DIE ANWENDUNG MULTIVARIATER STATISTISCHER VERFAHREN AUF DIE ABHÄNGIGKEIT DES LANDWIRTSCHAFTLICHEN ERNTEERTRAGES VOM KLIMA AM BEISPIEL VON NIEDERÖSTERREICH

Josef STROBL, Amstetten

(Mit 11 Textabbildungen)

INHALT

1. EINLEITUNG

Die generelle Bedeutung des Zusammenspiels klimatischer Einflußgrößen und landwirtschaftlicher Ertragsbildung kann vorneweg sicherlich außer Diskussion gestellt werden. Die durch die alljährliche Witterung verursachte Schwankungsbreite in der regionalen wie der weltweiten Nahrungsmittelproduktion wird uns regelmäßig vor Augen geführt. Eine Erfassung der steuernden Relationen kann nunmehr zwei völlig unterschiedliche Wege beschreiten. Auf der einen Seite steht der herkömmliche naturwissenschaftliche Ansatz, der durch Experimente und theoretische Überlegungen die Anwendung bzw. Erstellung physikalischer und chemischer Gesetzmäßigkeiten in

diesem vorgegebenen Forschungsbereich anstrebt. Dem gegenüber steht das sich rasch entwickelnde Instrumentarium der Statistik, das bekanntlich weit über den umgangssprachlichen Begriff im Sinne einer Datenauflistung, -aggregation und Mittelwertsbildung hinausgeht.

Das sind aber nun keineswegs konkurrierende Ansätze, sondern vielmehr sich ergänzende Forschungsstrategien mit unterschiedlichen Ausgangsvoraussetzungen, Anwendungsbereichen und Zielvorstellungen. Indem der Verfasser die erstzitierte Vorgangsweise mit ihrer langen Tradition im physiogeographischen Forschungsfeld als weitgehend bekannte Praxis hinstellt, möchte er kurz einige divergierende Punkte im Vergleich zum Ablauf einer Untersuchung mit statistischer Methodik herausarbeiten.

Zunächst sei als gemeinsames Ziel die numerische Quantifizierung der untersuchten Zusammenhänge und damit die Möglichkeit zu Vorhersagen akzeptiert. Somit stehen auf der einen Seite einer Vorhersagegleichung „alle" Einflußgrößen, auf der anderen Seite das erzielte Ertragsergebnis. Für eine statistische Untersuchung müssen nun Stichproben sowohl der als unabhängig („zufällig aufeinander folgend") betrachteten Eingangsgrößen wie auch der davon gesteuerten Variablen vorhanden sein. Aus diesem Datenmaterial mit seinen noch impliziten Informationen über bestehende (Formal-) Zusammenhänge sollen die Parameter, d. h. die Gewichtungen der Eingangsgrößen geschätzt werden. Hier zeigt sich, daß eine Untersuchung keineswegs mit der Applikation statistischen Instrumentariums allein durchgeführt werden kann. Zu Beginn stehen nicht nur eine ganze Reihe formaler a priori-Annahmen, sondern auch eine (mehr oder weniger bewußte) Basisvorstellung über die Struktur der untersuchten Zusammenhänge. In eben diese Starttheorie (die natürlich im Verlauf der Untersuchung modifiziert werden kann) fließen Fakten und Vorstellungen über ursächliche physische Relationen ein.

Es sind also 2 Abläufe denkbar: Ein formal festgestellter Zusammenhang dient als Grundlage für die Entwicklung von Erklärungsmodellen, oder ein deduktiv erstellter Ansatz wird mit Hilfe des inferenzstatistischen Instrumentariums auf der Grundlage der erwähnten Stichprobe geprüft und (unter einer definierten Wahrscheinlichkeit) akzeptiert oder verworfen. Diese letztere Vorgangsweise entspricht den Basisprämissen statistischer Verfahren, es werden also physische Zusammenhänge postuliert und nur deren formales (!) Zutreffen beurteilt bzw. quantifiziert. Damit ist also zu sehen, daß die herkömmliche naturwissenschaftliche Basis für einen derartigen formalwissenschaftlichen Ansatz unentbehrlich ist. Worin liegen also die Vorteile des Abgehens von einer „rein" naturwissenschaftlichen Arbeitsweise?

Befaßt man sich mit der Ergründung von Abhängigkeitsstrukturen der landwirtschaftlichen Ertragsbildung, so gelangt man bald in allen Dimensionen in Mikromaßstäbe. Damit findet man sich auch plötzlich in Bereichen, die unmittelbar von wenig geographischem Interesse sind, aber auch häufig außerhalb der fachlichen Kompetenz des durchschnittlichen Physiogeographen. Andererseits ist aber auch bald ersichtlich, daß bei den Agrarwissenschaften die Rückführung der erzielten Untersuchungsergebnisse auf größere räumliche und zeitliche Dimensionen sowie Versuche zur Artenaggregation offenbar nicht von primärem Interesse sind.

Gerade hier aber liegt das Arbeitsgebiet einer sich als Raumwissenschaft verstehenden Geographie. Wie in so vielen anderen Fällen erscheint also das Verarbeiten (das strapazierte Stichwort von der Integration ist vielleicht gar nicht so schlecht) von

Detailergebnissen aus analytischen und deskriptiven „Grundwissenschaften" als sinnvoll und von großem praktischen Interesse. Genau hier auch liegt der Grund für die neuerdings fast zwingend erscheinende Ergänzung geographischer Methodik aus dem statistischen Instrumentarium. In räumliche, zeitliche und sachliche Dimensionen, die einer direkten naturwissenschaftlich-experimentellen Untersuchung nicht oder kaum zugänglich sind, werden mit einer Art Analogieschluß Erkenntnisse aus dem mikromaßstäblichen Arbeitsbereich anderer Fachrichtungen übertragen. Die Gültigkeitsprüfung eines solchen Schlusses, die Parameterschätzung des Modells sowie die „Qualitätsprüfung" der Ergebnisse obliegen dem kompetenten Einsatz statistischer Verfahren.

Um diese Gedanken zu veranschaulichen, ein kurzes und einfaches Beispiel: eine Untersuchung auf einer quadratmetergroßen Testfläche einer Winterweizenart mit täglichen Messungen während des gesamten vegetativen und generativen Zyklus habe ergeben, daß der Einfluß der Strahlung in einer bestimmten Entwicklungsphase für den Endertrag von ganz entscheidender Bedeutung ist. Man versucht nun diese Erkenntnis zu verallgemeinern, d. h. in diesem Fall, in übergeordnete Maßstabsbereiche zu übertragen. Das ergäbe also etwa folgende Hypothese: Die Strahlungssumme des Monats, in dem üblicherweise diese entscheidende Entwicklungsphase erreicht wird, hat signifikant positive Auswirkungen auf den Ertrag von Winterweizen (aller Sorten) in einem definierten Anbaugebiet. Es wird also in zeitlicher, räumlicher und sachlicher Hinsicht der Gültigkeitsbereich der Ausgangsaussage überschritten. Es ist vorerst nicht bekannt, ob diese umfassende Aggregation im Datenmaterial noch genügend Zusammenhangsinformation beläßt, um vom formalen Standpunkt die hinterfragte Relation bestätigen zu können. Hier setzt die Anwendung statistischer Methodik ein und prüft die Hypothese, schätzt Modellparameter usw.

Die sich aus diesem umfassenden Skalenwechsel ergebenden Fragen und Probleme sind wohl offensichtlich und begleiten den gesamten Untersuchungsablauf. Sie äußern sich in Einschränkungen und Prämissen, die bei der Beurteilung der Ergebnisse immer berücksichtigt werden müssen, häufig aber ein wenig beachtetes Dasein fristen. Auf Details wird in den folgenden Abschnitten laufend verwiesen. Was war aber das konkrete Ziel der hier referierten Arbeit? (STROBL 1981) Im Verlauf vorhergehender Beschäftigung mit agroklimatischen Fragen trat ein äußerst unbefriedigender Zustand hinsichtlich der Verfahrensauswahl zur Beurteilung der postulierten Zusammenhänge ein. Nur eher oberflächlich geläufige Methoden wurden auf die vorliegende Stichprobe (s. Abschnitt 2.) angewandt und erst während der Untersuchung genauere Kenntnisse über Prämissen und Gesamtkonzept sowie mathematische Basis einzelner Verfahren erworben.

Genau der umgekehrte Weg ist aber für die Beurteilung einer vorliegenden Theorie über untersuchte Zusammenhänge erforderlich, andererseits ersetzen Einblicke in die Konzeption und mathematische Struktur eines Berechnungsablaufs nicht die sich aus der Anwendungspraxis ergebenden Erfahrungen hinsichtlich der Flexibilität, Aussagefähigkeit und praktischen Leistungsfähigkeit eines Verfahrens.

Um bei folgenden Arbeiten diesen divergierenden Forderungen gerecht werden zu können, war als Ziel vorgegeben, die Anwendbarkeit und den möglichen Einsatzbereich im agroklimatischen Fragenkomplex für die „üblichen" multivariaten Verfahren

zu prüfen. Als „üblich" wurde das Vorhandensein in den verbreitetsten statistischen Programmpaketen SPSS und BMDP definiert, multivariat deshalb, weil die komplexen Relationsstrukturen dies erforderten. Eine weitere Eingrenzung der Verfahrensanzahl ergab sich aus der Qualität der vorhandenen Daten, wodurch alle unter dem Intervallniveau operierenden Verfahren einen Informationsverlust bedeuten würden und deshalb nicht berücksichtigt wurden.

Wie sicherlich bereits zu erkennen war, ist eine derartige Arbeit weder als methodische zu bezeichnen noch den konkret problembezogenen Untersuchungen zuzuordnen. Hauptthema ist vielmehr die Schnittstelle zwischen diesen beiden Bereichen, es sollen ja die Anwendungsmöglichkeiten anerkannter Methoden auf einen klar umgrenzten Problembereich erkundet werden.

Nach einer kurzen Darlegung des sachlichen und theoretischen Bezugsrahmens sollen im Hauptabschnitt 5 die als Wichtigste erachteten Verfahren kurz vorgestellt und anhand von Beispielen aus dem verwendeten Datenmaterial einige ihrer möglichen Leistungen für die Agroklimatologie aufgezeigt werden. Soweit möglich bzw. sinnvoll wurde die räumliche Komponente mitberücksichtigt und nicht nur eine unilokale Datenreihe bearbeitet. Zur Veranschaulichung konnten auch einige der mit dem Zeilendrucker graphisch aufbereiteten Darstellungen in diesen Aufsatz einbezogen werden.

2. DATENMATERIAL DER ANWENDUNGSBEISPIELE

Hier soll nicht nur eine räumliche, zeitliche und sachliche Abgrenzung und Begründung der Auswahl der verwendeten Daten gegeben, sondern vor allem auch einige Probleme der Verfahrensanwendung aufgezeigt werden, die sich aus dem wohl selten völlig konsistenten Datenmaterial ergeben. Dies bezieht sich hauptsächlich auf die (ausgenommen im erwähnten Mikromaßstab von Detailuntersuchungen) meist nicht eindeutig lösbare Frage der Gegenüberstellung verfügbarer Klima- und Ertragsdaten.

Niederösterreich wurde als d a s agrarische Bundesland Österreichs und Hauptproduktionsgebiet vieler Feldfrüchte als Untersuchungsgebiet gewählt. Zum Zweck des regionalen Vergleichs wurden Ertragsdaten lt. Statistischem Zentralamt auf Basis der 21 politischen Bezirke (+4 Statutarstädte) herangezogen. Für den Bezirk Amstetten mit seinem agrarischen Schwerpunkt im Alpenvorland wurde eine 29jährige Datenreihe (1946–74), für den Rest der Bezirke die Jahre 1965–74 verwendet. Alternative Datenreihen würden auf der Basis der Kleinproduktionsgebiete (lt. Landes-Buchführungs-Gesellschaft) existieren, wegen besserer Konsistenz und Differenzierung der Daten, Erfaßbarkeit und Vergleichbarkeit wurde aber trotz der räumlichen Inhomogenität den Daten nach politischen Bezirken der Vorzug gegeben.

Es wurden nun Erträge für Getreide, Hackfrüchte und Futterpflanzen erhoben, um ein etwas gestreutes Spektrum an Feldfrüchten bearbeiten zu können. Die Differenzierung der Feldfruchtarten innerhalb dieser Gruppen entspricht der in den Datenquellen, es wurde keine weitere Aggregation vorgenommen.

