

Mitteilungen
des
Deutschen Wetterdienstes

Nr. 37
(Band 5)

DK 551.501.45:551.586:631.5

Versuch einer statistischen Analyse der Beziehung
zwischen Witterung und Ertrag mit Hilfe
multipler Regressionen

von

Hans Baumann und Ernst Weber

(mit 4 Abbildungen und 8 Tabellen im Text)



Offenbach a. M. 1966
Selbstverlag des Deutschen Wetterdienstes

Inhalt

	Seite
Zusammenfassung	3
Abstract	3
1. Einleitung	3
2. Untersuchungsmaterial	4
3. Methodik	4
4. Untersuchungsergebnisse	7
4.1. Der Ertrag in Abhängigkeit von Witterungsfaktoren, ausgewählt nach den bisherigen Erfahrungen mit dem Hafer	7
4.2. Der Einfluß des Witterungsfaktors x_4 (mittlere Temperatur der 1. Juli-Dekade) auf die Erhöhung der multiplen Bestimmtheit	11
4.3. Erweiterung der multiplen linearen Regression durch multiplikative Glieder	11
4.4. Der Ertrag in Abhängigkeit von schematisch ausgewählten Witterungsdaten	13
4.5. Zur Frage der Vorausschätzung von Erträgen	16
5. Zusammenfassung der wichtigsten Ergebnisse	21

Anschriften der Verfasser:

Prof. Dr. H. Baumann, Institut für Wasserwirtschaft und Meliorationswesen der Universität Kiel, 23 Kiel, Olshausenstraße 40-60

Priv.-Doz. Dr. E. Weber, Landwirtschaftliche Fakultät — Variationsstatistik — Universität Kiel, 23 Kiel, Olshausenstraße 40-60

Zusammenfassung

In der vorliegenden Arbeit wird der Versuch unternommen, das Modell der multiplen linearen Regression auf die Beziehung zwischen Witterungsfaktoren (als Einflußgrößen) und dem Ertrag (als Zielgröße) anzuwenden. Da nicht alle Einflußgrößen im gleichen Maße an der Erklärung der Zielgröße beteiligt sind, werden über ein Eliminierungsverfahren diejenigen Witterungsfaktoren ermittelt, die zu einem Minimum der Streuung der Residuen führen. Am Datenmaterial aus einem Dauerversuch (Hafer) wird demonstriert, daß mit sachlogischer Begründung ausgewählte Witterungsfaktoren wesentlich wirksamer sind als solche, die nach rein schematischen Gesichtspunkten ausgewählt wurden. Dabei zeigte es sich, daß sechs Witterungsfaktoren im wesentlichen den Ertrag formal bestimmen.

Die Eignung einer Funktion für eine Vorausschätzung wird durch eine Simulation überprüft. Größere Fehlschätzungen von 8—12 dz/ha kommen nur in Jahren mit extrem hohen oder extrem niedrigen Erträgen vor.

Abstract

The authors tried to investigate the relations between weather and yield by means of multiple regressions. By using the method of backward elimination a set of 6 regressors (out of 15 measurements of atmospheric influence) was selected, which was sufficient to estimate the yield. Variables chosen by an expert are more efficient than those selected schematically. It was possible to test the predictive value of the regression-functions by calculating the yield of years whose measurements of atmospheric influence were not included in order to estimate the regression-equations. The estimated values differed 0.8—1.2 t/ha from the observed yields only in years with extreme high or low yields. It is remarkable that extreme high temperature (20—21° C) in the first 10 days of July improve the yield.

1. Einleitung

Seit Holdefleiss (9) sind die verschiedensten Methoden zur Anwendung gekommen, um den Beziehungen zwischen Witterungsverlauf und Ernteertrag nachzugehen. Dabei spielte die Suche nach kritischen Perioden, die durch ihren Wetterablauf für den Ertrag (Pirotta [13], Brouwer [3]) oder für ihre Ausbildung der Ertragskomponenten (Zillmann [20]) im wesentlichen bestimmend sein sollten, eine besondere Rolle. Die Untersuchungen von Baumann (1) über die Methode signifikanter Witterungsdifferenzen zwischen Jahren mit guten und schlechten Erträgen ließen jedoch erkennen, daß jeder Abschnitt der Vegetationsperiode einen mehr oder weniger starken Einfluß auf den Ertrag nehmen kann. Da Gebiete oder Jahre mit höchsten Flächenerträgen ungefähr die optimalen Klimawerte bzw. den optimalen Witterungsverlauf für einzelne Nutzpflanzen zeigen, führen Abweichungen des Klimas bzw. des Jahreswetters von diesen Optimalwerten zu Ertragsminderungen. Thran (15) (16) errechnete daraus den „Wetterschadensfaktor“, der in Zusammenhang mit dem Ertrag (Wetterwirkungskurve) steht. Es liegt auch ein Einfluß der Witterung des Winters und des Vorjahres auf Arten vor, die erst im Frühjahr bestellt wurden. Es gibt eine nachweisbare, z. T. recht lang anhaltende Einwirkung

der Witterung auch auf den Boden, vor allem wohl, weil die Mobilisierung der bodenbürtigen Nährstoffe von der Witterung abhängt.

Da Pfau (12) eingehend die seitherigen Methoden der Wetterertragsstatistik diskutiert hat, ist es nicht erforderlich, im Rahmen dieser Arbeit näher darauf einzugehen. Bemerkenswert ist jedoch in methodischer Hinsicht die von Pfau durchgeführte Bereinigung der Ertragszahlen der Landkreise der Bundesrepublik Deutschland durch Berechnung eines sowohl zeitlich als auch geographisch variablen Trends. Auf diese Weise gelingt es, über Varianzanalysen Jahre mit ertragsmäßig günstigen bzw. ungünstigen Witterungsbedingungen festzulegen. Während Baumann (1) diejenigen Witterungsmerkmale feststellt, die signifikante Differenzen zwischen Jahren mit guten, mittleren und schlechten Erträgen aufweisen, geht Pfau noch einen Schritt weiter, in dem er diese einzelnen Witterungsmerkmale mittels diskriminanzanalytischer Linearkombination zu Gesamtmaßen verbindet. An Hand dieser Gesamtmaße kann festgestellt werden, ob ein zu prüfendes Jahr hinsichtlich der klimatischen Bedingungen eine gute, mittlere oder schlechte Ernte erwarten läßt.

Wenn aus den bisherigen Erfahrungen festzustehen scheint, daß der Ernteertrag von der Witterung während der ganzen Vegetationsperiode abhängt, war es naheliegend, die Ergebnisse eines Dauerversuches mit der Methode der multiplen Regression zu untersuchen, zumal Thran (17) deutlich zum Ausdruck gebracht hat, daß bei jedem Anbau in der Praxis die komplexe Wirkung des Wetters nicht außer Acht gelassen werden dürfe.

2. Untersuchungsmaterial

Den Untersuchungen liegen die mittleren Erträge der Volldüngungsparzellen mit Stallmist von 36 Jahren in $\text{dz/ha} \cdot 10$ bezogen auf 86% Trockensubstanz des Dikopshofer Dauerversuches für Hafer aus dem Zeitraum von 1906 bis 1950 zu Grunde sowie die täglichen Aufzeichnungen von Temperatur und Niederschlag, aus denen die Einflußgrößen ermittelt wurden.

3. Methodik

Snedecor (14) warnt in seinem Methodenbuch sehr eindringlich vor einer bedenkenlosen Anwendung multipler Regressionsanalysen. Es habe keinen Sinn, alle denkbar möglichen Einflußgrößen einzubeziehen, um dann irgend einen Aufschluß über die Wirkung einzelner Faktoren im Sinne einer Kausalanalyse zu erhalten. Nur vom Versuchsansteller mit sachlogischer Begründung ausgewählte Einflußgrößen können die Voraussetzung für eine sinnvolle Interpretation der numerischen Ergebnisse bilden (s. a. Kendall und Stuart [10]).

Aus diesem Grunde wurden zunächst die Witterungsdaten als Einflußgrößen genommen, die nach der Methode signifikanter Witterungsdifferenzen (Baumann [1]) aufgefunden wurden. Da anzunehmen ist, daß die einzelnen Witterungsfaktoren einen unterschiedlichen Anteil an der Höhe des multiplen Bestimmtheitsmaßes haben, ist es zweckmäßig, alle diejenigen Einflußgrößen auszuschneiden, die nicht wesentlich die multiple

Bestimmtheit herabsetzen. Moderne elektronische Rechenanlagen, wie z. B. die X1, legen es nahe, über ein Programm mit Reduktionsverfahren laufend diejenigen Einflußgrößen zu eliminieren, die nicht wesentlich zur Erhöhung des multiplen Bestimmtheitsmaßes beigetragen haben (Weber [19]).

Auf die in der Literatur beschriebenen Verfahren zur Auswahl von Einflußgrößen sei hier nicht näher eingegangen. Es sei insbesondere auf die Arbeiten von Hamaker (7) und Moore (11) hingewiesen.

