

Auswahl von Klonen für Intensivkulturen')

Von Dietmar W. Rose²⁾

(Eingegangen August / Oktober 1974)

Der erhöhte Bedarf nach forstlichen Produkten hat starkes Interesse am hochintensiven Plantagenbetrieb zur Erzeugung von Faserholz in extrem kurzen Umtriebszeiten hervorgerufen. Da für diesen Zweck eine beinahe endlose Zahl von Klonen verfügbar sind, sind eine wirksame Differenzierung und Auswahl auf Grund der für die Produktionszielsetzung bedeutsamen Merkmale erforderlich. Einfache Techniken der Klumpenanalyse, die häufig ohne spezifische Annahmen über die Verteilungsfunktionen der Populationscharakteristiken durchgeführt werden können, erscheinen für diesen Zweck geeignet. Die Zielsetzung einer Klumpenanalyse ist die Gruppierung von Beobachtungen in mehr oder weniger homogene Kategorien. Die anschließende Auswahl vielversprechender Klone und Klongruppen kann durch Messung ihres Abstandes von einer durch den Mittelpunkt der Beobachtungen gehenden Ebene im vieldimensionalen Raum erfolgen. Die Analyse wurde für vielvariable Stichproben von 26 Pappelklonpopulationen ausgeführt.

Einleitung

Der erhöhte Bedarf nach forstlichen Erzeugnissen hat die Entwicklung neuartiger waldbaulicher Systeme hervorgerufen. Ein Vorschlag zur Lösung des Problems ist der Plantagenbetrieb, der die Anwendung von Düngemitteln und Bewässerung in engen Verbänden schnell wachsender Laubbestände im Ausschlagbetrieb mit Umtriebszeiten von drei bis sieben Jahren einschließt (MCALPINE *et al.*, 1966). Der Anbau genetisch verbesserter Pappelklone ist besonders vielversprechend und hat das größte Interesse gefunden. Die baldige Einführung solcher waldbaulichen Systeme wird von dem Erfolg, das Wachstum solcher Bestände in mathematischen Modellen zu beschreiben, abhängen, um langfristige Feldversuche zu vermeiden und eine ökonomische Bewertung vornehmen zu können (PROMNITZ und Rosa, 1974).

Mathematische Modelle müssen weiterhin auf ihre Voraussagekraft, und wenn eine allgemeine Hypothese des Wachstums intensiver Kulturen erwünscht ist, auf ihre Erklärungskraft geprüft werden. Weitere wichtige Gesichtspunkte sind die Differenzierung und die wirksame Auswahl der für den Plantagenbetrieb vielversprechenden Klone und Baumarten, um das Testen beinahe zahlloser Faktorenkombinationen vermeiden zu können.

Das Problem

Die Differenzierung von Populationen auf Grund von Einzelcharakteristiken allein ist nicht zufriedenstellend. Je mehr Merkmale jeder Population erhoben werden, desto

¹⁾ Journal Paper No. J-7539 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project 1877.

²⁾ Assistant Professor of Forestry, Department of Forestry, Iowa State University, Ames. Der Autor dankt Dozent Dr. Manfred HÜHN für seine zahlreichen, wertvollen Kommentare zu diesem Artikel.

³⁾ Paul WRAY, wissenschaftlicher Assistent an der forstlichen Abteilung in Ames, Iowa führte diese Wachstumsexperimente im Rahmen einer Dissertation in Forstbotanik durch. In diesen Experimenten wurden 26 Klone unter den kontrollierten Bedingungen einer Klimakammer aufgezogen. Der Autor dankt Herrn WRAY für die Zuverfügungstellung der Daten.

mehr Information zu deren Differenzierung ist vorhanden. Die Anwendung statistischer Verfahren, die alle Merkmale gleichzeitig einbeziehen, verlangt häufig strikte Annahmen über die Verteilungsfunktionen der Merkmale und die Struktur der Varianz-Kovarianzmatrixen der Populationen. Klumpenanalyse ist eine Technik, die ohne strikte Annahmen angewandt werden kann und typisch der Beantwortung zweier Fragen dient:

- 1) Fallen die Beobachtungen in natürliche Gruppen, und wenn das der Fall ist,
- 2) wieviele Gruppen sollten gebildet und wie sollte die Einteilung in diese Gruppen vorgenommen werden?

Im vorliegenden Falle sollten 26 Pappelklone (Tab. 1) auf Grund von vier, am Ende eines Wachstumsexperimentes gemessenen Merkmalen differenziert werden³⁾:

- 1) Pflanzenhöhe
- 2) Gesamtes Blatttrockengewicht
- 3) Stammtr洛克engewicht
- 4) Zahl der Blätter

Zwei weitere Variable wurden definiert als:

- 5) Verhältnis von Blatt- und Stammtr洛克engewicht
- 6) Durchschnittliches Blattgewicht

Nach der Gruppierung muß eine Bewertung der Klone hinsichtlich der für die Produktionszielsetzung bedeutsamen Merkmale erfolgen. Maximale Trockengewichtsproduktion wurde als das Hauptziel dieses Waldbaus angenommen, da eine vollmechanisierte Ernte und Nutzung der gesamten Pflanze vorausgesehen werden.

Drei Koeffizienten sind bisher als geeignete Maßstäbe der Ähnlichkeit zwischen Paaren von Taxen vorgeschlagen worden (SOKAL und SNEATH, 1963):

- 1) Assoziationskoeffizienten
- 2) Korrelationskoeffizienten
- 3) Entfernung im euklidischen Sinne.

Assoziationskoeffizienten, in weitem Gebrauch in der Ökologie, sind normalerweise auf Arrangierung einer Vierfeldertafel gestützt und auf Merkmale beschränkt, die in zwei Zuständen, „gegenwärtig“ oder „missend“, sein können. Korrelationskoeffizienten dagegen erlauben Merkmale in mehr als zwei Zuständen. Das dritte Maß bestimmt die Ähnlichkeit zweier Punkte auf Grund ihrer Entfernung im mehrdimensionalen Raum. Die Entfernung zwischen Punkten stellt ein besonders anschauliches Maß dar, und wurde in der folgenden Analyse als Ähnlichkeitskoeffizient gewählt.