In diesen z. T. bis in die unmittelbare Nachkriegszeit zurückreichenden Ertragswerten ist natürlich eine Zunahme zu beobachten, die nicht etwa eine Klimaänderung indiziert, sondern auf Verbesserungen bei Sortenwahl, Bearbeitung, Düngung usw.

zurückzuführen ist. Da aber diese Fakten nicht Gegenstand der vorliegenden Untersuchung sind, wurden sie unter der Annahme eines linearen Trends subsummiert und dieser aus den Daten entfernt. Das heißt, in alle weiteren Berechnungen gingen nur mehr die Residuen einer linearen Einfachregression jeder Ertragsvariablen auf die Zeit ein. Inwieweit es in diesem und in einigen anderen Fällen sinnvoll ist, einen linearen Ansatz durch eine nichtlineare Funktion zu ersetzen, ist Gegenstand einer Reihe von Untersuchungen und auch theoretischer Überlegungen. Zu Beginn steht hier die Frage nach einer adäquaten funktionellen Form. Eine Möglichkeit dabei wäre, sich an bereits erstellte Konzepte zu halten, wie etwa den „abnehmenden Ertragszuwachs" nach MITSCHERLICH. Demgegenüber steht die Vorgangsweise, gegen die Zeitachse die Ertragswerte aufzutragen und aus dem Spektrum „gängiger" Funktionsformen die mit der besten Anpassungsleistung auszuwählen.

Diese Strategie führt sicherlich zu einer Verringerung der Residualsummen, wird aber von vielen wegen der schwachen theoretischen Fundierung abgelehnt. Außerdem reagieren viele nichtlineare Funktionen sehr flexibel, was einzelnen Schwankungen und auch Datenfehlern einen größeren Einfluß zukommen läßt.

Somit läßt sich für das vorliegende Problem der Trendentfernung wie auch für die folgende Verfahrensdiskussion sagen, daß generell lineare Ansätze zwar weniger gute Anpassung an das Datenmaterial liefern, dafür aber gegenüber den häufig einer adäquaten theoretischen Grundlage entbehrenden nichtlinearen Konzepten eine stabilere Vorhersage und allgemeinere Anwendbarkeit zum Vorteil haben. Soweit also linear-additive Zusammenhänge postuliert werden können, bewegt man sich jedenfalls im Bereich der „robusteren" Verfahren.

Nach diesem Exkurs in den methodischen Bereich noch einmal zurück zur Besprechung der verwendeten Klimadaten. Diese wurden ausschließlich dem von Meteorologischer Zentralanstalt und Hydrographischem Zentralbüro publizierten Material entnommen. Es handelt sich dabei um monatliche und jährliche Werte von Lufttemperatur und Niederschlag, Anzahl der Tage mit Mittel über 5 bzw. 10 Grad sowie Anzahl der Tage mit Niederschlag. Dazu kommen als Summen über die „kurze" Vegetationsperiode (IV–VIII) die mittlere Lufttemperatur, die 14-Uhr Temperatur, mittlere Maxima und, soweit vorhanden, Strahlung bzw. Sonnenscheindauer als deren Indikator. Diese Daten wurden an jeweils einer für den politischen Bezirk als repräsentativ erachteten Klimastation erfaßt, bei deren Auswahl natürlich einige Kompromisse erforderlich waren. Diese Vorgangsweise ist sicherlich nicht ideal, wurde aber als relativ beste herangezogen. Es ergeben sich daraus aber einige Probleme, die jedem klar sein werden, der sich mit den Prämissen für die Anwendung multivariater statistischer Verfahren auseinandersetzt: Bei den Klimadaten handelt es sich um Stichproben aus den Kontinua der Klimaelemente. Die räumliche Auswahl wurde zwar nicht nach den Kriterien einer Zufallsverteilung getroffen, kommt einer solchen aber mit einem nearest-neighbour Koeffizienten von 1.14 (ideal 1.0) sehr nahe. Demgegenüber sind die Ertragsdaten jeweils Teilflächen zugeordnet, für die sie aus Stichproben hochgerechnet wurden. Somit besteht zwischen den gegenübergestellten Daten eine gewisse strukturelle Inkongruenz, die aber hingenommen werden mußte.

Damit in engem Zusammenhang steht auch wieder die Frage eines adäquaten Maßstabs der Untersuchungen, wobei darunter das Niveau der räumlichen, zeitlichen

und auch sortenspezifischen Aggregation bzw. Disaggregation zu verstehen ist. Da, wie in 1. besprochen, dieser Maßstab im geographischen Interessensbereich sich nicht mit dem Datenangebot deckt, sind die erwähnten Nachteile in Kauf zu nehmen.

3. ZUR AGROKLIMATISCHEN GRUNDLAGE

Dieser Abschnitt kann natürlich keineswegs auch nur annähernd die Grundzüge agroklimatischer Beziehungsmuster darstellen, sondern soll lediglich einige Möglichkeiten der Beeinflussung agrarischer Ertragsbildung durch klimatische Parameter und die Spannweite möglicher Wirkungszusammenhänge aufzeigen. Die Bedeutung für eine Untersuchung der quantitativen Strukturen liegt dabei hauptsächlich im Vorverständnis der naturwissenschaftlichen Abhängigkeiten, das mehr oder weniger bewußt in der Anwendung des formalen Instrumentariums zum Tragen kommt. Dies bewirkt die Bedeutung der nachfolgend besprochenen Punkte:

Obwohl die kausale Abhängigkeit des landwirtschaftlichen Ertrags vom Klima als allgemein akzeptiert angesehen werden kann, sind beide Begriffe zu komplex, um direkt meßbar zu sein, da keiner dieser Terme eine klar definierte Ganzheit darstellt. Jegliche Untersuchung kann also nur bei eindeutig abgegrenzten Indikatoren ansetzen, wobei deren Abstraktionsniveau vom gewählten Arbeitsmaßstab mitbestimmt ist. Die Oberbegriffe „Ertrag" und „Klima" werden nach einem klar pragmatischen Aspekt – nach der Verfügbarkeit publizierter Daten – operationalisiert.

– Es existieren (nahezu) keine über den gesamten Variationsbereich der gemessenen Indikatoren stetigen (positiven oder negativen) Zusammenhänge, sondern es gibt Optima als Wendepunkte, wo dann bei linear postulierten Zusammenhängen die Funktionsparameter geändert werden müssen, solche Funktionen gelten also immer nur für einen bestimmten Bereich.

– Bivariate Abhängigkeiten sind aus dem gesamten Wirkungsgefüge nur schwer zu isolieren, können aber auch nicht einfach additiv zusammengefaßt werden, sondern entfalten wesentlich komplexere Wechselwirkungen (etwa im katalytischen Sinn). Dies ist eines der Hauptargumente für den Einsatz multivariater Verfahren zur simultanen Untersuchung von Einflußgrößen. Trotzdem ist meist die Verfahrensstruktur eine essentielle Vereinfachung des Interrelationsmusters.

– Derartige Simplifikationen des Wirkungsgefüges sind unumgänglich und die erzielten Ergebnisse unter diesem Aspekt zu beurteilen. Nur eine eng begrenzte Anzahl von (Indikatoren für) Einflußgrößen kann berücksichtigt oder explizit ausgeschaltet werden, der Rest der Ausprägung einer abhängigen (Ertrags-) Variablen wird als Zufallskomponente oder „statistisches Rauschen" interpretiert.

– Der gemessene Ertrag ist meist nur ein Teil der insgesamt produzierten Biomasse, es werden also nur die quantitativen Auswirkungen der Klimaindikatoren auf diese Subeinheit beurteilt. Dazu kommt beim in Gewichtseinheiten festgestellten Ertragsumfang die Frage des Gehalts an spezifischen, wertbestimmenden Inhaltsstoffen des Produkts.

Ertragssicherheit und Spitzenerträge sind selten völlig vereinbar. Somit nimmt (bei entferntem Trend) die Varianz der Ertragsergebnisse zu, was die formalen Zusam-

menhangswerte verschlechtert. Nach BAEUMER (1978) hängen „Menge und Qualität des Ertrags von der genetisch fixierten Reaktionsnorm der Pflanzen gegenüber den wechselnden Umweltbedingungen ab". Die Auswirkungen der in der Erbmasse erzielten Züchtungsfortschritte gehen also über die unter einem linearen positiven Trend subsummierten Einflüsse hinaus.

– Die Kompensation von klimatischen Nachteilen durch pflanzenbauliche Maßnahmen (im einfachsten Fall Bewässerung) ist nicht erfaßt und außerdem regional stark unterschiedlich, verringert also die aus dem Datenmaterial erzielbare Aussageschärfe.

– Bei regionalen Vergleichen können sich sortenspezifische Unterschiede als hinderlich erweisen, die meist ausgleichend auf unterschiedliche Umweltbedingungen reagieren, der Qualitätsunterschied wird in diesem Rahmen ja nicht erfaßt. Dazu kommen noch im gesamten pflanzenbaulichen Bereich wirksam werdende Fragen der zentralitäts- (= informations-) und kapitalgebundenen Innovationsvermittlung. Die Diffusion aller Neuerungen dauert immerhin so lange, daß hier ein zusätzlich regional differenzierender Faktor wirksam wird.

Hier sind also hauptsächlich Störfaktoren bei der Erklärung von Ertragsschwankungen aus der Klimavariation aufgelistet. Alle diese und andere Fragen der Substitution, Interdependenz und Überlagerung von Einflußgrößen stehen in engem Zusammenhang mit der formal und sachlich richtigen Anwendung statistischer Verfahren, weshalb der Übergang auf die folgenden Probleme wesentlich natürlicher ist, als er vorerst scheinen mag. Nahezu allen formalen Fragen liegen inhaltliche Probleme zugrunde, die formale Richtigkeit einer Verfahrensanwendung kann also nicht völlig aus der gegebenen Thematik herausgelöst werden.

4. ÜBERSICHT MULTIVARIATER TECHNIKEN

Allen diesen statistischen Verfahren liegt das probabilistische bzw. stochastische Erklärungsmodell zugrunde. Während im deterministischen Sinn für jeden Fall A behauptet wird, daß das Ereignis B folgt, spielt bei stochastischen Erklärungen die Zufallskomponente eine dominante Rolle, die über die bloße Addition eines Fehlerterms hinausgeht (GILCHRIST 1976).

Die statistische Analyse von Zusammenhangsstrukturen intervallskalierter Daten basiert weitgehend auf der einfachen (bivariaten) Korrelations- und Regressionsrechnung, die sich im wesentlichen dadurch unterscheiden, daß bei der Korrelationsrechnung keine gerichtete Abhängigkeit vorausgesetzt wird. Die geläufigen multivariaten Techniken können in zwei große Gruppen geteilt werden: die Verallgemeinerungen der Korrelations- und Regressionsanalyse, denen die faktorenanalytischen Techniken im weitesten Sinne gegenüberstehen. Diese beiden Hauptgruppen sollen nun kurz in ihrer konzeptuellen Struktur vorgestellt werden.

4.1. Korrelations- und regressionsanalytische Verfahren

Hiezu gehört zunächst als Übergang von den bi- zu den multivariaten Verfahren die partielle Analyse, die zwar auch nur den Zusammenhang zwischen zwei Variablen berechnet, dafür aber eine oder mehrere weitere Variable auspartialisiert, d. h. ihren Einfluß explizit eliminiert bzw. „konstant hält".

Demgegenüber bezieht die multiple Analyse auch diese Merkmale voll ein und versucht den simultanen Einfluß mehrerer unabhängiger Variabler auf eine abhängige Variable zu eruieren. Diese Technik ist wohl die mit den meisten Varianten und dem breitesten Anwendungsspektrum, wie nachstehend noch zu sehen sein wird.