Das Verfahren sei im folgenden dargestellt. Darin bedeuten:

y_j für $j = 1, 2, 3, \dots, N$ die beobachteten Werte der Erträge,

\hat{y}_j für $j = 1, 2, 3, \dots, N$ die aus der multiplen Regressionsgleichung geschätzten Werte der Erträge,

x_i für $i = 1, 2, 3, \dots, p$ die Einflußgrößen.

Die multiple Regressionsgleichung lautet:

$$\hat{y}_j = \bar{y} + b_{y1.2,3,\dots,p} (x_{1j} - \bar{x}_1) + b_{y2.1,3,\dots,p} (x_{2j} - \bar{x}_2) + \dots + b_{yp.1,2,3,\dots,p-1} (x_{pj} - \bar{x}_p),$$

oder

$$\hat{y}_j = b_{y.1,2,3,\dots,p}^* + b_{y1.2,3,\dots,p}^* \cdot x_{1j} + b_{y2.1,3,\dots,p}^* \cdot x_{2j} + \dots + b_{yp.1,2,3,\dots,p-1}^* \cdot x_{pj};$$

$b_{y.1,2,3,\dots,p}^*$ ist die additive Konstante und z. B. $b_{y1.2,3,\dots,p}^*$ der partielle Regressionskoeffizient, der die Regression zwischen y und der erklärenden Variablen x_1 , bereinigt vom Einfluß der übrigen Variablen x_2, x_3, \dots, x_p wiedergibt.

Im folgenden verwenden wir dafür die abgekürzte Schreibweise

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_p x_p.$$

Die Koeffizienten b_i für $i = 0, 1, 2, 3, \dots, p$ werden nach der Methode der kleinsten Quadrate so bestimmt, daß die Summe der Abweichungsquadrate von

$$\sum_{j=1}^N (y_j - \hat{y}_j)^2$$

zum Minimum wird. Die Koeffizienten b_i werden zweckmäßiger Weise über die Elemente der invertierten Matrix der Korrelationskoeffizienten berechnet, da aus der einmal vorhandenen Kehrmatrix ohne weiteres sofort die um eine Einflußgröße reduzierte Kehrmatrix gebildet werden kann (Goulden [6], Snedecor [14]). Wenn c_{ij} (für $i, j = 1, 2, 3, \dots, p$) die Elemente dieser Kehrmatrix sind, so sind die neuen Elemente der um eine Variable reduzierten Kehrmatrix

$$c_{ij}^* = c_{ij} - c_{iq} \cdot c_{qj} / c_{qq}, \quad [1]$$

wobei der Index q sich auf die zu streichende Variable bezieht.

Um diesen Vorteil zu nutzen, bezieht man die vorhandenen Einflußgrößen in die multiple Regression ein, wobei der Stichprobenumfang N größer als die Anzahl p der Einflußgrößen sein muß und streicht laufend

diejenigen Variablen, durch die der Ausdruck $\sum_{j=1}^N (y_j - \hat{y}_j)^2$ am geringsten erhöht wird, so daß das multiple Bestimmtheitsmaß $B_{y.1,2,3,\dots,p}$ am geringsten sinkt. Da die Werte der Einflußgrößen x_{ij} als repräsentative Zufallsstichproben aus normal verteilten multivariaten Grundgesamtheiten betrachtet werden können („correlation model“ nach Ezekiel und Fox [4]) ist die jeweils zu streichende Variable, die diese Forderung erfüllt, diejenige, deren partielles Bestimmtheitsmaß

$$B_{y_{i.1,2,3,\dots,(i-1),(i+1),\dots,p}} = c_{yi}^2 / (c_{ii} \cdot c_{yy}) \quad [2]$$

den geringsten Betrag hat.

Der Ablauf eines Reduktionsprogrammes gestaltet sich so, daß nach dem ersten Durchlauf folgende Varianzanalyse (VA) ausgedrückt wird:*)

Tab. 1
V A der multiplen linearen Regression

Ursache	SAQ	FG	var	Streuung s
total (Zielgröße)	$\sum_{j=1}^N (y_j - \bar{y})^2 = SAQ_y$	N-1	$SAQ_y / (N-1)$	$\sqrt{\text{var}_y}$
auf der Regression	$\sum_{j=1}^N (\hat{y}_j - \bar{y})^2 = SAQ_{\text{Regr}}$	p	SAQ_{Regr} / p	$\sqrt{\text{var}_{\text{Regr}}}$
um die Regression (Residuen)	$\sum_{j=1}^N (y_j - \hat{y}_j)^2 = SAQ_{\text{Res}}$	N-p-1	$SAQ_{\text{Res}} / (N-p-1)$	$\sqrt{\text{var}_{\text{Res}}}$

multiple Bestimmtheit $B_{y.1,2,3,\dots,p} = SAQ_{\text{Regr}} / SAQ_y$,
 $F = \text{var}_{\text{Regr}} / \text{var}_{\text{Res}}$ mit $n_1 = p$ und $n_2 = N - p - 1$ FG

Das multiple Bestimmtheitsmaß ist der Anteil der durch die Einflußgrößen erklärten Summe der Abweichungsquadrate (= SAQ_{Regr}) an der gesamten Summe der Abweichungsquadrate (= SAQ_y). Das Varianzverhältnis $F = \text{var}_{\text{Regr}} / \text{var}_{\text{Res}}$ gibt somit gleichzeitig an, ob die multiple Bestimmtheit für eine vorgegebene Irrtumswahrscheinlichkeit signifikant von Null verschieden ist.

Anschließend wird vom Programm her die zu eliminierende Einflußgröße x_i aufgesucht, d. h. diejenige, deren partielles Bestimmtheitsmaß zu der Zielgröße y hin den geringsten Betrag hat (s. Gleichung [2]).

Der Index i dieser zu streichenden Variablen wird jetzt mit q bezeichnet und in die Gleichung [1] eingesetzt. Die Kehrmatrix wird um diese q te Zeile und Spalte reduziert. Daraufhin wird sofort der Index q der eliminierten Einflußgröße und die entsprechende Varianzanalyse von Tab. 1 ausgedrückt. Dieses Verfahren wiederholt sich, bis nur noch eine Variable übrigbleibt, bzw. man bricht es ab, sobald die Differenz zwischen zwei auf-

*) Alle Berechnungen wurden auf der X1 des Rechenzentrums der Universität Kiel, Leiter Doz. Dr. B. Schlender, nach Programmen von E. Weber durchgeführt.

einanderfolgenden Bestimmtheitsmaßen für eine vorgegebene Irrtumswahrscheinlichkeit signifikant ist.

Die Prüfung, ob durch Eliminierung einer Einflußgröße im Laufe einer Reduktion das multiple Bestimmtheitsmaß signifikant erniedrigt wurde, erfolgt über eine Varianzanalyse der Ausdrücke von $\sum_{j=1}^N (y_j - \hat{y}_j)^2$ vor der Eliminierung (= SAQ_{Res}) und nach der Eliminierung (= SAQ_{Res}^*) (s. Tab. 2 nach Snedecor [14]).

Tab. 2
VA der SAQ_{Res}

Ursache	SAQ	FG	var
nach der Eliminierung	SAQ_{Res}^*	$N-(p-1)-1$	$SAQ_{Res}^*/(N-(p-1)-1) = var_{Res}^*$
vor der Eliminierung	SAQ_{Res}	$N-p-1$	$SAQ_{Res}/(N-p-1) = var_{Res}$
Differenz	$SAQ_{Res}^* - SAQ_{Res}$	1	$(SAQ_{Res}^* - SAQ_{Res})/1$

$F = (SAQ_{Res}^* - SAQ_{Res}) \cdot (N-p-1) / (SAQ_{Res})$ mit $n_1 = 1$ und $n_2 = N-p-1$ FG.

Da $n_1 = 1$, ist $\sqrt{F} = t$ mit $N-p-1$ FG.

Da bei der successiven Eliminierung immer diejenige Einflußgröße ausgeschlossen wird, die den Ausdruck $\sum_{j=1}^N (y_j - \hat{y}_j)^2$ am geringsten erhöht,

kann die Differenz $SAQ_{Res}^* - SAQ_{Res}$ zumindestens bei Beginn einer Reduktionsfolge kleiner sein als var_{Res} . In diesem Falle muß die Streuung der Residuen (= s_{Res}) in Abhängigkeit von der Anzahl p der jeweils verbliebenen Einflußgrößen so lange sinken, bis die Differenz $SAQ_{Res}^* - SAQ_{Res}$ größer als var_{Res} wird. Es wird also nach Erreichen eines Minimums der Streuung der Residuen bei einer bestimmten Anzahl von Einflußgrößen (= p_{min}) wieder ein Anstieg der Streuung der Residuen erfolgen (s. Abb. 1).