Berechnung der Entfernung

Angenommen, wir untersuchen zwei Pappelklone 1 und 2, für welche zwei Merkmale, Höhe (x) und Stammtr洛克engewicht (y), erhoben wurden. Ein Maß der Ähnlichkeit zwischen den beiden Klonen auf Grund der gemessenen Merkmale ist die Entfernung zwischen zwei Punkten P_1 und P_2 mit Koordinaten (x_1, y_1) und (x_2, y_2) im zwei-dimensionalen Raum. Das Quadrat der Entfernung kann einfach nach der bekannten Formel aus der analytischen Geometrie berechnet werden:

$$E^2_{1,2} = (x_1 - x_2)^2 + (y_1 - y_2)^2 \quad (1)$$

Die Berechnung von $E^2_{1,2}$ im Fall von drei oder p Merkmalen ist einfach eine Summierung der Quadrate der Ent-

Tab. 1. — 26 Pappelklone mit Kennziffern der Forstlichen Versuchsanstalt in St. Paul, Minnesota

Klon	Station Kennziffer	
1	4877	<i>Populus alba</i> L.
2	4878 (5327)	<i>Populus × euramericana</i> GUINIER (<i>deltoides</i> × <i>nigra</i>)
3	4879	<i>Populus × euramericana</i> GUINIER
4	5258	<i>Populus</i> sp. (<i>deltoides</i>)
5	5262	<i>Populus candicans</i> AIT. × <i>Populus berolinensis</i> DIPP.
6	5263	<i>Populus candicans</i> AIT. × <i>Populus berolinensis</i> DIPP.
7	5264	<i>Populus deltoides</i> MARSH. × <i>Populus plantierensis</i> SCHNEID.
8	5265	<i>Populus deltoides</i> MARSH. × <i>Populus trichocarpa</i> TORR. und GRAY
9	5266	<i>Populus deltoides</i> MARSH. × <i>Populus trichocarpa</i> TORR. und GRAY
10	5267	<i>Populus deltoides</i> MARSH. × <i>Populus caudina</i>
11	5271	<i>Populus charkoviensis</i> SCHROED. × <i>Populus deltoides</i> MARSH.
12	5272	<i>Populus nigra</i> L. × <i>Populus laurifolia</i> LEDEB.
13	5321	<i>Populus × euramericana</i> GUINIER
14	5322	<i>Populus × euramericana</i> GUINIER
15	5323	<i>Populus × euramericana</i> GUINIER
16	5324	<i>Populus × euramericana</i> GUINIER
17	5325	<i>Populus × euramericana</i> GUINIER
18	5326	<i>Populus × euramericana</i> GUINIER
19	5327 (4878)	<i>Populus × euramericana</i> GUINIER
20	5328	<i>Populus × euramericana</i> GUINIER
21	5331	<i>Populus betulifolia</i> DIPP. × <i>Populus trichocarpa</i> TORR. und GRAY
22	5332	<i>Populus betulifolia</i> DIPP. × <i>Populus trichocarpa</i> TORR. und GRAY
23	5334	<i>Populus deltoides</i> MARSH. × <i>Populus trichocarpa</i> TORR. und GRAY
24	5260	<i>Populus tristis</i> FISH. × <i>Populus balsamifera</i> L.
25	5377	<i>Populus × euramericana</i> GUINIER, Wisconsin No. 5
26		<i>Populus grandidentata</i> × <i>Populus alba</i> , Crandon

fernungen entlang der drei, bzw. p Achsen in dem nun dreier- oder p-dimensionalen Raum. Die Formel für p Merkmale ist:

$$E^2_{1,2} = \sum_{i=1}^p (x_{i1} - x_{i2})^2 \quad (2)$$

Diese Formel kann vereinfacht als eine Vektormultiplikation dargestellt werden oder:

$$E^2_{1,2} = (x_1 - x_2)' (x_1 - x_2) = d'd \quad (3)$$

wobei $d = (x_1 - x_2)$ ein Vektor mit Dimension (px) ist, und x_1 und x_2 zwei Vektoren mit der gleichen Dimension sind. Abb. 1 gibt eine Darstellung der Entfernung zwischen zwei Klonen, wenn drei Merkmale (x, y, z) erhoben wurden. Mehr als drei Merkmale können in einer konventionellen Abbildung nicht dargestellt werden, und eine algebraische Manipulation der Daten wird erforderlich.

Die Entfernung zwischen zwei Punkten ist notwendigerweise eine Funktion der Zahl der erhobenen Merkmale. Je größer die Zahl der Merkmale, desto größer ist auch die Entfernung. Daher ist ein relatives Maß der Entfernung in manchen Situationen vorteilhafter. Ein solches relatives Maß kann durch Dividieren der Entfernung mit p, der Zahl der Merkmale, gefunden werden.

$$\bar{E}^2_{1,2} = 1/p \sum_{i=1}^p (x_{i1} - x_{i2})^2 \quad (4)$$

kann als quadratisches Mittel der Differenzen zwischen Merkmalen der beiden verglichenen Punkte erkannt werden und gleicht der Varianz der Differenzen (SOKAL, 1961). CZEKANOWSKI (1932) schlug die durchschnittliche Differenz

$$\bar{E}^2 = 1/p \sum_{i=1}^p |x_{i1} - x_{i2}| \quad (5)$$

als ein Maß taxonomischer Ähnlichkeit vor. Die Nachteile dieser Statistik sind, daß die wahre euklidische Entfernung zwischen Punkten immer unterschätzt wird, und dieses Maß im Gegensatz zur quadrierten Entfernung nicht in Komponenten geteilt werden kann.