4.2. Faktorenanalytische Verfahren

Die im weitesten Sinne als Faktorenanalyse bezeichneten Verfahren gehen zwar auch von den Beziehungen zwischen verschiedenen Merkmalen aus, versuchen dann aber, eventuell zugrundeliegende gemeinsame Strukturen („common patterns") aufzudecken. Der vieldimensionale Merkmalsraum soll auf einige wenige Basisdimensionen reduziert werden. Dabei kann zwischen Faktorenanalyse im engeren Sinn und Hauptkomponentenanalyse unterschieden werden, wobei letztere eigentlich mehr eine Datentransformationstechnik zum Zweck der Redundanzausschaltung und Informationskonzentration auf wenige „Hauptkomponenten" ist. Dagegen stellt die Faktorenanalyse im engeren Sinn höhere theoretische Ansprüche, da die Basisdimensionen als zugrundeliegende, meist nicht direkt meßbare „höherrangige" Einflußgrößen betrachtet werden. Zwischen Variablen und Faktoren besteht also eine Unterscheidung in Indikatoren und von diesen in die Meßdimensionen abgebildete theoretische Konstrukte.

Als eine Mischtechnik zwischen Faktorenanalyse und Korrelationsanalyse kann die kanonische Korrelationsanalyse interpretiert werden. Für zwei Datensätze mit mehreren Variablen werden Faktoren berechnet und diese aber zugleich unter dem Gesichtspunkt bestmöglicher Korrelation zwischen den Datensätzen gewichtet, es werden also gemeinsame Basisdimensionen in zwei Variablengruppen gesucht.

Abschließend soll noch kurz auf die Gruppe der clusteranalytischen Verfahren eingegangen werden, wobei eine Menge von Objekten, die durch eine Menge von Merkmalen identifiziert sind, auf möglichst intern homogene und klar unterscheidbare Gruppen oder Cluster aufgeteilt werden sollen. Neben dem allgemein klassifikatorischen Einsatz ergibt sich bei der Wahl von Raumeinheiten als Objekte die Möglichkeit zur Regionalisierung.

4.3. Gemeinsamkeiten multivariater Verfahren

Allen diesen Techniken sind nun eine Reihe von Prämissen, Restriktionen und sonstigen Charakteristika gemeinsam, die vorher noch allgemeingültig zusammengefaßt werden sollen. So ist das mathematische Konzept zur simultanen Analyse von mindestens 3 Variablen in der Zusammenfassung dieser Variablen in einen oder mehrere kombinierte Werte, die „Linearkombinationen", begründet. Zur vereinfachten Darstellung werden diese Gleichungen meist zu Matrizen kombiniert, bestehen aber prinzipiell aus additiv verknüpften Termen der linear gewichteten (d. h. nicht potenzierte Parameter) unabhängigen Eingangsvariablen. Durch den Effekt der Informationsreduktion und Konzentration auf die wesentlichen Inhalte ermöglichen diese Linearkombinationen sowohl vereinfachte Operationen wie auch vereinfachte Darstellung der Verfahrensabläufe.

Lineare Additivität ist demnach ein gemeinsamer Nenner aller besprochenen Verfahren, da der nichtlineare Fall bereits zu Beginn ausgeklammert wurde. In Zusammenhang mit der anschließend kurz besprochenen Signifikanzprüfung steht die Forde-

rung nach einer definierten Verteilung der Daten, meist einer Normalverteilung, die für Wahrscheinlichkeitsangaben probabilistischer Schlüsse von Bedeutung ist. Weiters ist Homoskedastizität gefordert, was bedeutet, daß die Varianz konstant bleibt, also nicht mit einer der Variablen zu- oder abnimmt.

Eine für multivariate Verfahren typische Frage ist die nach Multikollinearität. Damit ist das Bestehen linearer Interdependenzen innerhalb eines Variablensatzes gemeint. Über die Strenge dieser Forderung nach Orthogonalität von Variablen für diverse Verfahren besteht keine Übereinstimmung. Jedenfalls aber führt das Auftreten von perfekten Korrelationen zum Abbruch der Berechnungen, muß also unbedingt vermieden werden.

Eine in ihrer Rigidität ebenfalls unterschiedlich bewertete Forderung ist die nach Multi-Normalverteilung. Demnach genügen jeweils für sich normalverteilte Variable nicht, sondern eine gemeinsame mehrdimensionale Normalverteilung der Objekte in allen Merkmalsdimensionen um ein Symmetriezentrum wäre erforderlich.

Abschließend gilt noch für alle Verfahren eine Tatsache, die sich aus den Eigenschaften von Linearkombinationen ableiten läßt. Sobald sich die Anzahl der Variablen der Zahl der Basiseinheiten (an denen die Variablen gemessen werden) nähert, konvergiert der erklärte Anteil jedenfalls gegen 100%, jegliche Vorhersage wird aber extrem instabil und die Gleichung gilt nur mehr für den gegebenen Datensatz. Diese formale Einschränkung sollte aber ohnehin nicht zum Tragen kommen, da die Auswahl der Variablen durch theoriegeleitetes Vorgehen und a priori – Formulierung von Hypothesen gekennzeichnet sein sollte. Dies alles spricht für eine knappe Auswahl der Variablen auf Grund der Voruntersuchung bzgl. deren linearer Interdependenz, eine vorgeschaltete Orthogonalisierung dagegen ist aber im Zuge der Interpretation häufig nur schwer zu verarbeiten.

4.4. Signifikanzprüfung

Für alle numerischen Ergebnisse multivariater statistischer Verfahren wie etwa Korrelationskoeffizienten lassen sich Maßzahlen berechnen, die Auskunft über die Verläßlichkeit und Aussagefähigkeit der errechneten Werte geben. Da empirische Untersuchungen i. A. nicht an der Gesamtheit des interessierenden Phänomens, sondern an einer Stichprobe daraus ausgeführt werden, geht es bei der Signifikanzprüfung primär um die Gültigkeit der Übertragung von ermittelten Parametern auf die Grundgesamtheit. Dies ist auf der Grundlage einer v o r h e r festgelegten Fehlerwahrscheinlichkeit (z. B. 5%) zu entscheiden. Meist werden diese das Ergebnis bewertenden Maßzahlen unter einer entsprechenden Verteilungsannahme aus der Anzahl der Freiheitsgrade (Anzahl der Basiseinheiten minus Anzahl der Variablen) und dem Ergebnis- (Korrelations-) Koeffizienten errechnet. Dies ist natürlich nur eine kurze Charakteristik der üblicherweise verwendeten Signifikanztests.

5. VERFAHRENSBEISPIELE

5.1. Die partielle Korrelation und Regression

Wie bereits erwähnt, bildet diese Technik den „Einstieg" in das Gebiet multivariater Verfahren. Gegenüber einer bivariaten Methode wird hier der Einfluß einer oder mehrerer weiterer Variabler konstant gehalten, also kontrolliert und damit eliminiert. In

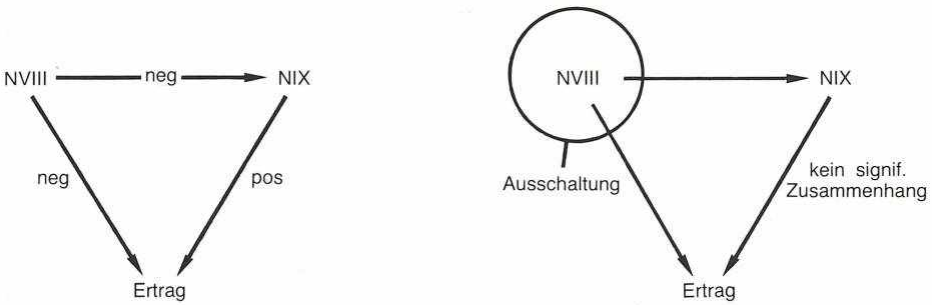

Abbildung 1: Wandel des Relationsabbildes durch Auspartialisieren

komplexen Interrelationsstrukturen können damit mögliche Störeffekte auf eine bivariate Beziehung ausgeschaltet und damit die wirkliche Zusammenhangsqualität ermittelt werden. Das sei kurz an dem folgenden Beispiel aus dem Bezirk Amstetten demonstriert. Eine auffällig enge positive Beziehung zwischen den Getreideerträgen und der Niederschlagssumme im September war vorerst sachlogisch nicht zu erklären, da Getreide sowohl für den Abschluß der generativen Entwicklung wie auch für den Einsatz moderner Erntetechnologie trockenes und warmes Wetter benötigt, außerdem praktisch die gesamte Ernte bereits im August eingebracht wird. Erst das Kontrollieren (= Auspartialisieren) der Niederschlagssumme im August führte die Beziehung Ertrag – Septemberniederschlag auf ein unsignifikantes Maß zurück. Offenbar folgt nämlich auf einen trockenen August ein feuchter September und umgekehrt, was die positive Scheinkorrelation zwischen Septemberniederschlag und Ertrag erklärt (siehe Abbildung 1).

Die verfahrenstechnische Vorgangsweise kann auch so erläutert werden, daß man die gesamte Stichprobe in Untergruppen nach den Ausprägungen der Kontrollvariablen (Niederschlagssumme August) ordnet. Für jede einzelne Untergruppe ergibt sich dann kein signifikanter Zusammenhang Septemberniederschlag-Ertrag (siehe Abbildung 2).

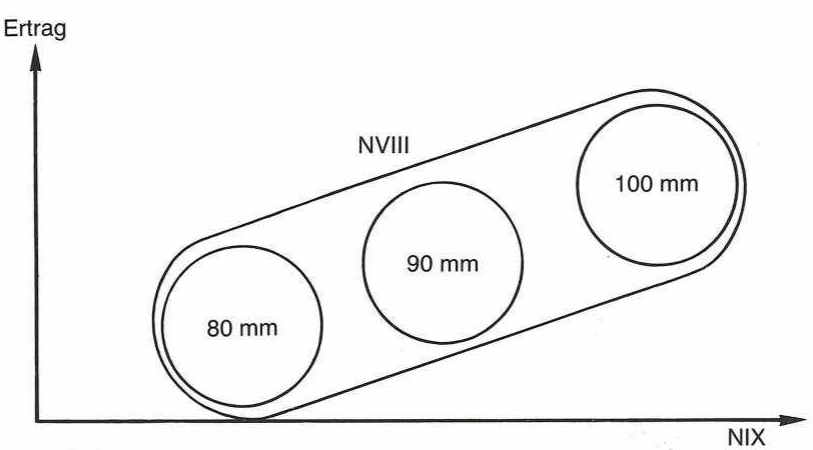

Abbildung 2: Elimination des Einflusses unterschiedlicher Ausprägungen der Kontrollvariablen

Nach BAHRENBERG-GIESE kann also der partielle Korrelationskoeffizient definiert werden als: „das Ausmaß, zu dem eine Variable mit einer anderen zusammenhängt, wenn man den Einfluß eines oder mehrerer bei der einfachen Korrelation implizit eingehender Faktoren ausschaltet."

Als Beispiel für ganz Niederösterreich soll der Zusammenhang zwischen dem Ertrag von Winterweizen (WW) und der Anzahl der Tage mit Niederschlag (TGN) herangezogen werden. Im Mittel liegt hier eine signifikant negative Beziehung vor, was auf ausreichende Feuchteverhältnisse und somit negativen Einfluß gesteigerten Niederschlags hinweist. Partialisiert man aber nun die durchschnittlich positive Einflußgröße der Tage mit einem Temperaturmittel über 10 Grad aus (TG 10), so kehrt sich in den trockeneren Räumen im Osten Niederösterreichs die ursprüngliche Beziehung um: dort ist nicht die Temperatur die entscheidende ertragsbestimmende Variable, sondern die im Minimum befindliche Feuchte, während der Ertrag etwa im Alpenvorland dominant von der thermischen Variation gesteuert wird (Abbildung 3).