Weiterhin folgt daraus, daß SAQ_{Res} , das zu dem Minimum der Streuung der Residuen gehört, nicht signifikant gegenüber dem vorhergehenden erhöht wurde (F-Werte < 1 , s. Weber und Brott [18]), d. h. man kann auf alle bis zu p_{min} hin eliminierten Einflußgrößen verzichten, da sie nicht wesentlich die multiple Bestimmtheit herabsetzen. Aus diesem Grunde wurde nur mit denjenigen Einflußgrößen weitergearbeitet, die von den ursprünglich einbezogenen zu einem Minimum der Streuung der Residuen führen.

4. Untersuchungsergebnisse

4.1. Der Ertrag in Abhängigkeit von Witterungsfaktoren, ausgewählt nach den bisherigen Erfahrungen mit dem Hafer

Es wurden folgende 15 Witterungsfaktoren in die multiple Regressionsgleichung einbezogen (Baumann [1]).

Temperaturmittel

- x₁ 2. März- bis 1. April-Dekade
- x₂ 2. März- bis 1. Mai-Dekade
- x₃ 3. April- bis 1. Mai-Dekade
- x₄ 1. Juli-Dekade
- x₅ 3. bis 6. Juni-Pentade (Summe d. \pm Abweichungen v. d. Norm)

Summe der Niederschläge

- x₆ Dez. bis Jan.
- x₇ 3. März- bis 3. April-Dekade
- x₈ Mai
- x₉ Juni
- x₁₀ früh., mittl., spät. Frühlingsanfang
- x₁₁ Abw. von 133 mm, Mai—Juni
- x₁₂ Juli—August

Temperatur

- x₁₃ Abw. vom opt. Mittel im März (6.1 °)
- x₁₄ Abw. vom opt. Mittel im Mai (13.5 °)
- x₁₅ Mittel der + Abweichungen d. Pentade v. d. Norm im Juni

Abb. 1 (kurzgestrichelte Linie) zeigt die Abnahme der multiplen Bestimmtheit und den Verlauf der Streuung der Residuen entsprechend dem vorher beschriebenen Eliminierungsverfahren. Das Minimum der Streuung der Residuen liegt bei $p_{\min} = 6$ Einflußgrößen. Bis zu diesem Minimum hin wurden die Witterungsfaktoren in der Reihenfolge x₁₄, x₁₁, x₃, x₁₀, x₁₂, x₅, x₁₃, x₈ und x₂ eliminiert. Sie reduzieren die Summe der Abweichungsquadrate, die durch die Einflußgrößen erklärt wird (= SAQ_{Regf}), nicht wesentlich. Es bleiben also die sechs Einflußgrößen x₁, x₄, x₆, x₇, x₉ und x₁₅ zurück, aus denen die multiple Regressionsgleichung in Tab. 3 berechnet wurde. Die multiple Bestimmtheit aus der VA besagt, daß diese sechs Witterungsfaktoren den Ertrag zu 59.4% formal erklären. Auf Grund der Rangfolge der partiellen Korrelationskoeffizienten ist formal der Einfluß von x₁ am wichtigsten und der von x₇ am geringsten. Von einer in gleicher Weise durchgeführten Kausalinterpretation wollen die Verfasser jedoch auf Grund der Erfahrungen anderer Autoren (z. B. Kendall und Stuart [10]) und im Blick auf diesen Problemkomplex vorerst absehen.

Tab. 3

Die multiple Regressionsgleichung mit p = 6 Einflußgrößen und Varianzanalyse.

Einflußgröße	part. Regr. Koeffizient	Streuung d. Koeff.	Stand. part. Regr. Koeff.	t-Wert* (FG=29)	r part. r _{yi...}	r einf. r _{yi}
0	+ 130.216	—	—	—	—	—
x ₁	+ 2.43544	0.61231	+ 0.515	3.977	+ 0.594	+ .472
x ₄	+ 1.30129	0.45833	+ 0.351	2.839	+ 0.466	+ .416
x ₆	— 0.10333	0.03239	— 0.390	3.189	— 0.510	— .343
x ₇	+ 0.07745	0.04865	+ 0.211	1.591	+ 0.283	+ .057
x ₉	— 0.08072	0.03844	— 0.294	2.100	— 0.363	— .016
x ₁₅	— 0.31765	0.16825	— 0.255	1.888	— 0.331	— .212

*) t_{0.05} = 2.045

Reihenfolge der Eliminierungen der Einflußgrößen x_i für $i =$

—	23	24	20	2	7	4	8	22	25	10	13	21	28	12	27	16	3	33	11	31	5	6	9	32	29	15	26	1	14	17	18	19	30
---																	15	14	9	10	1	11	3	7	6	8	13	5	12	2	4	4	4
----																			14	11	3	10	12	5	13	8	2	7	9	15	6	4	1

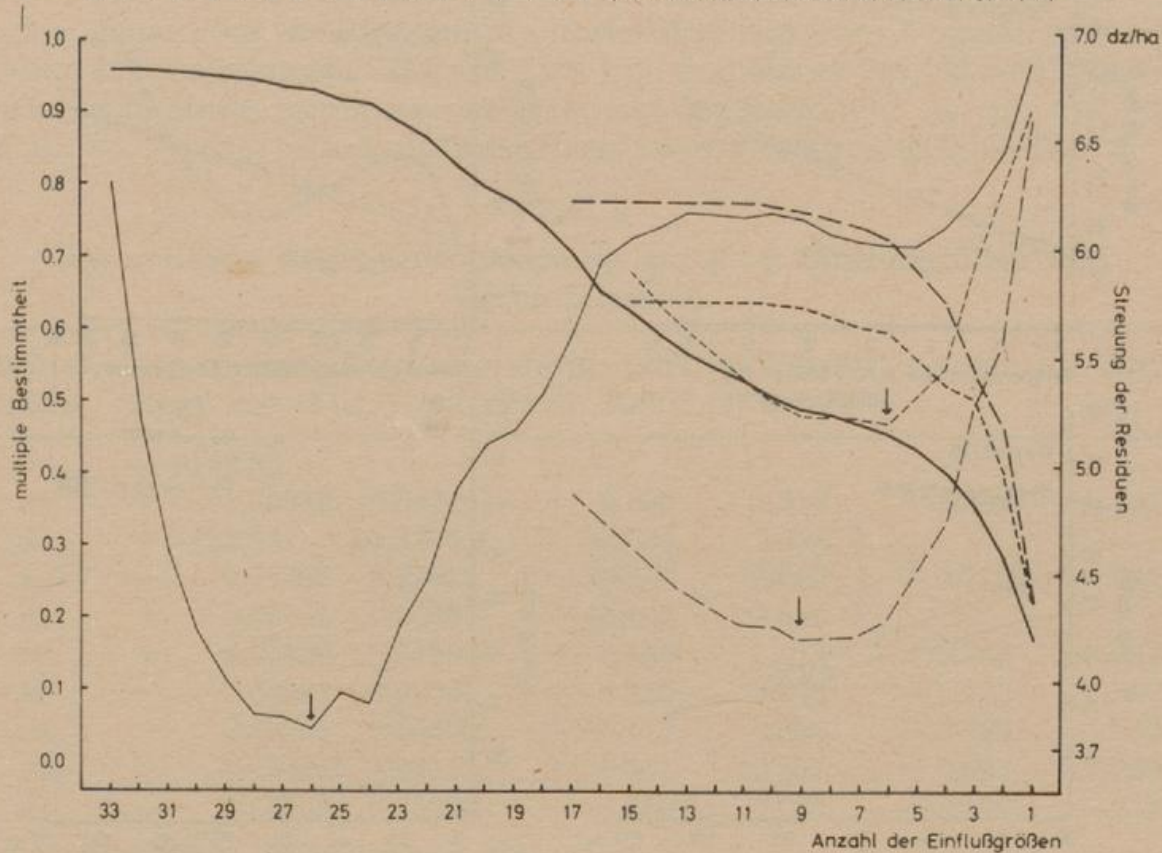


Abb. 1

Verlauf der multiplen Bestimmtheit und der Streuung der Residuen bei dem Eliminierungsverfahren.

Zeichenerklärung:

- mult. B. — Streuung d. Res. (Witterungsdaten schematisch ausgewählt)
 - mult. B. --- Streuung d. Res. (Witterungsdaten nach Baumann 1962)
 - · - mult. B. - · - Streuung d. Res. (Witterungsdaten nach Baumann 1962)
- ↓ pmin

VA

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	114590	6	19098.333	7.075 ($F_{0.05} = 3.81$)
um die Regression (Residuen)	78284	29	2699.448	

mult. B = $114590/192874 = 0.594$

Aus Abb. 2 (s. S. 10) ist die Abhängigkeit des Ertrages von den sechs ausgewählten Witterungsfaktoren zu ersehen.

Ein Linearitätstest ergibt, daß $y = f(x_4)$ am besten durch ein Polynom 3. Grades approximiert wird, während bei den übrigen nur eine lineare bzw. keine signifikante Beziehung vorliegt.