Ein theoretischer Nachteil der Entfernung, der von SOKAL (1961) erkannt wurde, ist, daß die Charakteristiken als an

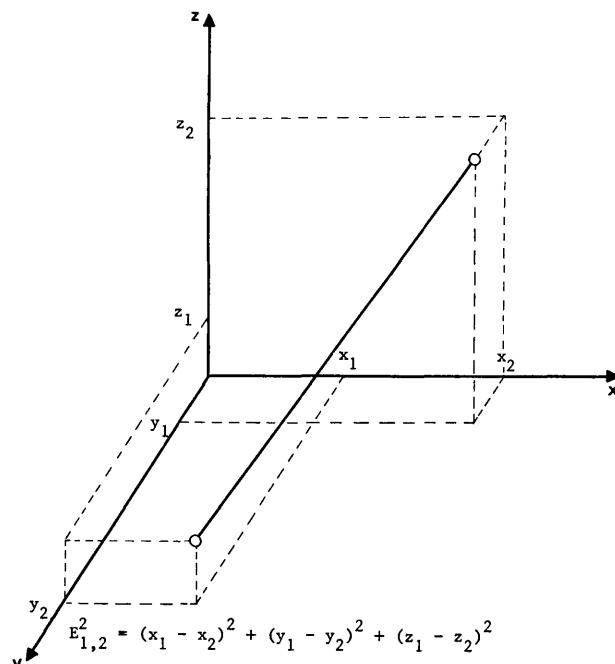


Abb. 1. — Entfernung zwischen zwei Punkten im dreidimensionalen Raum.

rechteckigen Achsen liegend betrachtet werden müssen, um den Pythagoras' Lehrsatz anwenden zu können. Da Populationmerkmale häufig in verschiedenem Maße korreliert sind, sind die entsprechenden Achsen nicht rechteckig zueinander. Die Anwendung von Formel 3 unter Annahme rechteckiger Koordinatenachsen führt zu einer Verzerrung des Raumes. Eine Lösung dieses Problems ist die rechteckige Wiedergabe der Originaldaten durch eine Rotation der Achsen oder durch eine Transformierung der Daten, die zur Unabhängigkeit der transformierten Variablen und damit zu einem euklidischen Raum für diese führt.

MAHALANOBIS (1936) schlug den Ausdruck

$$(u_j - u_k)' \Sigma^{-1} (u_j - u_k) \quad (6)$$

als ein Maß der Entfernung zwischen zwei Populationen j und k vor, wobei u_j und u_k die Vektoren der Populationsmittelwerte für p Merkmale sind und die Varianz-Kovarianzmatrix darstellt. Diese Gleichung führt zu der oben beschriebenen Transformierung der Daten durch Multiplikation der Originaldaten mit der Matrix T der charakteristischen oder Eigenvektoren für die Matrix Σ oder:

$$(T u_j - T u_k)' (T u_j - T u_k) = (u_j - u_k)' \Sigma^{-1} (u_j - u_k) \quad (7)$$

da $T' T = \Sigma^{-1}$. Dieser Ausdruck kann als ein gewichteter Koeffizient ähnlich der quadrierten Entfernung, in welchem die Variation einzelner Merkmale und die Korrelation zwischen Merkmalen in Betracht gezogen sind, angesehen werden. Gleichung 6 trägt davon Rechnung, daß die Achsen in dem p -dimensionalen Raum verschiedene Maßeinheiten haben können. Wenn nicht alle Variablen die gleiche relative Varianz haben, ist es möglich, daß zwei Variable mit großer Varianz die meisten Ähnlichkeitskoeffizienten dominieren, die euklidische Entfernung eingeschlossen. Eine Normierung der Daten führt zu Variablen mit vergleichbarer Varianz und versichert, daß bedeutend korrelierten Variablen nicht übermäßiges Gewicht beigegeben wird, und daß Gruppierungen, die auf hohen Korrelationen innerhalb der Gruppen beruhen, leicht entdeckt werden können (MARIOT, 1971).

Mit einer vielvariablen Stichprobe von zwei Populationen j und k kann

$$(\bar{x}_j - \bar{x}_k)' S^{-1} (\bar{x}_j - \bar{x}_k) \quad (8)$$

zur Schätzung der quadrierten Entfernung zwischen den beiden Populationen benutzt werden, wobei \bar{x}_j und \bar{x}_k die Vektoren der arithmetischen Mittelwerte für p Merkmale der zwei Populationen darstellen, und S die auf die Stichprobe basierte Schätzung von Σ ist. Unter der Annahme identischer Varianz-Kovarianzmatrix für k Populationen kann Σ von der gesamten Stichprobe

$$n = \sum_{h=1}^k n_h \text{ durch} \\ S = \frac{1}{n_1 + n_2 + \dots + n_k - k} (A_1 + A_2 + \dots + A_k) \quad (9)$$

geschätzt werden, wobei

$$A_h = (n_h - 1) \left[\begin{array}{cccccc} \sum_{i=1}^{n_h} (x_{i1} - \bar{x}_1)^2 & \dots & \sum_{i=1}^{n_h} (x_{i1} - \bar{x}_1) (x_{ip} - \bar{x}_p) \\ \vdots & & \vdots & & & \vdots \\ \sum_{i=1}^{n_h} (x_{i2} - \bar{x}_2)^2 & \dots & \dots & & & \vdots \\ \vdots & & \vdots & & & \vdots \\ \sum_{i=1}^{n_h} (x_{ip} - \bar{x}_p)^2 & \dots & \dots & & & \vdots \end{array} \right] \quad (10)$$

Die Hypothese $H_0: \Sigma_1 = \Sigma_2 = \dots = \Sigma_k$ wurde durch die vielvariable Analogie des Bartlett Test's (MORRISON 1967) getestet. Der Test ist recht sensibel hinsichtlich Abweichungen der Variablen von der Normalität und daher nicht sehr robust. Die Annäherung an eine Chi-Quadrat Verteilung ist auch nur gut, wenn k und $p \leq 5$ und $n_h \geq 21$, was im vorliegenden Fall nicht gegeben war. Daher wurde, obwohl die Nullhypothese verworfen wurde, Σ durch S nach Formel 9 geschätzt. REYMENT (1962) wies darauf hin, daß das verallgemeinerte Quadrat der Entfernung nicht sehr

sensibel mit Rücksicht auf Heterogenität in den Kovarianzmatrixen ist.