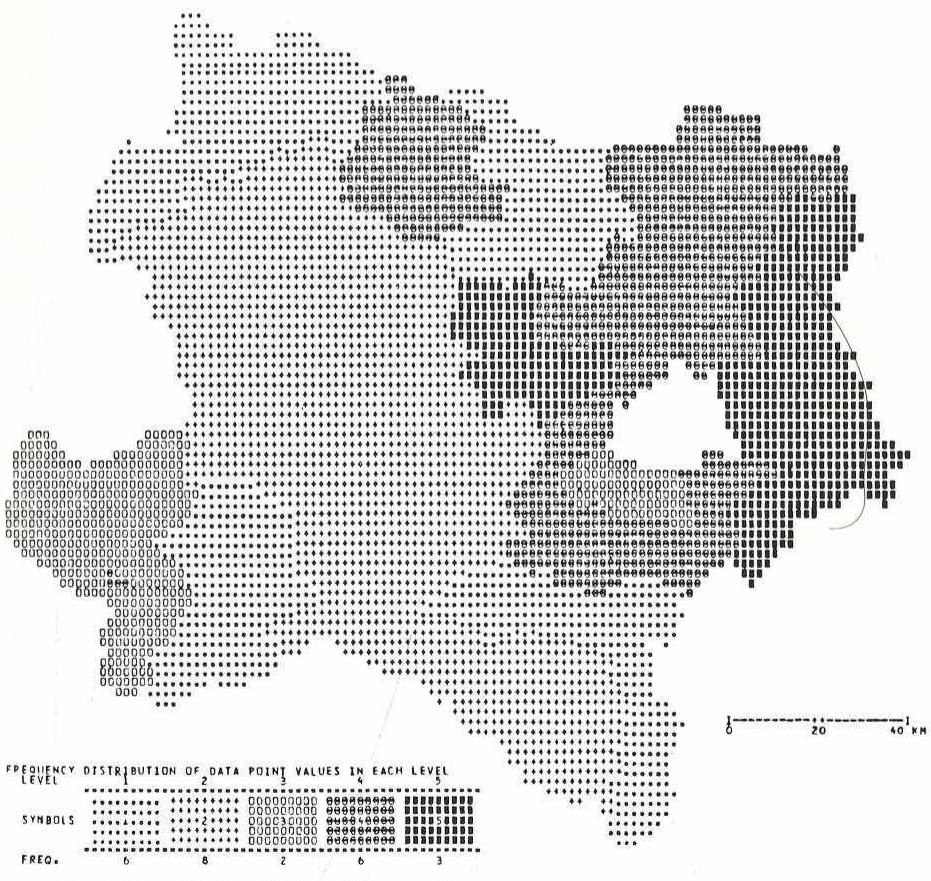

Abbildung 3: Der korrelative Zusammenhang von Winterweizen mit der Anzahl der Tage mit Niederschlag unter Ausschaltung der Anzahl der Tage mit einem Mittel über 10° C

Hier etwa wäre die Aussagefähigkeit eines multiplen Ansatzes geringer, da sich thermische und hygrische Variable durch ihr ausgeprägt antagonistisches Verhalten im Untersuchungsgebiet meist auf dieselbe Basisdimension abbilden und somit die Einbeziehung von TG 10 keinen wesentlichen Informationszuwachs gebracht hätte.

5.2. Multiple Korrelation und Regression

Bei der multiplen Analyse werden nun weitere unabhängige Variable nicht nur kontrolliert, sondern voll einbezogen. Damit versucht man mit mehreren unabhängigen Variablen (Prädiktoren) einen möglichst großen Varianzanteil der abhängigen Varia-

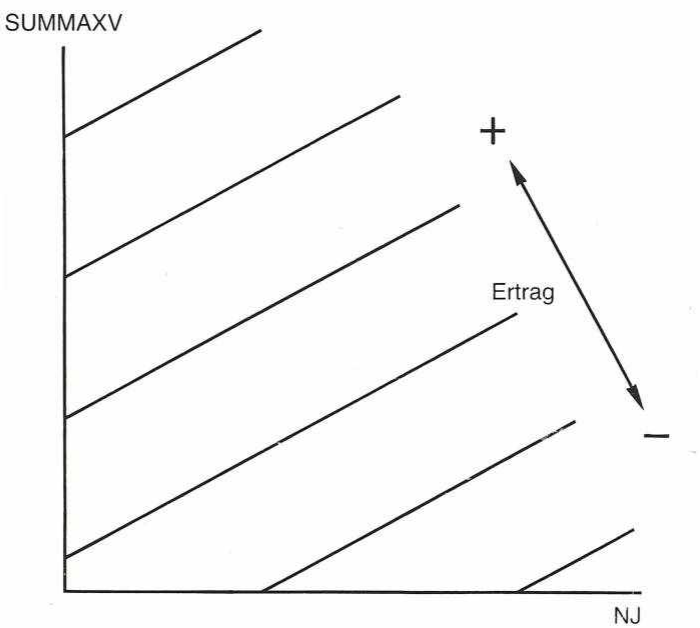

Abbildung 4: Reaktion des Zuckerrübenertrags auf die kombinierten Effekte der jährlichen Niederschlagssumme und der Vegetationszeitsumme der mittleren Temperaturmaxima

blen (Kriterium) zu erklären. Damit liegt gegenüber den bivariaten Methoden ein wesentlich realitätsnäheres Verfahren vor, da ja in der Natur auch nicht nur jeweils 2 Merkmale in wechselseitiger Beziehung stehen. Die Ergebnisse dieser simultanen Schätzung des Kriteriums aus mehreren Prädiktoren werden noch dadurch weiter verbessert, daß es eine ganze Reihe von Varianten und Modifikationen dieser Technik gibt. Die multiple Analyse hat also als eines der flexibelsten Verfahren überhaupt ein sehr weites Anwendungsspektrum. Eine „bestmögliche" Linearkombination aller Prädiktorvariablen bildet den linear abhängigen Anteil der Varianz der zu erklärenden Variablen, der Varianzrest wird wieder als eigenständiger „Fehlerterm" angesehen. Die multiple Korrelation mißt einfach das Ausmaß der gemeinsamen Varianz aller unabhängigen Variablen mit der abhängigen Variablen. Demgegenüber hat die multiple Regression eine optimale lineare Schätzung des Kriteriums auf Grund der Prädiktoren zum Ziel.

Im einfachsten Fall von 2 Prädiktoren kann dieser simultane Einfluß auf die Zielvariable (in einer Abbildung der 3. Dimension durch Isolinien) noch gut graphisch veranschaulicht werden. Bei Hinzufügen weiterer Prädiktoren würde man zu nur mehr analytisch erfaßbaren Hyperebenen gelangen. Für den Ertrag von Zuckerrüben (ZR) in Niederösterreich wurden die jährlichen Niederschlagssummen (NJ) und die Summe der mittleren täglichen Maxima in der Vegetationsperiode (SUMMAXV) herangezogen. Wie aus Abbildung 4 zu entnehmen ist, steigt der Ertrag mit SUMMAXV und sinkt mit NJ, wobei das Ausmaß der Variation durch Dichte und Steigung der „Isogeraden" festgelegt ist. (Die Parameter beruhen auf nicht standardisierten Daten!)

Weiters ließe sich analog zur vorliegenden Skizze ein Diagramm zur Verteilung der Residuen herstellen, das Auskunft über die Güte der Anpassung in den einzelnen Ausprägungsbereichen der Prädiktoren gibt.

Als Beispiel für die Anwendung einer multiplen Regression mit mehreren Variablen wurde der Rotklee-Ertrag (RK) im Bez. Amstetten gewählt. Als Prädiktorensatz dienen die bereits geläufigen Variablen NJ, TG 10 und SUMMAXV. Es ergibt sich ein Regressionskoeffizient von $r = .953$, der auf einem Signifikanzniveau von 95% gültig ist. Die Schätzgleichung dazu lautet:

$$RK = -.065\ NJ + .957\ TG\ 10 - .013\ SUMMAXV + 292.6$$

Eine Modifikation des Verfahrens liegt in der schrittweisen multiplen Regression vor, wobei die einzelnen Prädiktoren je nach der Größe ihres Beitrages zur Varianzerklärung des Kriteriums schrittweise einbezogen werden. Dies erleichtert etwa die Auswahl eines möglichst knappen optimalen Prädiktorensatzes, da jeder Prädiktor daraufhin untersucht wird, ob er einen signifikanten zusätzlichen Beitrag zur Vorhersage zu leisten imstande ist oder nicht. Die Reihenfolge der Einbeziehung für obiges Beispiel ist NJ – TG 10 – SUMMAXV, was die Bedeutung der Hydratur für Grünlandkulturen unterstreicht.

Hervorragende Vorhersageergebnisse können auch häufig mit Hilfe der polynomischen Regression erzielt werden. Dabei sind die Parameter der Gleichung nach wie vor linear, sie gewichten aber quadratische, kubische, quartische, quintische usw. Polynome, die additiven Termen mit Variablen zur 2., 3., 4. und 5. Potenz entsprechen. Die höchste vorkommende Potenz bezeichnet dann den Grad des Polynoms. Als Indiz für den Einsatz der polynomischen Regression können Diagramme mit Verteilungen dienen, die auf die typischen quadratischen, kubischen usw. funktionellen Erscheinungsbilder hinweisen, wie auch theoretische Überlegungen oder Analogien zu physikalischen Phänomenen.

In der Praxis wird man kaum über Terme 3. Grades hinausgehen, da mit Steigerung des Polynomgrads und damit der Anzahl der Terme in der Gleichung die Anpassung an die zur Eichung dienenden Daten zwar verbessert, die Stabilität der Schätzung jedoch verringert wird. Sofern ein Einbeziehen potenzierter Variabler in eine Vorhersagegleichung keine signifikante Verbesserung ergibt, kann dies als Indiz für die Linearität eines Zusammenhangs herangezogen werden, ebenso wie dies bei der Einbeziehung höhergradiger Terme für den Abbruch der Steigerung des Polynomgrads spricht.

Als einfaches (bivariates) Beispiel soll versucht werden, mit einer polynomischen Kurve den Verlauf des Zuckerrübenertrags im Bez. Amstetten 1946–74 anzunähern. Polynome bis zum 4. Grad wurden berechnet, der Verlauf der Zunahme des erklärten Varianzanteils ist aus Abbildung 5 ersichtlich:

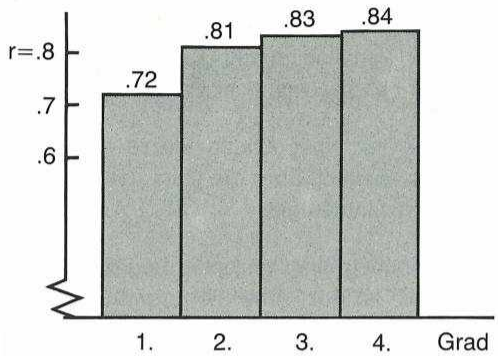

Abbildung 5: Steigerung des Anpassungsgrads eines Polynoms

Wie der „Qualitätssprung" zum quadratischen Polynom anzeigt, bildet eine solche Funktion den steilen Ertragsanstieg der Nachkriegszeit und die folgende Kurvenverflachung besonders gut ab (siehe Abbildung 6).

Allgemein formuliert würde die Gleichung lauten:

$$x = a_1t^2 + a_2t + c$$

Es liegt hier also trotz des relativ einfachen Ansatzes ein sehr anpassungsfähiges Verfahren vor. Diese zunehmende Flexibilität und potentielle Leistungsfähigkeit eines Verfahrens erfordert allerdings auch erhöhte Vorsicht und Problembewußtsein in der Anwendung, da naturgemäß einfachere Ansätze weniger formale und sachlogische Fehlerquellen in sich bergen.

5.3. Trendflächenanalyse

Eigentlich ist dieses Verfahren zwar „nur" ein Spezialfall der multiplen (polynomischen) Regression, jedoch sei es hier als eine Umsetzung in die räumliche Dimension gesondert herausgegriffen. Die räumliche Trendvariable (z) wird nicht wie etwa im letzten Beispiel als von e i n e r Variablen (Zeit) abhängig gedacht, sondern von ihrer Allokation im in die Ebene projizierten Raum, wo sie durch die geographischen Koordinaten x und y definiert ist.

Die Variable z soll nun als Funktion von x und y ausgedrückt werden, und zwar dient als Grundlage für die Berechnung der Funktionsparameter die multiple polynomische Regression, die im einfachsten Fall zu einer einfachen multiplen Regression (= Polynom 1. Grades) degeneriert. Bei der Zugrundelegung empirischer Daten wird postuliert, daß die abhängige Variable z aus einer deterministischen (funktionell erfaßbaren) und einer stochastischen Komponente besteht:

$$z = f(x, y) + e$$

Die deterministische Komponente ist dabei eine additive Kombination der gewichteten unabhängigen Variablen x und y sowie deren potenzierter Kombinationen. Eine räumliche Zufallsstichprobe dient zur Schätzung der Parameter der Funktionsgleichung.

Eine derartige Trendflächenanalyse kann nun (nach STEINER 1977) verschiedene Ziele verfolgen:
- vereinfachte Beschreibung der wichtigsten räumlichen Trendkomponenten im Untersuchungsgebiet.