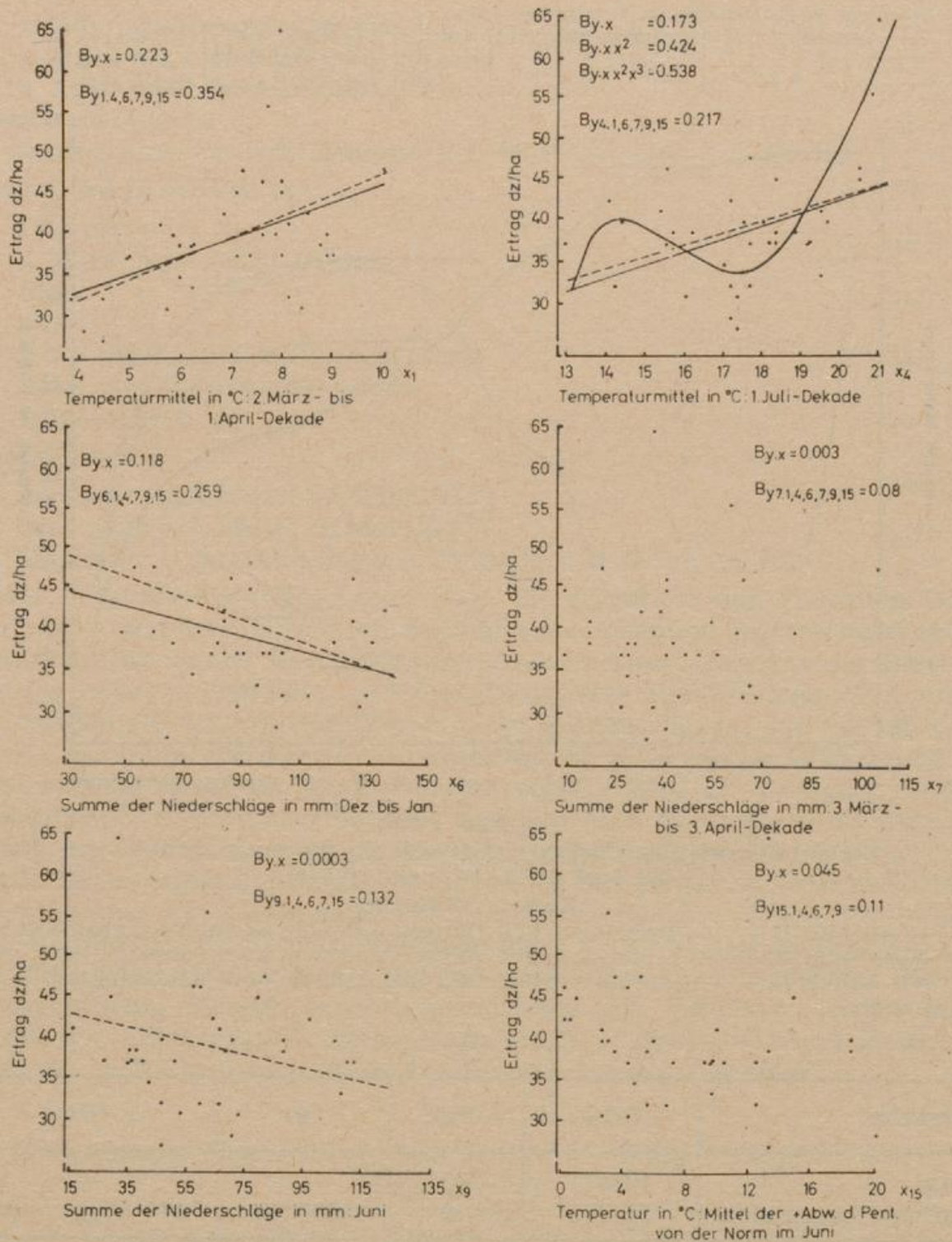


Abb. 2

Die Beziehung zwischen den sechs ausgewählten Witterungsfaktoren und dem Ertrag. (Regressionsgeraden sind nur eingezeichnet, wenn sie auf dem 5%-Niveau signifikant von Null verschieden sind.)*)

Zeichenerklärung: — einfache Regression — partielle Regression

*) Die Punktdiagramme wurden mit einem Zeichenprogramm von U. Rempe über die XI angefertigt.

4.2. Der Einfluß des Witterungsfaktors x_4 (mittlere Temperatur der 1. Juli-Dekade) auf die Erhöhung der multiplen Bestimmtheit

Entsprechend der guten Approximation durch ein Polynom 3. Grades bei x_4 in Abb. 2 wurde in einem weiteren Eliminierungsverfahren von den vorher genannten 15 Witterungsfaktoren und den zwei Transformationen x_4^2 und x_4^3 ausgegangen. Wie aus Abb. 1 zu entnehmen ist, führten folgende Variable zu einem Minimum der Streuung der Residuen: $x_2, x_4, x_4^2, x_4^3, x_5, x_6, x_8, x_{12}, x_{13}$. Ihre multiple Regressionsgleichung enthält Tab. 4.

Tab. 4

Die multiple Regressionsgleichung mit $p=9$ Einflußgrößen und Varianzanalyse.

Einflußgröße	partieller Regr. Koeff.	Streuung d. Koeff.	stand. part. Regr. Koeff.	t-Wert* (FG=26)	r part. $r_{yi \dots}$	r einf. r_{yi}
0	- 6142.62	—	—	—	—	—
x_2	+ 1.36579	0.54244	+ 0.330	2.518	+ 0.443	+ .261
x_4	+ 122.641	51.3205	+ 33.092	2.390	+ 0.713 ¹⁾	+ .732
x_4^2	- 0.77427	0.29686	- 72.848	2.608		
x_4^3	+ 0.00162	0.00057	+ 40.405	2.849		
x_5	+ 0.05886	0.04260	+ 0.160	1.272	+ 0.242	+ .057
x_6	- 0.03827	0.04041	- 0.125	0.947	- 0.183	+ .038
x_8	- 0.03259	0.04072	- 0.117	0.800	- 0.155	- .140
x_{12}	+ 0.00714	1.09113	+ 0.001	0.007	+ 0.001	- .243
x_{13}	- 0.13433	0.15260	- 0.108	0.880	- 0.170	- .212

*) $t_{0.05} = 2.056$ 1) s. Weber (19)

VA

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	130767	9	14529.667	6.083 ($F_{0.05} = 2.90$)
um die Regression (Residuen)	62107	26	2388.731	

mult. B = $130767/192874 = 0.678$

Die Differenz zwischen den mult. B von Tab. 3 mit 59.4% und von Tab. 4 mit 67.8% ist knapp auf dem 10% Niveau signifikant.

Läßt man jedoch in der letzten multiplen Regressionsgleichung der Tabelle x_4, x_4^2 und x_4^3 fallen, dann erhält man eine multiple Bestimmtheit von nur 34.6%. Die entsprechende VA führt zu einem nicht-signifikanten F-Wert.

4.3. Erweiterung der multiplen linearen Regression durch multiplikative Glieder¹⁾

Da zwischen den einzelnen Witterungsfaktoren Wechselwirkungen anzunehmen sind, ist eine Erweiterung durch multiplikative Glieder angebracht.

¹⁾ Diese Berechnung erfolgte auf Grund eines Diskussionsbeitrages von Herrn Prof. Dr. B. Schneider, Inst. für Dokumentation u. Statistik der Med. Akademie Hannover.

Da in der Regel jedoch der Zahl der damit zu berechnenden Parameter sehr bald durch den Stichprobenumfang eine Grenze gesetzt ist, beschränken die Verfasser sich auf Polynome 2. Grades (s. a. Hoffmann und Dörfel [8]).

Für nur 3 erklärende Variable würde der Ansatz lauten:

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{22} x_2^2 + b_{23} x_2 x_3 + b_{33} x_3^2$$

Von 6 Einflußgrößen ausgehend steigt die Anzahl der Variablen auf $p=27$.

Tab. 5

Die multiple lineare Regressionsgleichung mit Erweiterung durch multiplikative Glieder und Varianzanalyse.