Klumpenanalyse

Jede individuelle Messung einer vielvariablen Stichprobe kann als ein Punkt im mehrdimensionalen Raum dargestellt werden. Die Klumpenanalyse versucht, diese Punkte in homogene Einheiten zu gruppieren, in der Hoffnung, daß diese mit markanten Charakteristiken der Stichprobe übereinstimmen (EDWARDS und CAVALLI-SFORZA, 1965). Die Klumpenanalyse umfaßt eine Vielzahl numerischer Techniken, die auf Ähnlichkeitskoeffizienten basierte Gruppen definieren.

Für zwei Merkmale können Punkte leicht in einer zweidimensionalen Darstellung interpretiert und Gruppierungen erkannt werden. Wenn mehr als zwei Merkmale gemessen wurden, müssen andere Wege gefunden werden, um natürliche Gruppierungen zu erkennen. Ein erster Schritt in Klumpenanalyse ist die Arrangierung der Beobachtungen in einer symmetrischen Assoziationsmatrix (Tab. 2). Für die vorliegende Matrix wurden die paarweisen, auf sechs Merkmale basierten Entfernungswerte gewählt und mit Hilfe von Gleichung 8 berechnet.

Eine der einfachsten Methoden der Gruppierung von Punkten besteht darin, daß die zwei Elemente der Matrix, die am nächsten zusammenliegen oder den größten Ähnlichkeitskoeffizienten besitzen, in eine Gruppe zusammengefaßt werden. Im nächsten Schritt werden die Reihen (Spalten) der beiden entsprechenden Elemente durch eine Reihe (Spalte) ersetzt, und die kürzeste Entfernung aller anderen Punkte zu der gebildeten Gruppe wird dann berechnet (single-linkage method). In der so reduzierten Matrix werden wiederum die beiden naheliegendsten Elemente in einen Klumpen zusammengefaßt. In diesem Prozeß der Agglomeration werden mehr und mehr Punkte oder bereits geformte Klumpen durch Herabsetzung des Kriteriums der Zulassung zu einem Klumpen hinzugefügt, und

dieser Prozeß wird wiederholt, bis alle Punkte in den Klumpen eingeschlossen sind.

In der Berechnung einer neuen Assoziationsmatrix kann auch die längste Entfernung aller Punkte oder Gruppen zu einem Klumpen gewählt werden (complete-linkage method). Die Durchschnittsmethode (average-linkage method) unterscheidet sich von der „single-linkage“ Methode darin, daß nach der Formung eines neuen Klumpens die Entfernung aller anderen Elementen zum Mittelpunkt des geformten Klumpens berechnet wird.

Tab. 2. — Assoziationsmatrix für 26 Pappelklonen mit der auf 6 Merkmalen basierten Entfernung als Ähnlichkeitskoeffizienten

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	
1	0.00																										
2	2.62	0.00																									
3	5.46	2.92	0.00																								
4	3.95	9.52	9.09	0.00																							
5	4.42	9.29	10.14	6.03	0.00																						
6	0.25	2.55	4.99	5.38	4.86	0.00																					
7	11.66	5.86	10.31	17.09	18.79	11.87	0.00																				
8	5.42	5.75	12.83	7.52	12.85	7.25	6.99	0.00																			
9	10.03	6.77	11.74	13.24	18.58	10.95	4.05	4.58	0.00																		
10	10.96	10.67	17.31	14.78	12.49	12.31	10.32	8.15	18.25	0.00																	
11	8.17	6.91	11.67	18.56	6.29	7.85	15.76	15.34	18.72	12.78	0.00																
12	1.52	1.94	7.80	7.64	7.69	2.03	7.32	2.69	7.57	6.75	7.44	0.00															
13	4.82	2.77	5.21	7.32	7.38	5.60	5.36	4.21	6.97	6.30	7.91	3.55	0.00														
14	11.75	9.12	12.50	12.83	13.61	14.13	12.04	7.93	15.78	5.76	13.53	8.92	2.97	0.00													
15	0.82	1.34	4.48	4.31	6.03	1.46	7.33	2.92	5.80	9.46	8.39	1.03	2.23	8.33	0.00												
16	3.49	4.95	10.34	7.02	3.81	4.72	9.65	4.01	8.59	7.15	5.12	2.68	2.64	7.19	2.49	0.00											
17	2.75	5.31	3.05	6.00	2.14	9.41	5.70	12.26	10.34	2.33	3.11	9.61	0.39	2.85	0.00												
18	4.90	7.13	3.27	3.46	5.94	4.87	13.39	12.56	12.92	17.31	10.31	9.27	6.88	15.41	5.01	9.23	4.63	0.00									
19	5.17	0.30	2.60	10.69	12.51	5.30	3.78	6.18	5.87	10.78	10.30	3.76	2.30	7.83	2.61	7.00	4.13	7.37	0.00								
20	22.98	16.77	20.57	24.99	23.50	24.36	7.03	15.30	17.53	6.97	21.50	16.98	9.02	8.96	17.54	15.11	20.54	22.20	12.95	0.00							
21	1.27	1.38	3.51	4.86	1.77	4.48	6.63	6.55	10.24	6.28	1.99	1.74	8.05	2.05	0.72	4.18	2.65	17.04	0.00								
22	2.46	0.32	2.51	5.77	1.30	8.09	6.89	8.60	10.92	5.18	2.02	2.93	9.79	1.09	3.87	2.18	5.12	2.05	18.50	0.67	0.00						
23	13.90	6.75	9.66	18.61	22.38	14.19	1.31	8.91	2.61	17.21	19.14	10.32	6.75	14.53	8.61	12.43	9.90	13.89	4.00	11.30	8.72	9.36	0.00				
24	33.34	40.00	27.71	28.67	27.36	31.39	49.26	53.77	59.57	41.53	45.51	42.61	39.67	48.49	38.02	44.30	39.37	19.34	40.25	49.89	36.32	34.75	54.99	0.00			
25	1.77	3.16	2.99	2.90	4.14	2.19	9.73	6.61	7.65	13.31	9.56	4.21	3.20	10.72	1.27	3.88	0.98	1.63	4.17	19.58	0.73	1.85	10.15	30.20	0.00		
26	12.02	5.40	2.74	17.96	16.20	10.96	6.84	17.29	11.87	18.97	13.17	12.11	7.19	15.49	9.24	14.23	11.16	8.06	3.53	15.23	7.56	5.96	6.06	33.49	7.79	0.00	