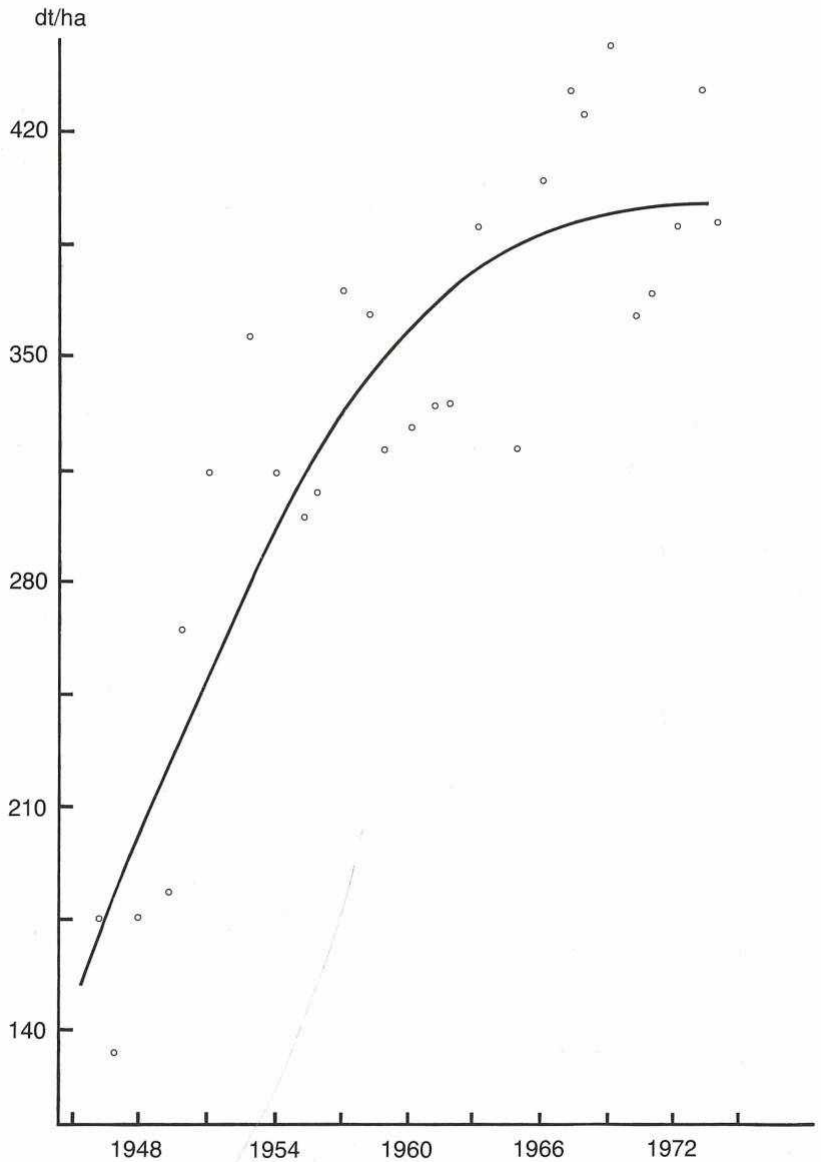

Abbildung 6: Anpassung eines quadratischen Polynoms an den Verlauf des Zuckerrübenertrags

– Separation von Variationskomponenten in regionale Trends und Residuen (unsystematische lokale Abweichungen davon).
– Interpolation durch Einsetzen beliebiger Koordinaten in die räumliche Prädiktionsgleichung. Ein derart von allen Datenpunkten ungeachtet der Entfernung gleich stark abhängiger Wert ist aber nur in den seltensten Fällen erwünscht.
– Generalisierung: aus denselben Gründen wie im letztgenannten Fall sind meist lokale Verfahren vorzuziehen.
– Kompakte Speicherung: sofern nur zur Reproduktion der ursprünglichen Datenpunkte verwendet, sind v. a. Polynome möglichst hohen Grades geeignet.

Den inferenzstatistischen Sachzusammenhang zu den anderen hier besprochenen Verfahren trifft eigentlich nur der Einsatz zur Separation von Variationskomponenten. Auch ist nur dann eine entsprechende Signifikanzprüfung möglich und sinnvoll.

Ein einfacher Ansatz zur Beurteilung der Anpassungsqualität der Fläche sind analog zur multiplen Regression berechnete Maßzahlen, die über die Verteilung der Gesamtvarianz Auskunft geben. Der Verlauf der Abnahme der Residualvarianz kann wieder als Kriterium für oder gegen den Abbruch der Steigerung des Polynomgrads herangezogen werden.

Als Beispiel dient hier die Verteilung der 10jährigen Mittel der Tage mit Niederschlag in Niederösterreich. Die regionalen Trends werden in einem Polynom 3. Ordnung wohl am besten nachgezeichnet – bei folgender Gewichtung der (lokalen) Koordinaten (siehe auch die Abbildung 7):

$$z = 0.11E4 - 0.61E1x - 0.52E2y - 0.12E1x^2 + 0.44xy + 0.51y^2 + 0.87E4x^3 - 0.94E\text{-}3x^2y - 0.19E\text{-}2xy^2 - 0.18E2y^3$$

Gebiete schlechterer Anpassung können anhand einer Residualkartierung abgegrenzt werden: dort überwiegen lokale Abweichungen vom regionalen landesweiten Trend (natürlich immer bezogen auf den jeweiligen Polynomgrad).

Bereits beim nächsthöheren Polynom treten Effekte wie zunehmend instabiles Verhalten der Fläche besonders an den Rändern auf, die für hohe Potenzen typisch sind und als Gegenmaßnahme etwa eine Ausdehnung der Stichprobenziehung über die Grenzen des Untersuchungsgebiets hinaus erfordern würden.

Als so günstig sich die Trendflächenanalyse z. B. zum Zweck räumlicher Veranschaulichung von Trends erweist, so wenig ist sie für eigentlich multivariate Untersuchungen geeignet. Für den Vergleich mehrerer Trendflächen etwa zur Gegenüberstellung von Klima- und Ertragsmerkmalen existieren eigentlich kaum formalisierte Ansätze, sodaß für derartige, über den explorativen Charakter hinausgehende Erklärungsversuche meist andere Techniken herangezogen werden müssen.

Verallgemeinert man die Trendflächenanalyse auf mehrere Prädiktoren, etwa durch Einbeziehung der Seehöhe zusätzlich zu den räumlichen Koordinaten, so ist das Ergebnis wieder eine Hyperfläche, wogegen die Annahme mehrerer Kriteriumsvariabler zur kanonischen Trendflächenanalyse führt. Hier handelt es sich um in der Geographie kaum praktizierte Techniken, die dann meist den Vorteil der einfachen räumlichen Visualisierung der Resultate entbehren. Auf die natürlich mögliche Änderung der funktionellen Form der Regressionsgleichung auf x und y soll nicht näher eingegangen werden.

5.4. Faktorenanalyse

Bei dieser großen Verfahrensfamilie werden die Einschränkungen einer Kurzcharakteristik von Techniken für einen bestimmten Einsatzbereich ganz besonders spürbar. Für die verbreitete Anwendung haben sich enorm viele Spielarten und fein differenzierte Ansätze entwickelt, sodaß hier nur kurz die beiden Hauptzweige, die Hauptkomponentenanalyse und die Faktorenanalyse i. e. S. vorgestellt werden können. Eine echte Bewertung der Einsatzmöglichkeiten ginge natürlich weit über den gesteckten Rahmen kurzer Anwendungsanregungen hinaus.

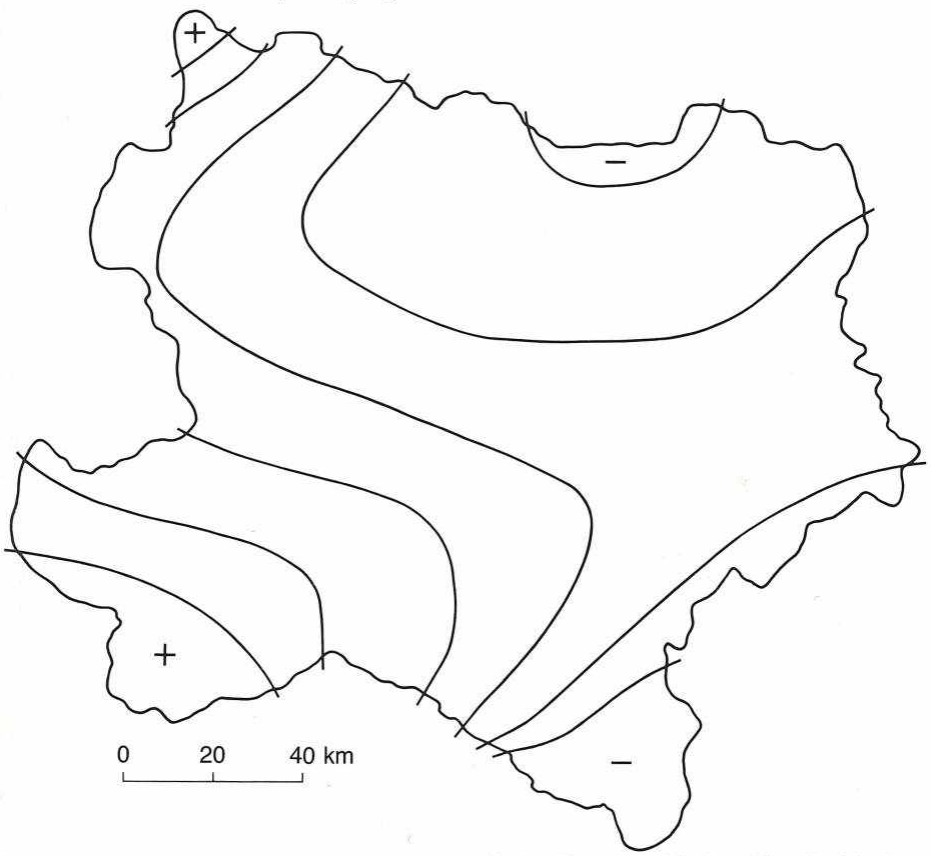

Abbildung 7: Kubische Trendfläche der Anzahl der Tage mit Niederschlag in Niederösterreich

Ganz allgemein besteht das Hauptziel der Faktorenanalyse darin, eine größere Anzahl vorgegebener Variabler auf eine möglichst kleine Anzahl neuer, voneinander linear unabhängiger hypothetischer Größen, sogenannte Faktoren, zu reduzieren. Dabei gilt die Grundhypothese, daß die beobachteten Merkmalswerte durch zugrundeliegende, vorläufig unabhängige Faktoren hervorgerufen werden, wobei angenommen wird, daß die Merkmale durch lineare Kombination dieser Faktoren entstehen (STEINER 1979). Diese aus einer Reihe beobachteter Merkmale (Variablen) abgeleiteten hypothetischen Größen oder Faktoren sollen möglichst einfach sein, andererseits aber die Beobachtungen ausreichend genau beschreiben und erklären. Die faktorenanalyti-

schen Verfahren sollen hier nur in ihrer Anwendung als Methode der deskriptiven Statistik und als explorativer Ansatz dargestellt werden, da eine Einbeziehung der hypothesentestenden (konfirmatorischen) Faktorenanalyse zusätzliche Prämissen und inferenzstatistische Überlegungen auf höherem Theorieniveau erfordern würde.

Grundziel der meisten Faktorenanalysen ist die Neuaufteilung der Gesamtvarianz aller einbezogenen Variablen unter Ausschaltung jeglicher Redundanz (Informationsüberlappung) und Abbildung möglichst großer Varianzanteile auf möglichst wenige Faktoren. Je größer also derartige partielle Redundanzen zwischen Merkmalen sind, umso eher können Faktoren mit großen Varianzanteilen entstehen. Als geometrische Interpretation kann man von einem vieldimensionalen, obliquen Merkmalsraum sprechen, dessen Informationsgehalt weitgehend auf wenige, meist orthogonale, Dimensionen transferiert werden soll.

Bei der Hauptkomponentenanalyse wird eine vollständige Lösung berechnet, was besagt, daß i. A. n Variable durch n Faktoren (bei umverteilter Varianz) abgebildet werden. Es handelt sich dabei also dominant um eine Datentransformationstechnik zur Informationskonzentration und Redundanzausschaltung unter Wahrung der gesamten vorhandenen Varianz.