Einflußgröße	partieller Regr. Koeff.	Streuung d. Koeff.	stand. part. Regr. Koeff.	t-Wert*) (FG=8)	r part. $r_{yi \dots}$	r einf. r_{yi}
0	+ 2110.39	—	—	—	—	—
x_1	+ 15.99220	11.94550	+ 3.382	1.339	+ .428	+ .472
x_4	— 27.75150	14.57520	— 7.488	1.904	— .558	+ .416
x_6	— 0.52820	1.04658	— 1.994	0.505	— .176	— .343
x_7	+ 1.40636	1.43364	+ 3.834	0.981	+ .328	+ .057
x_9	— 0.04759	0.62206	— 0.173	0.077	— .027	— .016
x_{15}	+ 2.65794	3.86386	+ 2.135	0.688	+ .236	— .212
$x_1 x_1$	— 0.03490	0.06432	— 0.991	0.543	— .188	+ .443
$x_1 x_4$	+ 0.02811	0.05718	+ 1.195	0.492	+ .171	+ .630
$x_1 x_6$	— 0.00977	0.00556	— 3.510	1.757	— .528	— .038
$x_1 x_7$	+ 0.00328	0.00835	+ 0.670	0.393	+ .138	+ .258
$x_1 x_9$	— 0.01035	0.00550	— 3.346	1.880	— .554	+ .168
$x_1 x_{15}$	— 0.02359	0.02385	— 1.239	0.999	— .330	— .032
$x_4 x_4$	+ 0.07541	0.03779	+ 7.095	1.996	+ .576	+ .448
$x_4 x_6$	+ 0.00205	0.00329	+ 1.419	0.624	+ .215	— .208
$x_4 x_7$	— 0.00675	0.00436	— 3.417	1.549	— .480	+ .141
$x_4 x_9$	+ 0.00008	0.00306	+ 0.054	0.027	+ .009	+ .068
$x_4 x_{15}$	+ 0.00692	0.01230	+ 1.026	0.563	+ .195	— .131
$x_6 x_6$	+ 0.00038	0.00031	+ 2.653	1.209	+ .393	— .295
$x_6 x_7$	— 0.00032	0.00030	— 0.874	1.090	— .360	— .187
$x_6 x_9$	+ 0.00044	0.00030	+ 1.786	1.446	+ .455	— .220
$x_6 x_{15}$	— 0.00038	0.00110	— 0.302	0.350	— .123	— .330
$x_7 x_7$	+ 0.00036	0.00065	+ 1.001	0.543	+ .189	+ .108
$x_7 x_9$	— 0.00039	0.00038	— 1.197	1.039	— .345	+ .066
$x_7 x_{15}$	— 0.00197	0.00132	— 0.868	1.490	— .466	— .157
$x_9 x_9$	+ 0.00037	0.00024	+ 1.925	1.554	+ .481	+ .019
$x_9 x_{15}$	— 0.00107	0.00178	— 0.504	0.601	— .208	— .339
$x_{15} x_{15}$	— 0.00537	0.00598	— 0.904	0.897	— .302	— .179

*) $t_{0.05} = 2.306$

VA

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	179061	27	6631.889	3.841 ($F_{6,05} = 3.08$)
um die Regression (Residuen)	13813	8	1726.625	

mult. B = 0.928

Die Approximation durch diesen Ansatz mit den hinzugezogenen quadratischen und gemischten Gliedern ist gut, wie vor allem Abb. 3 erkennen läßt. Der F-Wert der VA ist auf dem 5%-Niveau signifikant.

Um zu prüfen, ob diese hohe Beziehung zwischen Ertrag und Witterungsfaktoren tatsächlich vorhanden ist, oder ob sie nur formal auf die zum Stichprobenumfang mit $N=36$ relativ hohe Anzahl von Variablen mit $p=27$ zurückzuführen ist, wurden die einzelnen Erträge zufällig vertauscht. Die VA aus der multiplen Regression mit den auf die Witterungsfaktoren zufällig zugeteilten Erträgen führt zwar zu einem mult. B von 0.849. Ihr F-Wert ist aber nicht einmal auf dem 20%-Niveau signifikant. Dieses Ergebnis kann als Hinweis für einen Zusammenhang kausaler Art zwischen dem Ertrag und diesen sechs ausgewählten Witterungsfaktoren gedeutet werden.

VA mit zufällig vertauschten Werten der Zielgröße

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	163711	27	6063.370	1.663 ($F_{0,20} = 1.76$)
um die Regression (Residuen)	29163	8	3645.375	—

mult. B = 0.849

4.4. Der Ertrag in Abhängigkeit von schematisch ausgewählten Witterungsdaten.

Baumann (1) erwähnt, daß ein schematischer Vergleich von Witterungsdaten bei der Methode signifikanter Witterungsdifferenzen zwischen Jahren mit guten und schlechten Erträgen meist nicht zu Ergebnissen führt. Es liegt daher nahe, als Einflußgrößen in die multiple Regression einmal nur schematisch ausgewählte Witterungsfaktoren einzubeziehen, d. h. eine dekadenweise Folge von Temperatur- und Niederschlagsdaten von Februar bis August. Es wurden hierfür folgende 33 Witterungsfaktoren herangezogen:

Beobachtungen und Schätzwerte von Hafererträgen
(Dauerversuch Dikopshof)

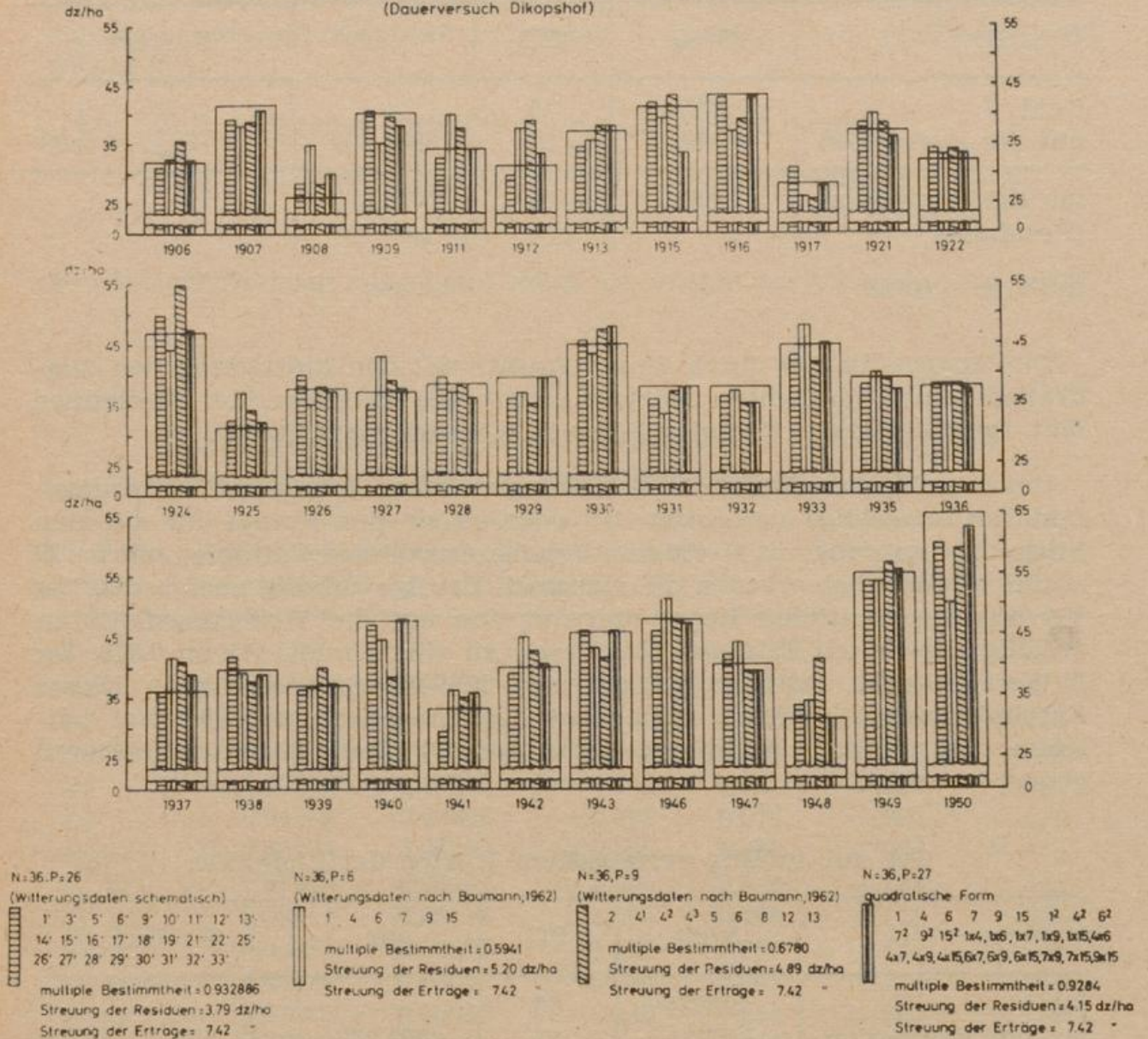


Abb. 3

Vergleich der beobachteten Erträge (nicht schraffierte Blöcke)
mit geschätzten Erträgen.