Die Teilung eines Klumpens, in welcher die entstehenden Klumpen gemäß eines bestimmten Kriteriums so dicht wie möglich sind, ist sinnvoll (MARRIOT, 1971; MAYER, 1971). EDWARDS und CAVALLI-SFORZA (1965) fanden in der Summe der Quadrate der Entfernungen analog zur Varianzanalyse ein Kriterium, das bedeutungsvoll in der Anwendung und einfach in der Berechnung ist. Die Summe der Quadrate der Entfernungen von Punkten auf einer Linie von ihrem Mittel kann in zwei Summen der Quadrate, „innerhalb“ und „zwischen“ den Gruppen, geteilt werden, wenn die Punkte in zwei Gruppen klassifiziert sind. Dieses Ergebnis ist für Punkte in jeglicher Zahl von Dimensionen wahr. Daher kann, wenn Punkte in zwei Klumpen geteilt werden, die Summe der Quadrate der Entfernungen der Punkte von ihrem Mittel in die Summe der Quadrate zu den entsprechenden Mittelwerten und die „zwischen-den-Klumpen-Summe-der-Quadrate“ (SS_z) geteilt werden. Das natürliche Kriterium für die Teilung ist offensichtlich SS_z , und die beste Teilung ist die, für welche diese Summe ein Maximum ist, und die „innerhalb-Summe-der-Quadrate“ (SS_i) folglich ein Minimum. EDWARDS und CAVALLI-SFORZA (1965) schlugen einen F-Test zum Vergleich der mittleren Quadrate SS_z und SS_i wie in einer herkömmlichen Varianzanalyse vor.

Die beschriebene Teilungstechnik ist einfach in Praxis und Theorie. Da n Punkte in $(2^{n-1}-1)$ Wegen in zwei Gruppen geteilt werden können, wird schon mit relativ wenigen Punkten eine elektronische Berechnung erforderlich. Eine Sequenzmethode, die das Testen aller Teilungen vermeidet, wurde bereits für mehr als 16 Punkte vorgeschlagen (EDWARDS und CAVALLI-SFORZA, 1965).

Das Ergebnis einer Klumpenanalyse kann in einem „Baumdiagramm“ dargestellt werden (McCAMON und WENNINGER, 1970; HARTIGAN, 1967) und wurde für sechs Variable nach der Durchschnittsmethode durchgeführt (Abb. 2). Die Abbildung deutet auf eine Klumpung der Klone (15, 17), (1, 6), (2, 22), (3, 26) und (7, 23) und auf eine Ähnlichkeit der Klone (15, 17, 21, 25, 1, 6, 12, 16, 13, 19, 22, 2) hin. Klon 24 hat am wenigsten Ähnlichkeit mit allen anderen Klonen, was auch durch die horizontale Achse oder die quadrierte Entfernung zwischen den Klumpen demonstriert wird.

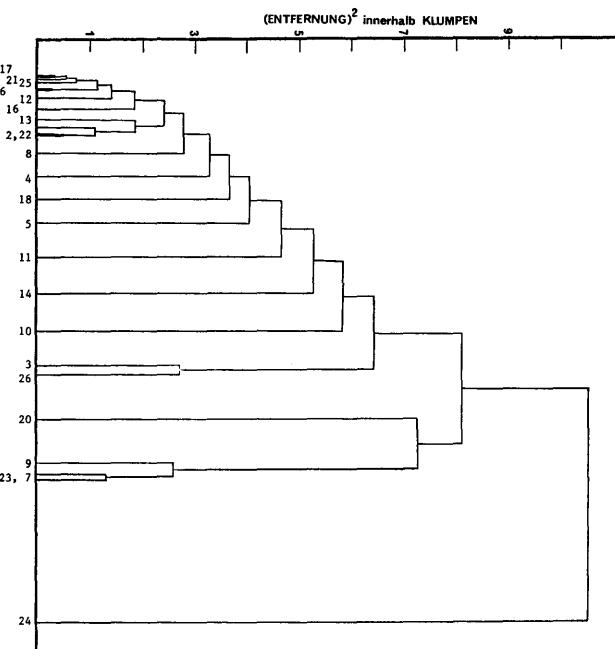


Abb. 2. — Klumpung von 26 Pappelklonen nach der Durchschnittsmethode und basiert auf 6 Variable.

Wahl der Klumpungsmethode

Es ist schwierig, die relativen Vor- und Nachteile der verschiedenen Klumpungstechniken zu beurteilen, da ein Klumpen nicht ein streng definiertes Konzept darstellt (GOWER, 1967), und verschiedene Methoden der Klumpenanalyse an der gleichen Stichprobe verschiedene geometrische Punktverteilungen annehmen können.

Ein objektives Maß zum Vergleich von Baumdiagrammen, die durch verschiedene Klumpungsmethoden oder auf Grund einer unterschiedlichen Anzahl von Variablen erhalten wurden, wurde von SOKAL und ROHLF (1962) entwickelt. Die sogenannte kophänetische Korrelation beruht auf einem Vergleich des Baumdiagramms mit den tatsächlichen Koeffizienten der Ähnlichkeitsmatrix und mißt die Verzerrung, welche die Klumpungsmethode dem System auferlegt hat. Die kophänetische Korrelation für die Klumpung nach der Durchschnittsmethode, basiert auf 4 Variablen, war 0.49 und deutete darauf hin, daß das Diagramm die originalen Ähnlichkeiten recht merklich verzerrt hat. Die kophänetische Korrelation für 6 Variable war 0.73 und ist ein Hinweis dafür, daß die Klumpung mit anwachsender Zahl der Variablen wirksamer wird, worauf auch SOKAL und SNEATH (1963) hingewiesen haben.

