SETHURAMAN, S. V.: The Urban Informal Sector: Concept, Measurement and Policy. World Employment Programme, Working Paper No. 7, International Labour Office, Genf 1976.
SINHA, D. P.: Culture Change in an intertribal market. The role of the Banari intertribal market among the hill peoples of Chota Nagpur. London 1967.
SMITH, R. H. T. (Hrsg.): Market-Place Trade – Periodic Markets, Hawkers, and Traders in Africa, Asia, and Latin America. Vancouver 1978.
STINE, J. H.: Temporal aspects of tertiary production elements in Korea. In: Urban Systems and Economic Development, hrsg. von Pitts, F. R., Eugene 1962.
SYMANSKI, R.: Periodic Markets in Southern Columbia. In: Market-Place Trade – Periodic Markets, Hawkers, and Traders in Africa, Asia, and Latin America, hrsg. von Smith, R. H. T., Vancouver 1978, S. 171–185.
–, WEBBER, M. J.: Complex Periodic Market Cycles. In: Ann. of the Ass. of American Geographers, 64, 1974, S. 203–213.
TAMASKAR, B. G.: The Weekly Markets of the Sagar-Damoh Plateau. In: National Geographical Journal of India, 12, 1966, S. 38–50.
TINKLER, K. J.: The topology of rural periodic market systems. In: Geografiska Annaler, 55 B, 1973, S. 121–133.
UHLIG, H.: Agrarlandschaften im westlichen Himalaya: Kulu-Mandi-Kangra (Himachal Pradesh). In: Kölner Geogr. Arb., Sonderband 1971, S. 458–481.
WANMALI, S.: Market Centres and Distribution of Consumer Goods in Rural India. A Case Study of Singhbum District, South Bihar. In: Mainzer Geogr. Studien, 10, 1976, S. 49–56.
WARD, R. G. et al.: Maket Raun: the introduction of periodic markets to Papua New Guinea. In: Market-Place Trade – Periodic Markets, Hawkers, and Traders in Africa, Asia, and Latin America, hrsg. von Smith, R. H. T., Vancouver 1978, S. 99–111.
WEBBER, M. J. und SYMANSKI, R.: Periodic Markets: An Economic Location Analysis. In: Economic Geography, 49, 1973, S. 6–46.
WEIGT, E.: Der trockene Südosten Indiens. Mensch und Wirtschaft im Tambraparni-Tal. In: Geographische Rundschau, 20, 1968, S. 405–414.
WIRTH, E.: Zur Theorie periodischer Märkte aus der Sicht von Wirtschaftswissenschaften und Geographie. In: Erdkunde, 30, 1976, S. 10–15.
World Agricultural Census, Salem District, 1970/71, Madras 1975.

BERICHTE UND MITTEILUNGEN

EINIGE ANMERKUNGEN ZUR REGRESSIONSANALYSE RAUMBEZOGENER DATEN – ERLÄUTERT AM BEISPIEL EINER NIEDERSCHLAGS-ABFLUSS-REGRESSION

Mit 4 Tabellen

ULRICH STREIT

Summary: Some remarks on regression analysis of spatial data – explained by an example of rainfall-runoff regression.

Regression analysis belongs to the most frequently applied statistical methods in geography. A basic but mostly ignored assumption of this model is the independence among the regression residuals. Concerning spatial variates this means the spatial autocorrelations to be zero. Otherwise all important, statistics of the regression equation may be misestimated, which leads to incorrect inferences and nonoptimal applications. Based on the geographically important concept of spatial autocorrelation, an appropriate procedure to detect spatial dependencies in regression residuals is summarized (CLIFF/ORD, 1973).

A hydrologic example of regression analysis between A = runoff and N = rainfall (mm/a) – defined on a 6 × 6 lattice in the upper Fulda region (FRG) – points out the methodological importance as well as the geographical implications of significant spatial autocorrelations among the residuals and the regression variables themselves. Hydrologic reflections and various computations recommend acceptance of a linear trend surface model combined with the predictor N. Another solution, including spatially lagged variables i. e. an autoregressive component, is discussed and shown to be effective too.