Summe der Niederschläge in mm

Temperaturmittel in °C

- x₁' 1. März-Dekade
- x₂' 2. März-Dekade
- x₃' 3. März-Dekade
- x₄' 1. April-Dekade
- x₅' 2. April-Dekade
- x₆' 3. April-Dekade
- x₇' 1. Mai-Dekade
- x₈' 2. Mai-Dekade
- x₉' 3. Mai-Dekade

- x₁₇' Februar
- x₁₈' 1. März-Dekade
- x₁₉' 2. März-Dekade
- x₂₀' 3. März-Dekade
- x₂₁' 1. April-Dekade
- x₂₂' 2. April-Dekade
- x₂₃' 3. April-Dekade
- x₂₄' 1. Mai-Dekade
- x₂₅' 2. Mai-Dekade
- x₂₆' 3. Mai-Dekade

Summe der Niederschläge in mm

Temperaturmittel in °C

x_{10}' 1. Juni-Dekade
 x_{11}' 2. Juni-Dekade
 x_{12}' 3. Juni-Dekade

x_{13}' 1. Juli-Dekade
 x_{14}' 2. Juli-Dekade
 x_{15}' 3. Juli-Dekade

x_{16}' 1. August-Dekade

x_{27}' 1. Juni-Dekade
 x_{28}' 2. Juni-Dekade
 x_{29}' 3. Juni-Dekade

x_{30}' 1. Juli-Dekade
 x_{31}' 2. Juli-Dekade
 x_{32}' 3. Juli-Dekade

x_{33}' 1. August-Dekade

Aus Abb. 1 ist das Minimum der Streuung der Residuen bei $p=26$ zu entnehmen. Die Einflußgrößen x_2' , x_4' , x_7' , x_8' , x_{20}' , x_{23}' und x_{24}' bewirken also keine wesentliche Erhöhung der SAQ der Residuen. Die Varianzanalysen der einzelnen Eliminierungsschritte weisen jedoch erst ab $p=29$ für $P=5\%$ signifikante F-Werte auf.

Tab. 6

Die multiple Regressionsgleichung mit $p=26$ Einflußgrößen (schematisch) und Varianzanalyse.

Einflußgröße	partieller Regr. Koeff.	Streuung d. Koeff.	stand. part. Regr. Koeff.	t-Wert* (FG=9)	r part. $r_{yi} \dots$	r einf. r_{yi}
0	+ 1597.5	—	—	—	—	—
x_1'	— 0.73836	0.13563	— 1.02050	5.44	— .876	— .106
x_3'	+ 0.49022	0.11468	+ 0.78352	4.27	+ .819	— .077
x_5'	— 0.57602	0.11327	— 1.22770	5.09	— .861	— .025
x_6'	+ 1.50000	0.27736	+ 2.48310	5.41	+ .874	— .013
x_9'	— 0.88107	0.16106	— 1.74350	5.47	— .877	+ .214
x_{10}'	— 0.11650	0.05971	— 0.29043	1.95	— .545	— .060
x_{11}'	— 0.42656	0.08047	— 0.93556	5.30	— .870	+ .051
x_{12}'	+ 0.13686	0.05658	+ 0.37252	2.42	+ .628	— .008
x_{13}'	— 0.16304	0.07491	— 0.39218	2.18	— .587	— .221
x_{14}'	+ 0.46920	0.11457	+ 1.40920	4.10	+ .807	+ .218
x_{15}'	+ 0.65146	0.12719	+ 1.57680	5.12	+ .863	— .301
x_{16}'	— 0.12632	0.05530	— 0.29437	2.28	— .606	— .304
x_{17}'	+ 1.70370	0.52997	+ 0.69033	3.21	+ .731	+ .165
x_{18}'	— 3.59500	0.73822	— 1.40440	4.87	— .851	— .238
x_{19}'	+ 6.57300	1.08390	+ 1.93980	6.06	+ .896	+ .303
x_{21}'	+ 1.12710	0.66202	+ 0.29893	1.70	+ .494	+ .409
x_{22}'	+ 0.67907	0.48790	+ 0.20296	1.39	+ .421	+ .144
x_{25}'	+ 0.81271	0.54535	+ 0.28163	1.49	+ .445	— .210
x_{26}'	— 4.39720	0.93655	— 1.35940	4.70	— .843	— .042
x_{27}'	— 2.75080	0.66954	— 0.77570	4.11	— .808	+ .002
x_{28}'	+ 1.49850	0.79277	+ 0.47713	1.89	+ .533	— .173
x_{29}'	— 7.54760	1.37420	— 2.05450	5.49	— .878	— .071
x_{30}'	+ 4.07510	0.78628	+ 1.09960	5.18	+ .865	+ .416
x_{31}'	— 5.80490	1.13560	— 1.54160	5.11	— .862	+ .079
x_{32}'	+ 10.13700	2.02160	+ 3.00620	5.01	+ .858	+ .019
x_{33}'	— 6.55190	1.32010	— 1.64910	4.96	— .856	+ .215

*) $t_{0.05} = 2.262$

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	179929	26	6920.346	4.812 ($F_{0.05} = 2.90$)
um die Regression (Residuen)	12945	9	1438.278	—

mult. B = 0.933

Aus Abb. 1 ist zu entnehmen, daß die sechs Einflußgrößen, die aus der Eliminierung sachlogisch ausgewählter Daten (Baumann [1]) resultieren, zu einer gleich hohen mult. Bestimmtheit führen, wie 15 schematisch ausgewählte Faktoren. Das gilt auch für die multiple Regression mit dem Polynom 3. Grades für die mittlere Temperatur der 1. Juli-Dekade, bei der 9 Variable zu dem selben Ergebnis führen wie 19 schematisch ausgewählte. Diese Gegenüberstellung demonstriert deutlich den Vorteil sachlogisch begründeter Einflußgrößen.

4.5. Zur Frage der Vorausschätzung von Erträgen.

Abb. 3 zeigt den Vergleich zwischen Beobachtung und Schätzung aus den verschiedenen Gleichungen. Die je nach Auswahl und Anzahl der Einflußgrößen zum Teil recht gute Approximation, auch von extrem hohen wie niedrigen Erträgen, legt den Gedanken an eine Vorausschätzung nahe. Die Frage lautet also, mit welcher Genauigkeit kann man mit einer aus der Stichprobe von 36 Jahren gewonnenen multiplen Regressionsgleichung den Ertrag auf Grund der Witterungsfaktoren vorausberechnen? Da über das Jahr 1950 hinaus keine Ergebnisse zu erhalten waren, wurde eine Vorausschätzung simuliert. Das Algol-Rechenprogramm wurde für diese Fragestellung so abgeändert, daß es die Daten je eines Jahres bei der Berechnung der multiplen Regressionsgleichung nicht berücksichtigt, d. h. es wird zuerst eine multiple Regressionsgleichung aus den Stichproben Nr. 2 bis Nr. 36 berechnet und diese auf die Daten der Stichprobe Nr. 1 angewendet. Dann wird Nr. 2 ausgelassen und die multiple Regressionsgleichung aus den Stichproben Nr. 1 und Nr. 3 bis Nr. 36 auf die Daten des zweiten eben ausgelassenen Jahres extrapoliert usw..

Eine Vorausschätzung ist nur sinnvoll, wenn es gelingt, den Ertrag mit befriedigender Genauigkeit an einem bestimmten Termin vor der Ernte zu schätzen. Für dieses Beispiel könnte man das Ende der 1. Juli-Dekade wählen.

Die zuerst berechnete multiple Regression mit den Variablen x_1 , x_4 , x_6 , x_7 , x_9 und x_{15} enthält zwar keine Daten über den 11. Juli hinaus, aber die relativ niedrige multiple Bestimmtheit von 59.4% läßt den Erfolg einer Vorausschätzung sehr fraglich erscheinen, wie andere Berechnungen gezeigt haben. Die schematisch ausgewählten Witterungsdaten lassen ebenfalls keine Vorausschätzung zu, wenn man die Witterungsfaktoren von der 1. August-Dekade bis einschließlich 2. Juli-Dekade eliminiert, wie folgende Aufstellung zeigt:

p	eliminiertes Witterungsfaktor	mult. B	F-Wert der VA	Grenzwert $F_{0.20}$
33	—	0.959	1.406	4.48
32	x_{16}' (1. August-Dekade, N)	0.934	1.324	2.98
31	x_{33}' (1. August-Dekade, T)	0.912	1.334	2.43
30	x_{15}' (3. Juli-Dekade, N)	0.883	1.258	2.13
29	x_{32}' (3. Juli-Dekade, T)	0.846	1.133	1.96
28	x_{14}' (2. Juli-Dekade, N)	0.837	1.279	1.84
27	x_{31}' (2. Juli-Dekade, T)	0.816	1.313	1.75
26	x_{13}' (1. Juli-Dekade, N)	0.800	1.384	1.70
25	x_{30}' (1. Juli-Dekade, T)	0.700	0.935	1.66

Die F-Werte der Varianzanalysen erreichen in keinem Stadium die Signifikanzschwelle von 20%. Besonders auffallend ist der starke Abfall der mult. B von $p=26$ auf $p=25$ Einflußgrößen bei Eliminierung der mittleren Temperatur der 1. Juli-Dekade. Eine Vorausschätzung unter Ausschluß der Witterungsdaten vom 11. Juli bis zum 10. August scheidet auch hier aus.

Es bleibt noch eine Vorausschätzung zu versuchen mit der multiplen Regressionsgleichung, bestehend aus den sechs nach Baumann (1) ausgewählten Einflußgrößen einschließlich ihrer multiplikativen Glieder wie in Abschnitt 4.3. Da jedoch das vorher beschriebene Verfahren sehr zeitraubend ist, wurden aus der multiplen Regressionsgleichung von Tab. 5 die Variablen eliminiert, die zu einem Minimum der Streuung der Residuen führen. Mit dieser in Tab. 7 aufgeführten multiplen Regression wurde die Vorausschätzung simuliert.