HARTIGAN (1967) untersuchte Klumpenstabilität und Wege zur Bewertung dieser Eigenschaft. Klumpen wurden als stabil angesehen, wenn sie in zufälligen Unterstichproben der Gesamtdaten beständig blieben. Im Gebrauch ist diese Methode sehr subjektiv, da die Ausdrücke „Beständigkeit“ und „Stabilität“ keine präzisen Definitionen haben.

Auswahl vielversprechender Klone und Klongruppen

Die Klumpenanalyse hat zu einer Differenzierung der Klone in Gruppen mit ähnlichen Charakteristiken geführt. Ein weiterer, wichtiger Schritt ist die Erstellung einer Rangordnung der Klone und Klongruppen auf Grund der für die Produktionszielsetzung bedeutsamen Merkmale. Die Erstellung einer Rangordnung auf Grund von Einzelmerkmalen führt oft dazu, daß Klone verschiedene Ränge für verschiedene Merkmale annehmen (Tab. 3, Spalten 1—4). Die Summe der Einzelränge ist daher ein besseres Maß.

Die Auswahl auf Grund einzelner Merkmale kann zur Vernachlässigung wichtiger Beziehungen zwischen den Merkmalen und zu falschen Schlußfolgerungen über die Leistungsfähigkeit eines Klones führen. Vielvariable Techniken erscheinen daher wünschenswert, obwohl diese eine visuelle Interpretation schwieriger machen. Möglichkeiten der Erstellung einer, auf alle erwünschten Merkmale basierten, Rangordnung bestehen in Entfernungsmessungen der Klone von einem Bezugspunkt im vieldimensionalen, normierten (euklidischen) Raum (Abb. 3):

- 1) (Quadrierte) Länge des Merkmalsvektoren für Klon i oder Abstand des Punktes $P_1(x_1, y_1, z_1)$ vom Koordinatenursprung (Spalte 5, Tab. 3).
- 2) Die quadrierte Entfernung E^2 zwischen dem Gesamtprobenmittelpunkt $P(a, b, c)$ und dem einzelnen Klon i Spalte 6, Tab. 3).
- 3) Der vertikale Abstand der Klone von einer durch den Mittelpunkt gehenden und zum Mittelwertvektoren rechtwinkligen Trennungsebene (Spalte 7, Tab. 3).
- 4) Klassifizierung der Klone durch eine Einteilung des Raumes in Unterräume und Beobachtung, in welchen Raum die Klone fallen (Spalte 8, Tab. 3):
 - a) Raum A unterhalb einer zum Mittelwertvektoren rechtwinkligen Ebene $ax + by + cz = D$, wobei $D = a^2 + b^2 + c^2$;
 - b) Raum B, für den $x > a, y > b, z > c$;

c) Raum C, für den $x > a^*, y > b^*, z > c^*$;
 d) Räume oberhalb der Trennungsebene und nicht in den Räumen B und C, insbesondere

Raum D: x, y
 E: x, z
 F: y, z
 G: x
 H: y
 I: z

größer als Durchschnitt

Alle Maße sind in jeglicher Dimension anwendbar und wurden im vorliegenden Falle auf drei Merkmale basiert.

Die Maße 2 und 3 lassen nicht erkennen, auf welcher Seite des Mittelwertes ein Klon liegt. Im Zusammenhang mit der Einteilung in die Räume A-I und der Vektorenlänge kann diese Frage beantwortet werden. Klone oberhalb der Trennungsebene sind besser als der Durchschnitt, in den Räumen B und C mit Rücksicht auf alle Merkmale. Klone unterhalb der Trennungsebene sind ebenfalls besser als der Durchschnitt, solange der entsprechende Vektor länger als der Durchschnittsvektor ist, und alle Merkmale das gleiche Gewicht erhalten. Klone im Raum C sind besser als alle Klone unterhalb der Trennungsebene. Der Vergleich von Klonen im Raum A, deren Vektoren länger als der Durchschnittsvektor sind, mit Klonen in den Räumen B und D-I muß auf einen Vergleich der Vektorenlänge gestützt werden. Dieser Vergleich ist auch zwischen Punkten in den Räumen D-I und B und C anwendbar.

Der Winkel δ zwischen dem Mittelwertvektor und den Einzelvektoren (Spalte 9, Tab. 3), zusammen mit dem Raum, in den der Einzelkロン fällt, gibt einen weiteren, nützlichen Hinweis auf die räumliche Anordnung der Punkte, zumindest solange drei Dimensionen nicht überschritten werden. Ein kleiner Winkel deutet darauf hin, daß Klone einfachheitshalber auf Grund der Länge ihrer Merkmalsvektoren und des Mittelwertvektors verglichen werden könnten, auch wenn den einzelnen Merkmalen unterschiedliche Gewichte beigemessen würden.

Auf Grund der Rangsummen sind die Klone 18, 24, 9, 4, 1 und 17 zu empfehlen. Diese Empfehlungen sind in diesem Fall in Übereinstimmung mit den Entfernungskriterien im vieldimensionalen Raum. Diese Übereinstimmung kann nicht in jedem Fall erwartet werden, da Rangsummen nicht vollständig die Beziehungen zwischen den Merkmalen in Betracht ziehen. Am wenigsten zu empfehlen sind die Klone im Raum A, für die alle Merkmale kleiner als der Durchschnitt sind oder die Klone 10, 11, 13, 14, 16 und 20. Keiner der Klone fiel in den Raum C. Die besten Klone im Raum B sind die, die am weitesten vom Koordinatenursprung entfernt sind oder die Klone 24, 6, 18, 9, 1, 4 und 17. Einige Klone in den Räumen D-I wie die Klone 7, 8, 12, 19, 22 und 26 haben Vektorenlängen, die mit den Klonen 15, 1 und 4 vergleichbar sind. Der meist kleine Winkel zwischen den Klonen und dem Mittelwertvektor erklärt, warum die Mehrzahl der Punkte oberhalb der Trennungsebene in den Raum B fallen. Der vertikale Abstand der Punkte von der Trennungsebene reflektiert ebenfalls die obigen Ergebnisse und ist ein weiterer Hinweis darauf, daß Klone in den Räumen D-I mit Klonen im Raum B oder C vergleichbar sein können.