Bei der Faktorenanalyse i. e. S. steht ein theoretisches Konzept dem Variablensatz zugrundeliegender gemeinsamer Einflußgrößen im Vordergrund, die durch Faktoren repräsentiert werden. Es wird also streng zwischen den mehreren Faktoren gemeinsamen und den diesen spezifischen Einflüssen unterschieden. In der Regel wird also nicht die gesamte Varianz durch die Faktoren erklärt, der gemeinsame Anteil wird als „Kommunalität" bezeichnet.

Formelles Hauptziel einer jeden solchen Berechnung ist die Erzielung einer Einfachstruktur, worunter man die möglichst gute Interpretierbarkeit der einzelnen Faktoren versteht. Das wird erreicht durch besonders hohe oder niedrige Beteiligung einzelner Variabler an einem Faktor („Faktorladungen"), wobei als Angelpunkte für die Interpretation „Leitvariable" mit hoher Ladung fungieren.

Die möglichen Anwendungen einer derart vielfältigen Verfahrensgruppe können natürlich nicht umfassend beschrieben werden, sondern nur einige Ideen möglicher Ansätze kurz dargelegt werden. Ein Beispiel für den Einsatz einer Hauptkomponentenanalyse wäre die Anwendung auf den Satz der vorliegenden Ertragsvariablen. Informationskonzentration wäre dabei in dem Sinn zu sehen, daß eine Mehrfachgewichtung von Reaktionsmustern (auf bestimmte klimatische Parameter) durch ähnliches Verhalten verschiedener Feldfrüchte vermieden wird, indem diese auf den gleichen Faktor transformiert werden.

Diese Analyse wurde am Bezirk Gänserndorf durchgeführt und brachte eine Zusammenfassung zu 3 Typen unterschiedlichen Ertragsverlaufs im Untersuchungszeitraum. Faktor 1 kombinierte niederschlagsorientierte Feldfrüchte wie Rotklee und Grünmais. Faktor 2 ist ein ausschließlicher Getreidefaktor, 3 dagegen unterstreicht die speziellen Ansprüche der Zuckerrübe durch Isolation von den anderen Arten. Mit diesen Typen von Feldfrüchten könnten nun weitere Analysen formal vereinfacht und ohne ungleiche Gewichtung durch asymmetrisch ausgewählte Variable durchgeführt werden.

Als Beispiel für eine Faktorenanalyse i. e. S. dient eine Analyse aller erhobenen Klimavariablen im Bezirk Amstetten. Es soll eine Extraktion von thermischem und hygrischem Faktor stattfinden, den „Einzelrestfaktoren" der Variablen wird wenig Bedeutung zugemessen. Die Einflußgrößen Temperatur und Niederschlag führen ja erst zur Realisation der diversen Meß- und Zähldaten. Es wurden 3 signifikante Faktoren errechnet, davon ist einer eindeutig den Niederschlagsvariablen zuzuordnen. Die beiden thermischen Faktoren können durch die Gruppierung der Variablen in (Jahresmittel, Vegetationszeitmittel, Tage mit Mittel über 5 Grad) und (Tage über 10 Grad, Vegetationszeitsummen der 14 Uhr-Temperatur und der Strahlung) als luftmassen- bzw. strahlungsorientiert interpretiert werden.

Im 3. Beispiel werden bereits 2 Verfahren integriert, und zwar wird eine Regression des Ertrags von Winterweizen auf die Hauptkomponenten der einzelnen Klimavariablen in den niederösterreichischen Bezirken durchgeführt, was natürlich eine unschärfere Aussage ergibt, dafür aber die großräumige Identifikation der steuernden Klimagrößen unter der gegebenen Voraussetzung perfekter Orthogonalität aller Prädiktoren ermöglicht.

Die Interpretation einer derartigen Analyse wäre hier für alle Bezirke zu umfangreich und ergab für ganz Niederösterreich die höchste Gewichtung für den Faktor, der die Tage mit Niederschlag hoch positiv, die Niederschlagssumme aber negativ berwertet: kontinuierliche Hydratur fördert im Mittel die Ertragsbildung, während viel Niederschlag allgemein die Erträge verschlechtert. Unter den thermischen Merkmalen wurde die Vegetationszeitsumme der 14 Uhr-Temperatur am höchsten bewertet.

Diese wenigen Punkte sind sicherlich nur ein ganz grober Abriß des Bereichs der Faktorenanalyse, konnten aber vielleicht doch die Stellung im Rahmen der und im Vergleich zu den anderen multivariaten Verfahren verdeutlichen und die wichtigsten Charakteristika zum Ausdruck bringen.

5.5. Kanonische Korrelationsanalyse

Dabei handelt es sich um den allgemeinsten Fall einer Korrelationsanalyse, der bereits faktorenanalytische Konzepte mit einbezieht und deshalb hier eingeordnet wurde. Es geht um die korrelative Zusammenhangsmessung zwischen 2 Variablengruppen, deren interner Informationsgehalt faktorenanalytisch orthogonalisiert wird. Der Unterschied zu einer Anwendung beider Einzeltechniken liegt in einer simultanen Durchführung, die „Faktoren" werden bereits unter dem Gesichtspunkt maximaler wechselseitiger Übereinstimmung gebildet.

Somit ist es möglich, in einem Verfahren jeweils mehrere Klima- und Ertragsvariable zueinander in Beziehung zu setzen. Die theoretischen Terme Klima und landwirtschaftlicher Ertrag können also unmittelbar über ihnen zuzuordnende Variable („Indikatoren") korreliert werden. Zwei sehr unterschiedliche Beispiele sollen das breite Anwendungsspektrum verdeutlichen und die Interpretation illustrieren.

Zunächst wurde eine Untersuchung zum Wandel der Ertragsstruktur in den einzelnen Bezirken durchgeführt. Die beiden Variablensätze wurden aus den Erträgen 6 verschiedener Feldfruchtarten aus den Jahren 1966 (1965 war ein zu extrem schlechtes Jahr) und 1974 gebildet. Die 1. kanonische Korrelation (zwischen den jeweils 1. Faktoren der Variablensätze) war mit über 95%iger Wahrscheinlichkeit signifikant, die

kanonischen Gewichte gaben den Einfluß der einzelnen Feldfrüchte auf die Faktoren beider Jahre an, die kanonischen Werte geben Auskunft über den wechselnden Einfluß der einzelnen Bezirke. Ein 1974 günstigerer Ertrag von Sommergetreide gegenüber Wintergetreide und der Zuckerrübe ist abzulesen, ebenso ein weiterer Ausbau der Vorherrschaft ohnehin schon begünstigter Bezirke (Bruck, Gänserndorf, Mistelbach, Tulln und Wien-Umgebung). Im 2. Beispiel wird für jeden politischen Bezirk eine kanonische Korrelation zwischen einigen Ertrags- und Klimamerkmalen berechnet. Daraus ergibt sich die Möglichkeit zu einer vergleichenden Bewertung aller niederösterreichischen Bezirke auf Grund der Enge des Zusammenhangs Klima/Ertrag, wo also die klimatischen Parameter den größten Einfluß auf die Ertragsbildung ausüben. Die folgende Karte soll die regionale Verteilung dieser Zusammenhangsbeziehungen verdeutlichen (Abbildung 8).

Gegenüber der an sich vielversprechenden Konzeption dieses Verfahrens stellt allerdings die häufig problematische Interpretation der Ergebnisse einen schwerwie-

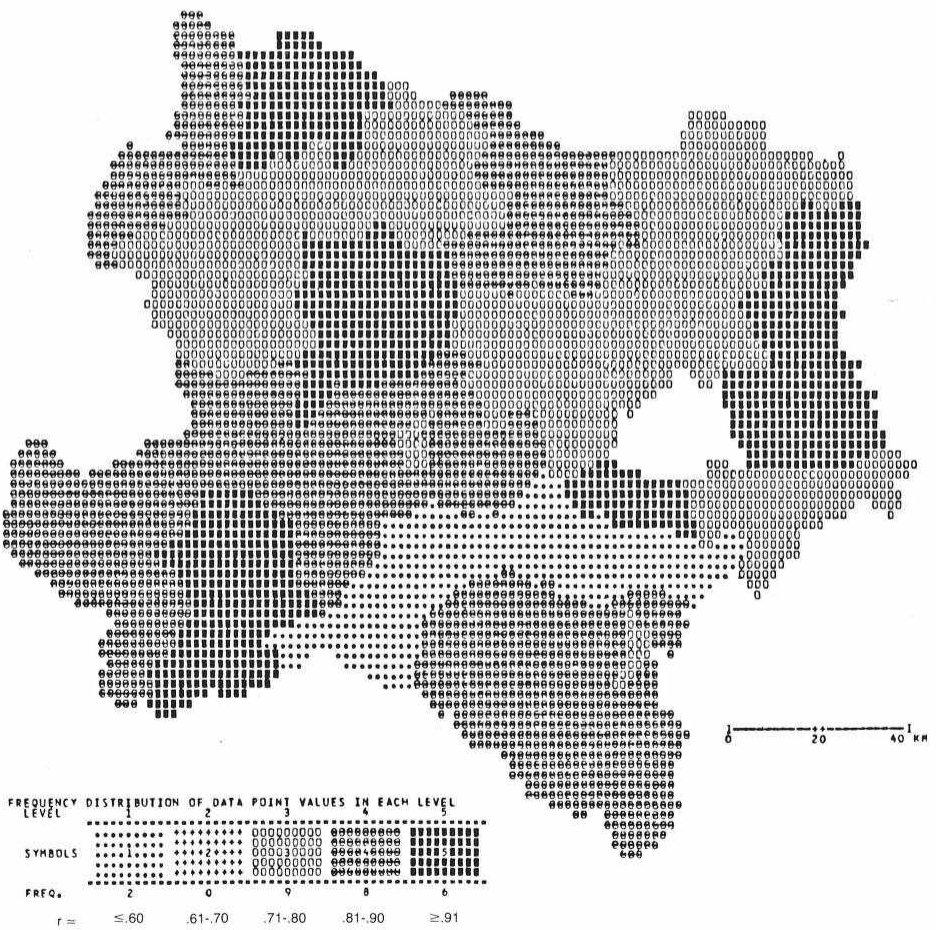

Abbildung 8: Verteilung der kanonischen Korrelationskoeffizienten zwischen Klima- und Ertragsvariablen in Niederösterreich

genden Nachteil dar. Dies ist sicher auch durch den komplexen Ansatz und die nur schwer nachzuvollziehenden Datentransformationen bedingt. In vielen Fällen werden mit der kanonischen Korrelation nicht die erhofften Resultate erzielt, was sicherlich mit ein Grund für die eigentlich recht seltene Anwendung ist. Der Einfluß der (hier nicht explizit angegebenen) Verfahrensprämissen ist nicht leicht einzuschätzen und wird unterschiedlich bewertet. Daher wird es günstig sein, Ergebnisse aus diesem Verfahren durch andere Techniken zu validieren, um die Reliabilität und Sicherheit der Interpretation zu erhöhen.

5.6. Clusteranalyse

Auch bei der Clusteranalyse handelt es sich nicht um eine einheitliche Technik, sondern um eine Gruppe von Methoden, wobei ich hier nur ihre gemeinsame Grundstruktur darstellen und einige Anwendungsmöglichkeiten im gegebenen thematischen Rahmen demonstrieren kann. Das Ziel einer Clusteranalyse besteht generell darin, eine Menge von Objekten (OTUs – operational taxonomic units), die durch eine Anzahl von Merkmalen (Variablen) identifiziert sind, auf in sich homogene Gruppen, Klassen oder Cluster aufzuteilen.