Tab. 7

Die multiple Regressionsgleichung mit $p=19$ Einflußgrößen für die Vorausschätzung und Varianzanalyse.

Einfluß- größe	partieller Regr. Koeff.	Streuung d. Koeff.	stand. part. Regr. Koeff.	t-Wert*) FG=16	r part $r_{yi \dots}$	r einf. r_{yi}
0	+1696.12	—	—	—	—	—
x_1	+ 16.6130	4.06927	+ 3.51381	4.08	+ .714	+ .472
x_4	— 22.9093	5.95732	— 6.18167	3.85	— .693	+ .416
x_6	— 1.0760	0.41850	— 4.06292	2.57	— .541	— .343
x_7	+ 2.3633	0.57812	+ 6.44251	4.09	+ .715	+ .057
x_{15}	+ 3.0936	1.28003	+ 2.48446	2.42	+ .517	— .212
$x_1 x_6$	— 0.0073	0.00274	— 2.62173	2.66	— .554	— .038
$x_1 x_9$	— 0.0128	0.00273	— 4.15158	4.71	— .762	+ .168
$x_1 x_{15}$	— 0.0179	0.01078	— 0.94273	1.66	— .384	— .032
$x_4 x_4$	+ 0.0676	0.01619	+ 6.36144	4.18	+ .722	+ .448
$x_4 x_6$	+ 0.0034	0.00173	+ 2.31810	1.94	+ .436	— .208
$x_4 x_7$	— 0.0087	0.00267	— 4.41671	3.27	— .633	+ .141
$x_6 x_6$	+ 0.0004	0.00015	+ 3.05834	3.02	+ .603	— .295
$x_6 x_7$	— 0.0004	0.00016	— 1.06758	2.41	— .516	— .187

Einfluß- größe	partieller Regr. Koeff.	Streuung d. Koeff.	stand. part. Regr. Koeff.	t-Wert*) FG=16	r part $r_{yi...}$	r einf. r_{yi}
$x_6 x_9$	+ 0.0004	0.00013	+ 1.64626	3.08	+ .610	— .220
$x_7 x_9$	— 0.0003	0.00015	— 0.98449	2.12	— .469	+ .066
$x_7 x_{15}$	— 0.0020	0.00079	— 0.88771	2.55	— .538	— .157
$x_9 x_9$	+ 0.0005	0.00012	+ 2.58230	4.18	+ .723	+ .019
$x_9 x_{15}$	— 0.0015	0.00056	— 0.69692	2.63	— .549	— .339
$x_{15} x_{15}$	— 0.0039	0.00337	— 0.65517	1.15	— .277	— .179

*) $t_{0.05} = 2.120$

VA

Ursache	SAQ	FG	var	F-Wert
total	192874	35	—	—
auf der Regression	176641	19	9296.895	9.163 ($F_{0.05} = 2.42$)
um die Regression (Residuen)	16233	16	1014.563	—

mult. B = 0.917

Es fällt auf, daß bei dieser reduzierten multiplen Regressionsgleichung mit $p=19$ nur die t-Werte der Variablen $x_1 x_{15}$, $x_4 x_6$ und $x_{15} x_{15}$ nicht den Grenzwert des 5%-Niveaus überschreiten. In Tab. 6 hingegen mit $p=27$ tritt kein signifikanter t-Wert auf.

Einen Maßstab für die Brauchbarkeit dieser Vorausschätzung kann man erhalten, wenn man die SAQ der Residuen von Tab. 7 mit der Summe der Quadrate der Residuen aus der Vorausschätzung vergleicht.

Tab. 8

Beobachtete Erträge und Residuen aus der multiplen Regression von Tab. 7 und aus der Vorausschätzung in dz/ha.

Jahr	Beobachtung	Residuen aus Tabelle 7	Residuen aus Vorausschätzung
1906	32.1	— 0.806	— 1.246
1907	41.9	+ 1.242	+ 2.301
1908	26.5	— 4.958	— 11.064
1909	39.8	+ 1.719	+ 4.613
1911	33.8	+ 0.215	+ 0.273
1912	31.0	— 1.652	— 2.311
1913	36.7	— 1.511	— 7.843
1915	40.4	+ 7.119	+ 9.730
1916	42.6	+ 0.402	+ 1.244
1917	27.6	+ 0.757	+ 4.609
1921	37.1	— 0.846	— 2.382
1922	31.9	— 1.726	— 3.570
1923	46.7	— 0.859	— 2.834
1924	30.3	— 2.663	— 4.718
1926	37.6	— 0.065	— 0.128

Jahr	Beobachtung	Residuen aus Tabelle 7	Residuen aus Vorausschätzung
1927	36.5	— 0.625	— 0.789
1928	38.0	+ 0.715	+ 0.876
1929	38.8	+ 1.847	+ 2.975
1930	44.6	— 3.623	— 6.246
1931	37.7	+ 0.501	+ 1.283
1932	37.5	+ 2.094	+ 2.880
1933	44.4	— 1.161	— 3.807
1935	39.1	+ 2.303	+ 9.455
1936	37.2	+ 0.897	+ 6.419
1937	36.8	— 1.986	— 2.466
1938	40.2	+ 1.119	+ 6.776
1939	37.0	+ 2.621	+ 3.542
1940	47.8	+ 0.302	+ 4.539
1941	32.4	— 3.574	— 9.825
1942	39.8	+ 0.789	+ 1.023
1943	45.9	— 0.821	— 2.428
1946	47.6	+ 0.495	+ 1.224
1947	39.3	+ 0.518	+ 6.955
1948	31.1	— 0.099	— 1.104
1949	54.9	— 1.269	— 11.852
1950	64.3	+ 2.589	+ 7.985
Summe der Quadrate		16233	103188

Subtrahiert man die Summe der Quadrate der Residuen der Vorausschätzung von der SAQ_y , so erhält man eine SAQ auf der Regression, die der Vorausschätzung zu Grunde liegen würde. Dieser Anteil der SAQ auf der Regression aus der Vorausschätzung an der SAQ_y entspricht der multiplen Bestimmtheit B und soll daher mit B' bezeichnet werden.

SAQ_y	+ 192874
SAQ_{Res} d. Vorausschätzung (s. Tab. 8 unten rechts)	— 103188
SAQ_{Regr} d. Vorausschätzung	= 89686

Für die Vorausschätzung erhalten wir folgende Bestimmtheit:
 $B' = 89686/192874 = 0.465$

Dieses ungenügend erscheinende Ergebnis der Vorausschätzung ist in erster Linie auf die der Erträge folgender Jahre zurückzuführen:

Jahr	Beobachtung in dz/ha	Residuen in dz/ha
1908	26.5	— 11.604
1913	36.7	— 7.843
1915	40.4	+ 9.730
1935	39.1	+ 9.455
1941	32.4	— 9.825
1949	54.9	+ 11.852
1950	64.3	+ 7.985

Von diesen sieben unbefriedigenden Vorausschätzungen mögen drei auf die extrem niedrigen bzw. extrem hohen Erträge aus den Jahren 1908, 1949 und 1950 zurückzuführen sein. Betrachtet man die Ergebnisse in Abb. 4, so ist ein gewisser Erfolg der Vorausschätzung trotz sieben negativer Ergebnisse von insgesamt 36 Jahren zu verzeichnen.