Die Klumpung, die in der Klumpenanalyse mit 6 Variablen abgeleitet wurde, ist im allgemeinen auch in den verschiedenen Rangordnungsmaßen von Tabelle 3 wiedergegeben. Abweichungen sind jedoch, da die Zahl der Variablen für die Klumpung und die Klonauswahl unterschiedlich waren, vorhanden.

Tab. 3. — Rangordnungen von 26 Pappelkronen auf Grund von drei Merkmalen

Klon	Höhe	Rangordnung		Summe der Ränge	$x_i S^{-1} x_i'$	$(x_i - \bar{x}) S^{-1} (x_i - \bar{x})'$	Entfernung von Ebene	7	8	9
		Blatt- gewicht	Stamm- gewicht							
1	24	16	16	56	48.60	0.58	0.05	B	2	
2	21	4	8	33	45.24	0.91	0.17	G	7	Quadrierte Länge des Mittelwertvektoren:
3	23	10	21	44	47.27	2.21	0.90	E	11	$\bar{x} S^{-1} \bar{x}' = 39.75$
4	13	25	24	62	47.64	2.00	0.40	B	10	Raum A: $ax + by + cz < D$
5	7	11	19	37	32.91	1.92	0.92	A	12	Raum B: $x > a, y > b, z > c$
6	25	15	18	58	52.54	1.31	0.18	B	5	Raum C: $x > a^*, y > b^*, z > c^*$
7	10	21	11	42	46.06	2.05	1.01	D	11	Räume D-I:
8	8	22	9	39	46.05	3.56	1.30	H	15	$ax + by + cz > D$ und ein oder zwei Merkmale größer als der Durch- schnitt, insbesondere
9	20	26	22	68	53.93	3.89	1.02	B	14	D: x, y
10	3	3	2	8	29.56	3.85	1.01	A	17	E: x, z
11	4	1	3	8	27.30	3.61	0.35	A	15	F: y, z
12	19	12	5	36	46.03	1.34	0.70	G	9	G: x
13	5	8	6	19	30.04	0.84	0.17	A	4	H: y
14	1	2	1	4	19.17	4.98	0.46	A	12	
15	17	17	13	47	44.62	0.19	0.21	B	2	
16	6	13	7	26	30.93	1.02	0.32	A	6	
17	14	20	20	54	45.13	0.32	0.08	B	3	
18	22	24	25	71	51.53	3.67	1.12	B	14	
19	15	9	10	34	43.36	0.48	0.22	G	5	
20	2	7	4	13	23.58	5.64	0.92	A	19	
21	9	14	14	37	39.81	0.03	0.13	E	1	
22	18	5	12	35	43.48	0.71	0.21	G	6	
23	11	23	15	49	47.43	1.73	0.80	B	10	
24	26	18	26	70	71.26	11.84	2.91	B	29	
25	12	19	23	54	44.51	0.89	0.52	B	8	
26	16	6	17	39	43.07	1.36	0.73	E	10	

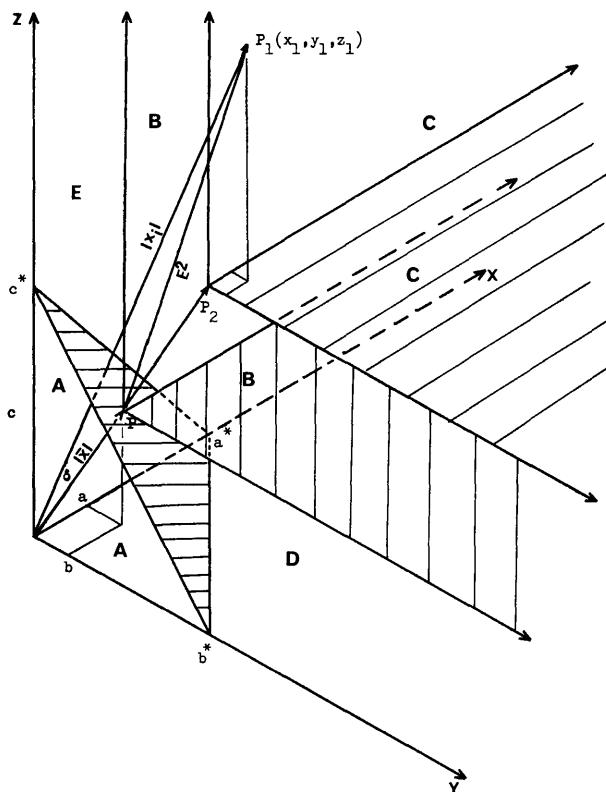


Abb. 3. — Einteilung eines drei-dimensionalen Raumes in Unter-räume zur Klassifizierung von Pappelklonen.

Die Merkmalsvektoren für jeden Klon repräsentieren Durchschnittswerte einer Stichprobe von Klonen. Ein Vertrauenskreis (-sphäre) könnte als eine Funktion der erwünschten Vertrauensebene und des Stichprobenfehlers um jeden Punkt konstruiert werden. Die Durchmesser der Vertrauenskreise oder -sphären könnten z. B. als die Wurzel der Varianz der Vektorenlängen für einen Klon berechnet werden. Damit können genauere Einteilungen der Punkte in die Unterräume vorgenommen werden.

