strategien auszuwählen. Hierbei gibt A an, in welchem Verhältnis (in Prozent) der optimale zu erwartende Selektionsgewinn in der einen zu dem in der anderen LC Strategie steht. Dies liefert ein Maß für die Effizienz der ersten im Vergleich zu der zweiten LC Strategie. Der Term δP stellt die algebraische Differenz zwischen der optimalen Prüfertsanzahl in der ersten und der zweiten LC Strategie dar. Damit erhält man Hinweise darauf, wieviele Prüferte wohl benötigt werden, um eine Effizienz von A Prozent durch Verwendung der ersten LC Strategie zu erhalten.

Die Herleitung des LC-abhängigen zu erwartenden Selektionsgewinnes wird am Beispiel des lineare-Varianz (LV)-Bodenheterogenitätsmodells an einigen ein-dimensionalen LC Strategien - die mit folgenden vier Anlagesituationen zusammenhängen - veranschaulicht:

- d_0 keine Gruppierung der Parzellen, keine Kontrollparzelle,
- d_1 Gruppierung der Parzellen in fixe unvollständige Blöcke,
- d_2 systematische Kontrollparzellenanlage, und
- d_3 Verwendung von Kontrollparzellen innerhalb unvollständiger Blöcke.

Für jede betrachtete LC Strategie werden die analytischen Ausdrücke für die ersten und die zweiten Ableitungen des LC-abhängigen zu erwartenden Selektionsgewinnes bezüglich der Einflußgrößen hergeleitet.

Wenn keine Adjustierung bezüglich d_0 erfolgt, zeigen die Ergebnisse, daß P^* und T^* für fixe $N=TP$ eine Funktion der relativen Prüfkapazität $N_r = (N/s)$ und des Varianzverhältnisses $V = (1+w_q) + (w_p + w_{pq} + w)$ darstellen. Hierbei sind $w_p = \sigma_{tp}^2 / \sigma_t^2$, $w_q = \sigma_{tq}^2 / \sigma_t^2$, $w_{pq} = \sigma_{tpq}^2 / \sigma_t^2$, $w = \sigma_e^2 / \sigma_t^2$, worin σ_t^2 , σ_{tp}^2 , σ_{tq}^2 , σ_{tpq}^2 , σ_e^2 die genetische, Genotyp-Ort-Interaktions-,

Genotyp-Jahr-Interaktions-, Genotyp-Ort-Jahr-Interaktions- bzw. die Parzellenfehlervarianzen sind. Der Wert von P^* nimmt zu, und derjenige für T^* nimmt ab, wenn N_r für feste Werte von V zunimmt. Im Gegenteil dazu nimmt P^* ab und T^* zu, wenn V für feste Werte von N zunimmt. Die P^* Werte werden als Funktion von N_r und V graphisch dargestellt.

Wenn wir zur Adjustierung den gleitenden Mittelwert bezüglich d_0 oder den Blockmittelwert bezüglich d_1 - der die zu adjustierende Beobachtung nicht einschließt - betrachten, ist hinsichtlich des LV Modells eine allgemeine Lösung des optimalen Blockumfangs (ζ^*) erhalten worden. Es wird gezeigt, wie der Wert ζ^* verwendet werden kann, um P^* und T^* für fixe $N=TP$ direkt aus der graphischen Lösung für den Fall ohne Adjustierung bezüglich d_0 zu ermitteln.

Zur Lösung der Optimierungsprobleme, die bei LC Strategien bezüglich d_2 und d_3 auftreten, werden zwei Methoden vorgeschlagen, nämlich der analytische Ansatz ohne Berücksichtigung der Ungleichheitsrestriktionen, und die Interior-Penalty-Funktionsmethode.

Mit Hilfe von in der Literatur erschienenen Daten wird für die Versuchsanlagen d_0 und d_1 veranschaulicht, wie das $(A, \delta P)$ -Kriterium verwendet werden kann.