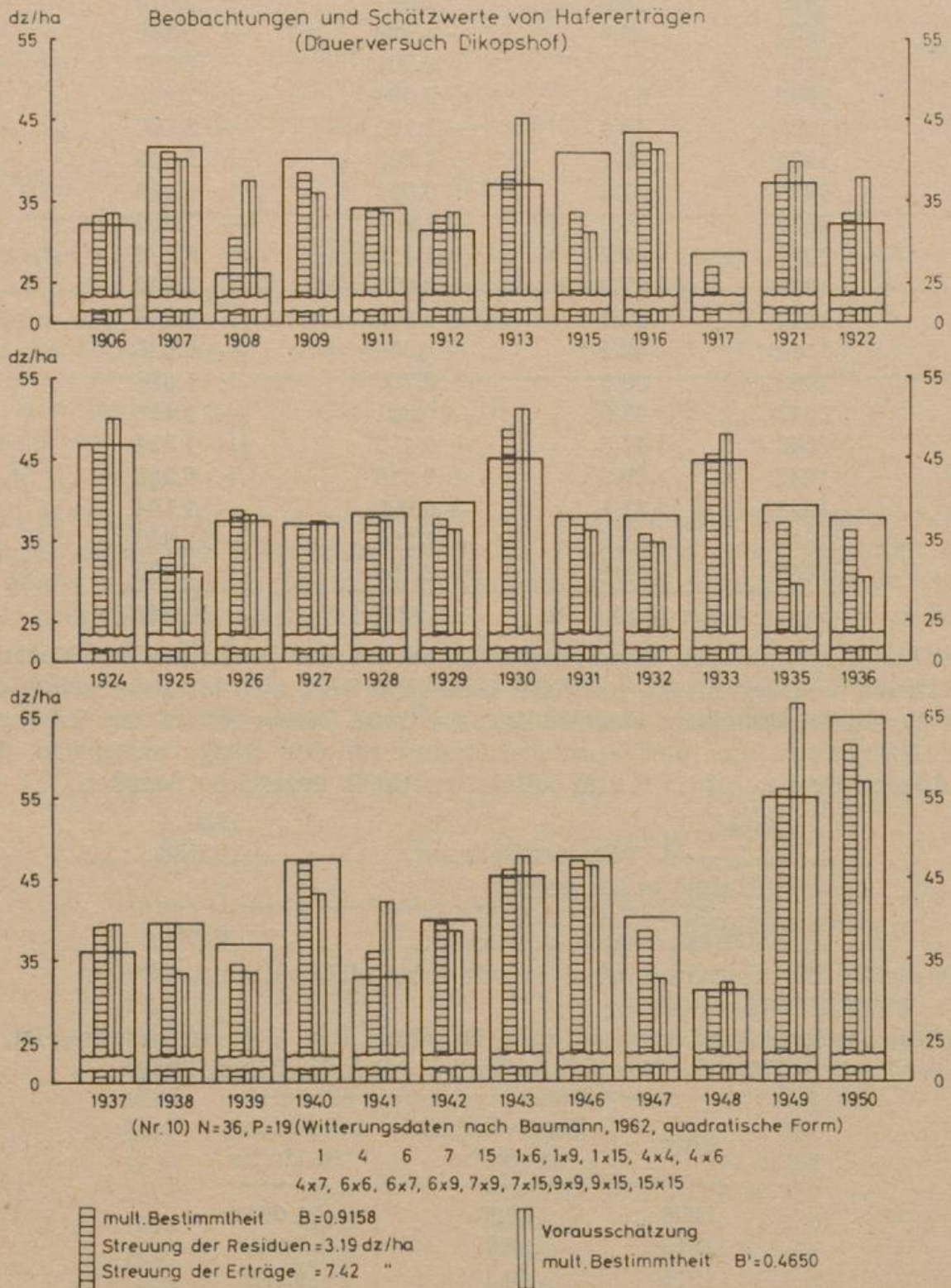


Abb. 4

Vergleich der beobachteten Erträge (nicht schraffierte Blöcke) mit geschätzten Erträgen der multiplen Regression aus Tab. 7 und aus der Vorausschätzung (Tab. 8).

5. Zusammenfassung der wichtigsten Ergebnisse

In der vorliegenden Arbeit wird der Versuch unternommen, das Modell der multiplen linearen Regression auf die Beziehung zwischen Witterungsfaktoren (als Einflußgrößen) und dem Ertrag (als Zielgröße) anzuwenden. Da nicht alle Einflußgrößen im gleichen Maße an der Erklärung der Zielgröße beteiligt sind, werden über ein Eliminierungsverfahren diejenigen Witterungsfaktoren ermittelt, die zu einem Minimum der Streuung der Residuen führen. Am Datenmaterial aus dem Dikopshofer Dauerversuch (Hafer) wird demonstriert, daß mit sachlogischer Begründung ausgewählte Witterungsfaktoren wesentlich wirksamer sind als solche, die nach rein schematischen Gesichtspunkten ausgewählt wurden.

Dabei zeigte es sich, daß von 15 Witterungsmerkmalen (die nach der Methode signifikanter Witterungsdifferenzen zwischen Jahren mit guten und schlechten Erträgen ausgewählt wurden), 6 zum Minimum der Streuung der Residuen führen. Die multiple Bestimmtheit zwischen dem Ertrag und diesen 6 Einflußgrößen ist nicht wesentlich niedriger als die zu allen 15 Variablen. Der Ertrag und die Höhe der Winterniederschläge stehen in negativer Beziehung zueinander (s. Abb. 2). Die mittlere Temperatur der 2. März- bis 1. April-Dekade schwankt in den untersuchten Jahren von $+4$ bis $+10^{\circ}\text{C}$. Höhere Temperaturen in Verbindung zu den Niederschlägen wirken sich in dieser Zeit günstig auf den Ertrag aus. Die einfache und partielle Regression zwischen dem Ertrag und den Temperaturen der mittleren positiven Abweichung der Pentaden von der Norm im Juni ist zwar nicht signifikant von Null verschieden, aber in Verbindung mit den Juni-Niederschlägen tragen sie wesentlich zu der Erklärung des Ertrages, d. h. zur multiplen Bestimmtheit bei. Besonders auffallend ist die positive Wirkung extrem hoher Temperaturen von 20 bis 21°C der 1. Juli-Dekade auf den Ertrag.

Die Möglichkeit der Vorausschätzung wird am Beispiel eines linearen Ansatzes mit multiplikativen Gliedern gezeigt. Die Eignung einer Funktion für eine Vorausschätzung wird durch eine Simulation überprüft, indem zuerst eine Funktion ohne die Daten des ersten Jahres berechnet wird. Diese Funktion wird auf die Daten des ausgelassenen ersten Jahres angewandt und das so ‚vorausgeschätzte‘ Ergebnis mit der Beobachtung verglichen. Als nächstes wird die Funktion aus dem ersten, dritten bis letzten Jahr berechnet und auf die Daten des zweiten Jahres extrapoliert usw. Größere Fehlschätzungen von 8 bis 12 dz/ha kommen nur in Jahren mit extrem hohen oder extrem niedrigen Erträgen vor.

Literatur

- (1) Baumann, H.: Die Erträge von Wintergerste, Hafer und Zuckerrüben im Dikopshofer Dauerdüngungsversuch und in der Kölner Bucht in Beziehung zur Witterung. *Z. Acker- u. Pflanzenbau* 114 (1962) S. 281 bis 294
- (2) v. Boguslawski, E. u. Schneider, B.: Die Anwendung der Parameter in der Mitscherlichschen Ertragskurve bei der Auswertung von Feldversuchen. *Z. Acker- u. Pflanzenb.* 2 (1964) S. 101—109
- (3) Brouwer, W.: Die Beziehungen zwischen Ernte und Witterung in der Landwirtschaft. *Landw. Jahrb.* 63 (1933) S. 1—81
- (4) Ezekiel, M. and Fox, K. A.: *Methods of Correlation and Regression Analysis.* New York 1959
- (5) Fisher, R. A.: *Contributions to Mathematical Statistics.* New York 1950
- (6) Goulden, C. H.: *Methods of Statistical Analysis.* New York 1960
- (7) Hamaker, H. C.: On Multiple Regression Analysis. *Statistica Neerlandica* 16 (1962) S. 31—56
- (8) Hoffmann, E. u. Dörfel, H.: Die funktionale Betrachtungsweise bei der Auswertung von Feldversuchen. *Landw. Versuchs- u. Untersuchungsw.* 2 (1963) S. 75—107
- (9) Holdefleiß, P.: Über den Einfluß der Witterungsfaktoren auf die Ernteerträge. *Kühn-Archiv* 9 (1925) S. 53—78
- (10) Kendall, M. G. and Stuart, A.: *The Advanced Theory of Statistics.* London 1961
- (11) Moore, P. G.: Regression as an Analytical Tool. *Applied Statistics* 2 (1962) S. 106—119
- (12) Pfau, R.: Ein Beitrag zur Wetterertragsstatistik von Halm- und Hackfrucht. *Ber. Dt. Wetterd.* 13 (1964) Nr. 94
- (13) Pirotta: Zit. bei Holdefleiss, *Agrarmeteorologie.* Berlin 1930
- (14) Snedecor, G. W.: *Statistical Methods.* Iowa State University Press 1961
- (15) Thran, P.: Beurteilung der landwirtschaftlichen Produktion mit Hilfe der Wetter-Ertragsrechnung. *Ber. Dt. Wetterd. US-Zone* Nr. 32 (1952) S. 97—98
- (16) Thran, P.: Kurze Erklärung der Wetter-Ertragsrechnung. *Ber. Dt. Wetterd. US-Zone* Nr. 32 (1952) S. 99—101
- (17) Thran, P.: *Agrarmeteorologie und Pflanzenzüchtung. Überblick über agrarmeteorologische Methoden und zukünftige Möglichkeiten.* Vortrag gehalten am 12. Januar 1965 DLG-Tagung in Wiesbaden
- (18) Weber, E. u. Brott, C.: Ein Linearitätstest mit Hilfe elektronischer Datenverarbeitungsanlagen. *Biometr. Z.* 5 (1963) H. 3, S. 188—205
- (19) Weber, E.: Partielle Bestimmtheitsmaße bei Einbezug von Polynomen in multiple lineare Regressionsanalysen. *Biometr. Z.* 6 (1964) S. 262 bis 269
- (20) Zillmann, K. H.: Über die witterungsbedingte Korntragsbildung beim Petkuser Winterroggen. *Habilitationsschrift,* Berlin 1960