Alle drei Merkmale erhielten in der vorliegenden Analyse das gleiche Gewicht. Die Analyse kann jedoch auch mit gewichteten Merkmalswerten erfolgen und könnte zu unterschiedlichen Schlußfolgerungen führen. Die Frage der korrekten Gewichtung muß mit Rücksicht auf das Produktionsziel beantwortet werden.

Zusammenfassung

Ohne Annahmen über die Verteilungsfunktion der Populationssmerkmale muß die Klumpenanalyse eine rein beschreibende Technik bleiben, und ihre Nützlichkeit liegt in der Kondensation umfangreichen Datenmaterials in eine leicht interpretierbare Form. Die Ergebnisse der Klumpenanalyse erlauben Schlüsse über die Ähnlichkeit von Klonen und vermitteln Richtlinien für intensivere Untersuchungen von vielversprechenden Klongruppen. Die Klumpung kann den Umfang des Problems, nämlich Testen aller Klone dadurch wirksam reduzieren, daß die Untersuchung nur einzelner Vertreter vielversprechender Gruppen hinreichend sein mag. Größere Stichproben für ausgewählte Klonen machen vielvariable statistische Analysen, die quantitative Schlußfolgerungen über die Gruppierungen und Ähnlichkeiten von Objekten zulassen, möglich (ANDERSON, 1958).

Die Auswahl von Klonen auf Grund der für die Produktion bedeutsamen Variablen, unter Berücksichtigung der komplexen Beziehungen zwischen diesen Variablen, ist

Techniken, die sich auf Einzelmerkmale stützen, vorzuziehen. Entfernungsmessungen im vieldimensionalen, normierten Raum erscheinen vorteilhaft, da eine einzige Zahl den Einfluß aller Variablen wiedergeben kann.

Das Wachstum von Gewächshaus- oder Klimakammerpflanzen lässt keine endgültigen Schlüsse über die Leistung von Klonen unter Feldbedingungen zu. WRAY (1974) wies auf die Stabilität von Klonen unter verschiedenen Umweltbedingungen als ein wichtiges Auswahlkriterium hin. Die Klone, die letztlich für die Produktion gewählt werden sollten, sind die, die stabil über mehrere Umweltbedingungen bleiben und überdurchschnittlich hinsichtlich ihrer Produktionseigenschaften sind. Die Leistungsfähigkeit der Einzelklone sollte daher in Feldversuchen unter verschiedenen Umweltbedingungen überprüft werden. Mit der Anwendung der oben beschriebenen Auswahlkriterien kann dann festgestellt werden, ob ein Klon in anderen Umweltsituationen seine Stellung im Raum relativ zum Vergleichspunkt oder zur Vergleichsebene ändert oder nicht. Diese Information ist erforderlich, wenn die Kosten falscher Entscheidungen schwierigend sind.

Schlagworte: Klumpenanalyse, Entfernungsmäße, Pappelklone.

Abstract

Selection of clones for intensive cultures of woody plants

The increased demand for forest products has created strong interest in intensive cultures of woody species for fiber production in extremely short rotations. An effective differentiation and selection of clones based on characteristics important for the production goal are essential because an almost unlimited number of clones is available for such intensive cultures. Simple cluster-analysis techniques can often be applied to group observations into categories with similar characteristic values without having to make specific assumption about the distribution functions of population characteristics. The subsequent selection of promising clones can be accomplished by measuring distances of the clones from a (hyper-) plane through the statistical average of all observations in the multi-dimensional space. The analysis was performed for multi-variate samples of 26 *Populus* clones.

Literaturhinweise

ANDERSON, T. W.: Introduction to multivariate statistical analysis. John Wiley & Sons, New York (1958). — CZEKANOWSKI, J.: Coefficients of racial likeness und durchschnittliche Differenz. *Anthrop. Anz.* 9, 227—249 (1932). — EDWARDS, A. W., and CAVALLI-SFORZA, L. L.: A method for cluster analysis. *Biometrics* 2 (21), 362—385 (1965). — GOWER, J. C.: A comparison of some methods of cluster analysis. *Biometrics* 4 (23), 623—637 (1967). — HARTIGAN, J. A.: Representation of similarity matrices by trees. *J. Am. Stat. Assoc.* 62, 1140—1158 (1967). — MAHALANOBIS, P. C.: On the generalized distance in statistics. *Proc. Natl. Inst. India* 12, 49—55 (1936). — MARRIOT, F. H. C.: Practical problems in a method of cluster analysis. *Biometrics* 3 (27), 501—514 (1971). — MAYER, L. S.: A method of cluster analysis when there exist multiple indicators of a theoretic concept. *Biometrics* 1 (27), 143—155 (1971). — McALPINE, R. G., BROWN, C. L., HERRICK, A. M., and RUARK, H. E.: Silage sycamore. *Forest Farmer* 26, 6—7 (1966). — McCAMMON, R. R., and WENNINGER, G.: The Dendrogram. Computer Contribution 48, Kansas State Geological Survey, Univ. of Kansas, Lawrence, Kansas (1970). — MORRISON, D. F.: Multivariate statistical methods. McGraw-Hill Book Co., London (1967). — PROMNITZ, L. C., and Rose, D. W.: A mathematical conceptualization of a forest stand simulation model. *Angew. Botanik* 48, 97—108 (1974). — RAO, C. R.: Advanced statistical methods in biometric research. John Wiley & Sons, New York (1952). — REYMENT, R. A.: Observations on homogeneity of covariance matrices in paleontologic biometry. *Biometrics* 1, 1—11 (1962). — SOKAL, R. R.: Distance as a measure of taxonomic similarity. *Syst. Zool.* 2 (10), 70—79 (1961). — SOKAL, R. R., and ROHLF, F. G.: The comparison of dendograms by objective methods. *Taxon* 11 (2), 33—39 (1962). — SOKAL, R. R., and SNEATH, P. H.: Principles of numerical taxonomy. W. H. Freeman, San Francisco and London (1963). — WRAY, P. H.: Peroxidase and growth in hybrid poplars. Unpubl. Ph. D. dissertation submitted to Dept. of Forestry, Ames, Iowa (1974).