Im Rahmen geographischer Anwendungen ergeben sich neben allgemein klassifikatorischen Fragen noch solche der Regionalisierung, sobald räumliche Einheiten als OTUs fungieren. Man spricht von Raumtypen, wenn ähnliche Raumeinheiten ohne Rücksicht auf ihren räumlichen Zusammenhang zusammengefaßt werden. Dagegen ergibt sich die eigentliche Regionalisierung bei der Gruppierung benachbarter Einheiten zu Regionen höherer Ordnung („Kontingenzrestriktion"). Diese Möglichkeit ist nicht in allen clusteranalytischen Programmsystemen vorgesehen, steht aber etwa im CLUSTAN-Paket zur Verfügung. Weiters können noch 2 Typen von Regionen unterschieden werden: einstellige Prädikate (Attribute) von OTUs führen nach einem regionaltaxonomischen Prozeß zu homogenen Regionen mit ähnlichen Merkmalsausprägungen. Dagegen erzeugen 2stellige Prädikate (= Relationen) als Merkmale der OTUs funktionelle Regionen, weisen also die Zonen intensivster Interaktionsverflechtungen aus. Wie daraus zu ersehen ist, führt die Clusteranalyse schon ihrem Konzept nach keine Zusammenhangsanalyse im Sinn der vorhergehenden Verfahren durch, sondern nur eine Ähnlichkeitsbeurteilung von Objekten hinsichtlich der Merkmale, aber nicht der Merkmale selbst. Eine Kombination mit anderen Verfahren wird daher häufig der Fall sein.

Von den einzelnen Merkmalen wird wieder minimale inhaltliche Überschneidung (Redundanz) gefordert. Anhand dieser Merkmale soll eine Clusteranalyse dann (1) maximale interne Homogenität und (2) maximale externe Separation der Klassen erzielen. Während es viele Varianten gibt, so ist der typische Verlauf einer Clusteranalyse doch die schrittweise Aggregation von OTUs zu Clustern mit jeweils zunehmender Merkmalsunähnlichkeit (hierarchische Agglomeration), meist auf Basis einer verallgemeinerten euklidischen Metrik, jedoch mit unterschiedlichen Ähnlichkeitsforderungen (diverse linkage-Verfahren).

Zwei Beispiele sollen jetzt kurz den möglichen Ablauf bzw. die Anwendung einer Clusteranalyse veranschaulichen. Zunächst soll versucht werden, die Bezirke Nieder-

österreichs hinsichtlich ihrer (Un-) Ähnlichkeit bzgl. des Ertragsverhaltens anhand von 6 Feldfrüchten zu gruppieren. Der Verlauf des Aggregationsprozesses ist aus dem nachstehenden Dendrogramm ersichtlich (Abbildung 9):

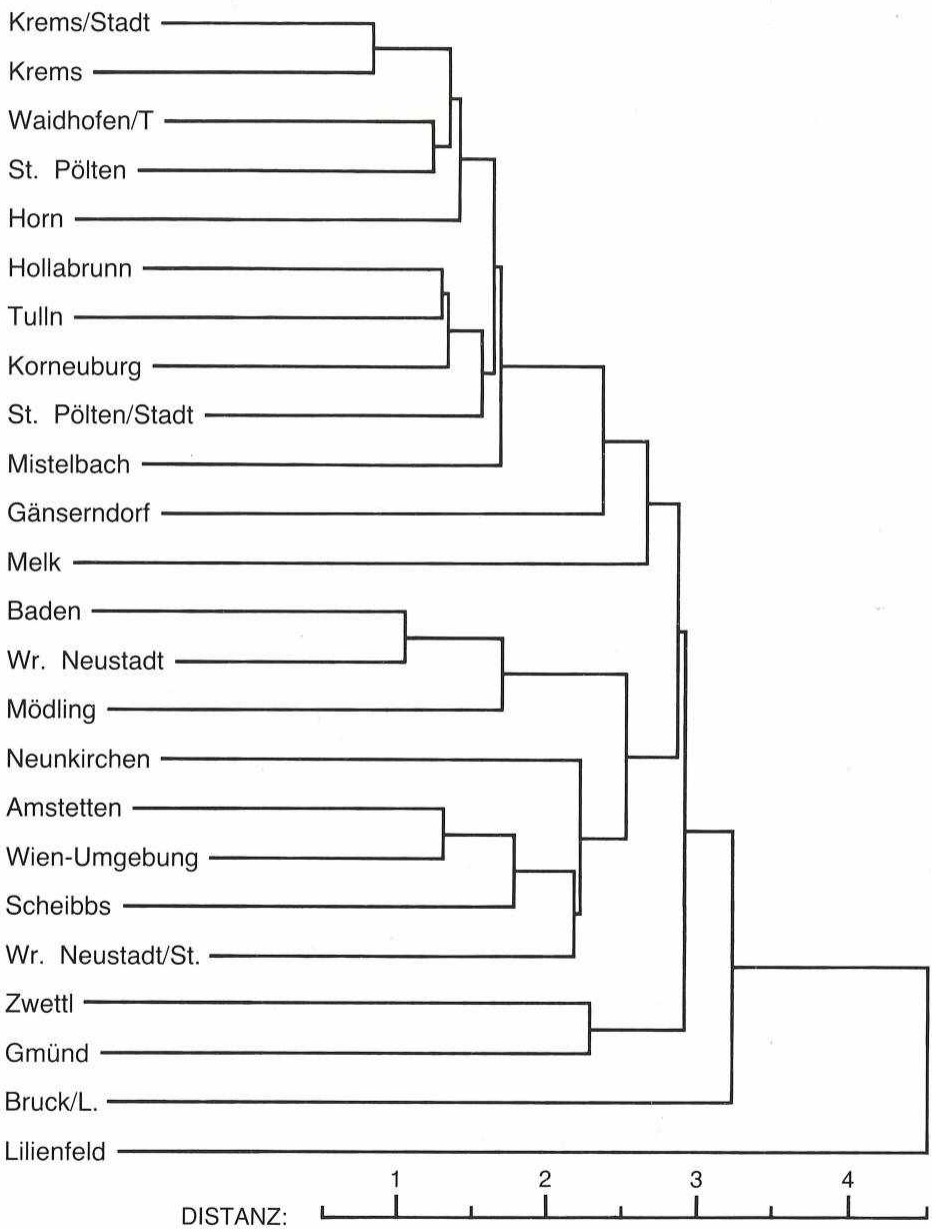

Abbildung 9: Dendrogramm einer Clusteranalyse niederösterreichischer Bezirke hinsichtlich der Ähnlichkeit des Ertragsverhaltens

Dabei ergibt sich die Frage nach der optimalen Clusterzahl und dem Punkt im Verlauf der Agglomeration mit möglichst intern homogenen und wechselseitig differenzierten Gruppen. Eine Entscheidung ist anhand numerischer Kriterien unter Beachtung sachlogischer Aspekte möglich. Eine Hilfe dabei ist ein Graph des internen Inhomogenitätszuwachses von Clustern, wobei steile Anstiege der Kurve zu einem Schnitt auf dem jeweiligen Distanzniveau einladen (Abbildung 10).

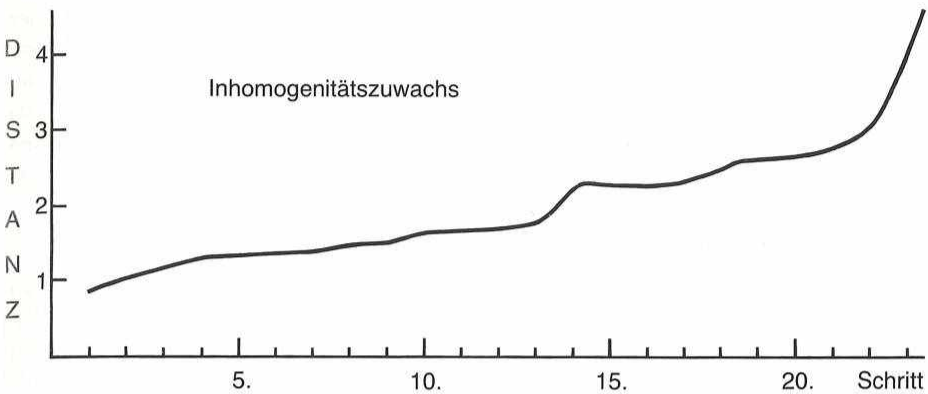

Abbildung 10: Inhomogenitätszuwachs der Klassen im Verlauf der Aggregation

In einem 2. Beispiel sollen die Merkmale (Variable) selbst auf Grund ihrer Ausprägungen in den einzelnen Bezirken, d. h. ihrer wechselseitigen Interkorrelation, zusammengefaßt werden. Es sollen also 6 verschiedene Feldfrüchte nach der Ähnlichkeit ihres Ertragsverlaufs über 10 Jahre in Niederösterreich geordnet werden. Das Resultat ist ein einfaches Dendrogramm ohne Substrukturen (Abbildung 11):

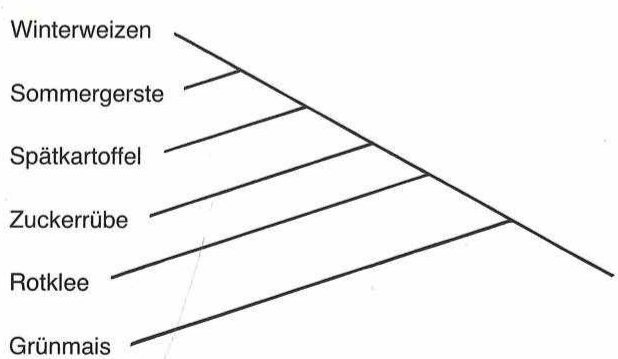

Abbildung 11: Gruppierung der Feldfrüchte nach der Ähnlichkeit ihres Ertragsverlaufs

Es läßt sich also eine Sequenz ähnlicher Reaktionen im Ertragsverlauf auf die Umwelteinflüsse aufstellen: Winterweizen – Sommergerste – Spätkartoffel – Zuckerrübe – Rotklee – Grünmais, wobei die Verlaufskurven der Ertragsausprägung zunehmend unähnlich werden. Damit konnten sicherlich nur kurze Hinweise auf die Mög-

lichkeiten im Rahmen der Clusteranalyse gegeben werden, die mit ein wichtiges Instrument statistischer Ernteertragsanalyse sein kann. Eine gewisse Vertrautheit mit klassifikatorischen Grundproblemen ist für eine sinnvolle Anwendung und Auswahl der vielen zur Verfügung stehenden Optionen von großer Bedeutung.

6. ABSCHLIESSENDE BEMERKUNGEN

Bei der gegebenen Breite des Verfahrensspektrums kann natürlich keine detaillierte Charakteristik individueller Techniken gegeben werden. Eine tiefergehende Herleitung und Behandlung der Verfahren ist in meiner Diplomarbeit enthalten, während hier der Schwerpunkt auf die Anwendungsbeispiele gelegt werden sollte. Diese haben demonstrativen Charakter, sind also z. T. nicht voll durchgearbeitet und daher die numerischen Ergebnisse auch unter diesem Aspekt zu sehen. Schärfere Aussagen wären wohl in vielen Fällen durch weitere Disaggregation zu erzielen gewesen, es wurde aber der Möglichkeit zu flächendeckenden Aussagen der Vorzug gegeben.

Nahezu alle diskutierten Techniken sind unter 2 Gesichtspunkten zu betrachten:

1. Auch multivariate statistische Verfahren sind häufig „nur" Instrumente zur Datenaufbereitung, Aggregation und inhaltlichen Visualisierung. In diesem Fall ist das Resultat einer Statistikprozedur natürlich keineswegs mit dem eigentlichen Untersuchungsergebnis ident, sondern auf eine reine Hilfsrolle beschränkt.

2. Für die eigentliche Identifikation und Quantifizierung von Zusammenhängen beruht die Verfahrensauswahl auf der jeweiligen Fragestellung. Es existiert kein allgemein der Untersuchung des Zusammenhangs Klima/Ertrag zuzuordnendes Verfahren. Der hier gegebene Überblick soll den Weg zu einer rein problem- und sachorientierten Auswahl des Instrumentariums auf der Basis der gemachten Erfahrungen erleichtern. Allzuoft noch ist die Verfahrensauswahl von persönlichen Vorkenntnissen und formellen Analogien bestimmt und wäre bei größerem Überblick ganz anders ausgefallen.

Anschließend an und ergänzend zur Wertung der diversen Methoden in den entsprechenden Abschnitten sollen hier einige Aussagen, natürlich relativiert auf den gewählten Problemkontext und Maßstab, kurz gegenübergestellt werden:

Wie ja durch die verbreitetste Anwendung ohnehin ersichtlich, sind alle Arten der Korrelations- und Regressionsanalyse, ob es sich nun um bi- oder multivariate, partielle oder schrittweise Anwendungen handelt, wohl kaum auf bestimmte Gebiete einzuschränken und von universeller Brauchbarkeit. Viele Detailfragestellungen können unmittelbar beantwortet werden, während sonst der Applikationsschwerpunkt bei Datentransformation und im initial-deskriptiven Stadium einer Arbeit liegen wird. Die vergleichsweise einfache Konzeption bietet große Sicherheit bei der Anwendung und die geringsten Probleme bei der Interpretation der Ergebnisse.

Das der multiplen Regression eng verwandte Gebiet der Trendflächenanalyse ist vor allem durch die Möglichkeit zur direkten graphischen Umsetzung einer geographischen Untersuchung besonders gelegen. Beim Übergang zum nicht nur formal, sondern auch inhaltlich multivariaten Fall geht dieser Vorzug allerdings verloren und das Verfahren wird zu einer durch Hinzufügen der räumlichen Dimension komplexer gestalteten Verallgemeinerung regressiver Ansätze. Dem großen Gebiet der Faktorenanalyse kann man in diesem Rahmen auch hinsichtlich eines einzigen Anwendungsgebiets kaum gerecht werden und es nur vergleichend im Gesamtzusammenhang mit

anderen multivariaten Techniken diskutieren. Als explorativer Ansatz herrschen Anwendungen zur Datenauswahl und -transformation zu Beginn einer Untersuchung vor, während beim Übergang zur konfirmatorischen Analyse weitere Prämissen erfüllt sein müssen. Der Einsatz als vorgeschaltete Datentransformationstechnik erschwert die Interpretation der Endergebnisse beträchtlich und ist deshalb nicht unbedingt zu empfehlen.

Ein formal vielversprechendes, aber nicht unproblematisches Verfahren ist die kanonische Korrelationsanalyse. Die sehr allgemein gehaltene Konzeption ermöglicht weite Anwendungsbereiche, die komplexe verfahrenstechnische Struktur erschwert jedoch eine sinnvolle Interpretation in vielen Fällen und läßt die Resultate als enttäuschend erscheinen. Eine fundierte inhaltliche Theorie muß noch mehr als bei anderen Verfahren mit der Analyse einhergehen, um ein thematisch orientiertes Nachvollziehen der einzelnen Verfahrensschritte zu ermöglichen.

Speziell geographische Ziele sind für viele Spielarten der Clusteranalyse denkbar, auch hier konnte aber nur ein sehr allgemein gehaltener Überblick vermittelt werden. Eine anschauliche Darstellung von Ähnlichkeitsstrukturen und klassifikatorisch aufbereitete Ergebnisse sind in nahezu allen Stadien einer Untersuchung von großer Bedeutung. Dazu kommen noch spezifische Anwendungen als Typisierungs- und Regionalisierungsinstrument.

In vielen dieser besprochenen Verfahren steht die Komplexität der theoretischen und verfahrenstechnischen Basis in keiner Relation zur äußerlich-formalen Anwendbarkeit im Rahmen vorgefertigter Statistikprogrammsysteme. Gerade dies verleitet zu einem unreflektierten „probeweisen" Berechnen statistischer Maßzahlen ohne entsprechende Grundlage für deren Interpretation. Es erfordert also jede einzelne der hier vorgestellten Techniken noch eingehende Vorbereitung und Theorieformulierung.

Generell ist ein paralleles Verhalten zwischen den möglichen Leistungen eines Verfahrens und dem Komplexitätsgrad seiner Konzeption zu beobachten, dem meist eine erschwerte Interpretierbarkeit und erhöhte Theorieerfordernisse entgegenstehen. Qualität der Daten und theoretischen Basis sowie Vertrautheit mit formaler Konzeption der einzelnen Techniken werden also die Wahl zwischen einfacheren und robusteren sowie technische anspruchsvolleren und damit potentiell leistungsfähigeren Verfahren bestimmen. Eine Absicherung von Ergebnis und Interpretation durch vergleichende Analyse mit unterschiedlichen Verfahren ist, soweit möglich, jedenfalls anzuraten. Anwendungsorientierte methodische Arbeiten wie die hier kurz dokumentierte Diplomarbeit erachte ich jedenfalls für ein wichtiges Bindeglied zwischen Formalwissenschaft und problembezogener Einzeluntersuchung, deren theoretisch und verfahrenstechnisch bessere Fundierung auch zu besseren Resultaten führen sollte. Damit soll aber auch der Mittelweg zwischen methodisch und theoretisch schwach unterbauter Untersuchung und andererseits Verfahrensartistik mit wenig thematischem Bezug hervorgehoben werden.

LITERATURVERZEICHNIS

ABLER, R., J. S. ADAMS und P. GOULD (1977): Spatial Organisation – The Geographers View of the World. Prentice Hall, New Jersey.
ANDERSON, T. W. (1958): An Introduction to Multivariate Statistical Analysis. Wiley, New York.
BAEUMER, K. (1978): Allgemeiner Pflanzenbau. Ulmer, Stuttgart.

BAHRENBERG, G. et. al. (1978): Methoden der Quantitativen Geographie. In: Werkstattpapiere, Nr. 5. Gg. Institut der Justus Liebig-Universität, Gießen.
BAHRENBERG, G. und E. GIESE (1975): Statistische Methoden und ihre Anwendung in der Geographie. Teubner, Stuttgart.
BERRY, B. J. L. und D. F. MARBLE (Hrsg.), (1968): Spatial Analysis, A Reader in Statistical Geography. Prentice Hall, Englewood Cliffs, N. J.
BEUTEL, P., KÜFFNER, H., RÖCK, E. und W. SCHUBOE (1978): SPSS 7, Statistik-Programmsystem für die Sozialwissenschaften. G. Fischer, Stuttgart.
BINDER, J. und K. M. ORTNER (1978): Die Abhängigkeit der Erträge vom Witterungsverlauf. Österreichischer Agrarverlag, Wien.
BLÜTHGEN, J. (1966): Allgemeine Klimageographie. Walter de Gruyter, Berlin.
CEHAK, K. (1978): Allgemeine Meteorologie. Prugg, Wien.
CLARK, D. (1975): Understanding Canonical Correlation Analysis. CATMOG Nr. 3, Geo Abstracts, Norwich.
COOLEY. D. E. und J. K. LOHNES (1971): Multivariate Data Analysis. Wiley, New York.
DAULTREY, S. (1976): Principal Components Analysis. CATMOG Nr. 8, Geo Abstracts, Norwich.
DE GEER V. S. (1971): Introduction to Multivariate Analysis for the Social Sciences. Freeman, San Francisco.
DIXON, W. J. und M. B. BROWN (1979), Hrsg.: BMDP-79, Biomedical Computer Programs, P-Series. University of California Press, Berkeley.
DROTH, W. und M. M. FISCHER (1980): Zur Theoriebildung und Theorietestung. Eine Diskussion von Grundlagenproblemen am Beispiel der Sozialraumanalyse. Hamburg und Wien.
FERGUSON, R. (1977): Linear Regression in Geography. CATMOG Nr. 15, Geo Abstracts, Norwich.
FISCHBECK, G., K.-U. HEYLAND und N. KNAUER (1975): Spezieller Pflanzenbau. Ulmer, Stuttgart.
FISCHER, M. M. (1978): Theoretische und methodische Probleme der regionalen Taxonomie. In: Bremer Beiträge zur Geographie und Raumplanung, Heft 1: Quantitative Methoden in Geographie und Raumplanung (Hrsg.: G. BAHRENBERG und W. TAUBMANN). Universität Bremen.
GAENSSLEN, H. und W. SCHUBOE (1976): Einfache und komplexe statistische Analyse. Ulmer, München.
GILCHRIST, W. (1976): Statistical Forecasting. Wiley, London.
GODDARD, J. und A. KIRBY (1976): An Introduction to Factor Analysis. CATMOG Nr. 7, Geo Abstracts, Norwich.
HARD, G. (1973): Die Geographie – Eine wissenschaftstheoretische Einführung. Walter de Gruyter, Berlin.
HUGGETT, R. (1980): Systems Analysis in Geography. Clarendon Press, Oxford.
JOHNSTON, R. J. (1976): Classification in Geography. CATMOG Nr. 6, Geo Abstracts, Norwich.
KING, L. J. (1969): Statistical Analysis in Geography. Prentice Hall, Englewood Cliffs, N. J.
LACKO, L. und H. EXNER (1977): Die kanonische Korrelationsrechnung – eine Methode für die Regionalanalyse. AMR-Forum, Wien.
LOTZE, K. P. (1975): Korrelationsanalyse. In: Methoden der empirischen Sozialforschung (2. Teil). Veröffentlichungen der Akademie für Raumforschung und Landesplanung, Forschungs- und Sitzungsberichte Bd. 105, Hannover.
NIE, N. H., C. H. HULL, J. G. STEINBRENNER und D. H. BEND (1975): SPSS-Statistical Package for the Social Sciences. McGraw Hill, New York.
NORCLIFFE, G. B. (1977): Inferential Statistics for Geographers. An Introduction. Hutchinson, London.
NORUSIS, M. J. (1980): SPSS 8.0 – Statistical Algorithms. McGraw Hill, New York.
OPITZ, O. (1980): Numerische Taxonomie. G. Fischer, Stuttgart.
OSTHEIDER, M. (1979): Multivariate Analyse für Geographen, Teil 1: Statistische Grundlagen, Lineare Regression. In: Berichte und Skripten zur Quantitativen Geographie. Geographisches Institut, ETH Zürich.
RAO, C. R. (1973): Lineare Statistische Methoden und ihre Anwendung. Berlin.
RASE, W. D. (1976): SYMAP-Handbuch, Version 5.20. EDV-Report 2/76 Bundesforschungsanstalt für Landeskunde und Raumordnung, Bad Godesberg.
STEINER, D. (1977): Einführung in die Quantitative Geographie (Vorlesungsskriptum). Geographisches Institut der ETH Zürich.
STEINER, D. und H. GILGEN (1977): Beiträge zur Trendflächenanalyse. Geographisches Institut der ETH Zürich.
STEINER, D. (1979): Multivariate Analyse für Geographen, Teil II: Faktorenanalyse. In: Berichte und Skripten zur quantitativen Geographie. Geographisches Institut der ETH Zürich.
STROBL, J. (1981): Anwendungen multivariater statistischer Verfahren auf die Abhängigkeit des landwirtschaftlichen Ernteertrags vom Klima am Beispiel von Niederösterreich. Diplomarbeit: Wien, Institut für Geographie.
ÜBERLA, K. (1968): Faktorenanalyse. Springer, Berlin.
UNWIN, D. (1975): An Introduction to Trend Surface Analysis. CATMOG Nr. 5, Geo Abstracts, Norwich.
WIRTH, E. (1979): Theoretische Geographie. Teubner, Stuttgart.

Summary

The Application of Multivariate Statistical Methods on the Dependency of Agricultural Yields from Climate in Lower Austria.

As a methodological contribution to agroclimatology the presented paper intends to compare a number of widely used statistical algorithms. Examples of possible applications are emphasized, whereas the reader is referred to the complete text by

STROBL (1981) for further details. The main purpose of all these numerical operations is the quantification of structurally postulated relations and as a distant aim the forecasting of seasonal weather effects on agricultural yields.

To show the pros and cons of standard multivariate techniques in the given context, the process and possible results of different analyses are shortly described. There is a need to differenciate between a merely descriptive application aiming at preparation, aggregation and visualizing of contents of data and the actual identification and quantification of structural relationships. The potentially far-reaching conclusions from simultaneous consideration of multiple relations within arrays of variables by complex methods are opposed by the restrictions of more rigorous premises and frequently difficult interpretation. These facts must be kept in mind even if the computation of highly aggregated statistics is simplified and speeded up by packaged programs.

Certainly no single technique can be generally preferred as specially suitable for agroclimatical issues, so that a proper aquaintance with available routines is desirable. Knowing the special features of different methods through extensive applications regarding the particular structure of data and relations in agroclimatology enables a far better founded choice of techniques for acute questions. Well-balanced standards of methodological knowledge and insight into causal relations are certainly the most important prerequisites for well reproducible results in quantitative analyses.

Österreichischen Geographischen Gesellschaft

Jahr/Year: 1983

Band/Volume: 125

Autor(en)/Author(s): Strobl Josef

Artikel/Article: Die Anwendung multivariater statistischer Verfahren auf die Abhängigkeit des landwirtschaftlichen

Ernteertrages vom Klima am Beispiel von Niederösterreich 31-57