Die Anwendung statistischer Verfahren erfordert stets die Annahme einschränkender Bedingungen, unter denen das statistische Modell Gültigkeit besitzt. Selbst geringe Abweichungen von diesen Voraussetzungen führen – bei strenger Sichtweise – zum Einsturz des darauf aufbauenden theoretisch-mathematischen Konstruktes. Reale geowissenschaftliche Gegebenheiten genügen solchen Bedingungen in aller Regel nur unvollkommen; dennoch wird sich der Fachwissenschaftler häufig derartiger Instrumente bedienen müssen, um wenigstens eine approximative Lösung des anstehenden Problems zu erreichen. Als Anwender statistischer Verfahren wird er sich aber stets der latenten Gefahr einer damit verbundenen Fehleinschätzung erzielter Resultate bewußt sein.

In diesem Sinne sollten die folgenden Ausführungen als Hinweis auf ein gelegentlich übersehenes, häufiger jedoch schlicht verdrängtes und nur vordergründig rein statistisch-methodisches Problem betrachtet werden.

1. *Problemstellung*

Die Regressionsanalyse zählt zu den in der Geographie recht häufig angewendeten Methoden der Statistik. Bei Annahme einer bestimmten, meist bzgl. der Parameter linearen Form des Zusammenhanges zwischen einer zufallsbeeinflußten Zielvariablen Z und einer oder auch mehreren Prädiktorvariablen P_j ($j = 1, \ldots m$) ermöglicht sie unter gewissen Voraussetzungen eine optimale Schätzung von Zielvariablenwerten bei gegebenen Prädiktorenwerten im Sinne einer minimalen Schätzfehlervarianz.

Der formale Ansatz der multiplen linearen Regression in einer Stichprobe von n Elementen (z. B. Raumeinheiten) lautet:

$$z_i = b_o + b_1p_{1i} + b_2p_{2i} + \ldots + b_mp_{mi} + e_i (i = 1, \ldots, n) \quad (1)$$

Aus den gegebenen n Wertetupeln für die Zielvariable Z und die Prädiktoren $P_1, \ldots, P_m$

$$\left\{(z_i, p_{1i}, \ldots, p_{mi})/i = 1, \ldots, n\right\} \quad (2)$$

können die unbekannten Regressionsparameter $b_o, b_1 \ldots, b_m$ so berechnet werden, daß die Summe der quadrierten Fehler (Residuen) e_i möglichst klein ist:

$$\sum_{i=1}^{n} e_i^2 = \text{minimal} \quad (3)$$

Die Herleitung der Berechnungsformeln für die Regressionskoeffizienten erfolgt auf rein mathematischem Wege. Zu einem statistischen Verfahren wird diese numerische Technik erst durch die Betrachtung der Stichprobe (2) als eine repräsentative Teilmenge der betreffenden Grundgesamtheit und der damit verbundenen Auffassung, daß die Regressionskoeffizienten und alle anderen aus der Teilmenge (2) berechneten statistischen Kennwerte lediglich Schätzungen für die unbekannten statistischen Parameter der zugehörigen Grundgesamtheit sind. Geeignete Testverfahren erlauben dann Aussagen über die Signifikanz des Regressionsmodells und die Güte einer damit errechneten Schätzung. Die dazu notwendigen Modellannahmen über gewisse Eigenschaften in der Grundgesamtheit beziehen sich vor allem auf die Regressionsresiduen, für die außer Normalverteilung mit Mittelwert Null und homogener Varianz[1] auch die paarweise Unkorreliertheit gefordert werden muß (z. B. Bahrenberg/Giese, 1975, 139 und 200f; Gaensslen/Schubö, 1973, 45ff.).

Bei der Regression zeitlich oder räumlich bezogener Variablen bedeutet dies, daß zwischen benachbarten Residualwerten keine systematisch auftretende Ähnlichkeit, also Autokorrelation existieren darf. Verletzt ist diese Voraussetzung sicherlich z. B. dann, wenn bei der Regression zwischen Zeitreihen die Residuen einen Trend aufweisen, oder wenn bei der regressiven Verknüpfung raumbezogener Variablen eine Tendenz zur räumlichen Klumpung gleichartiger Werte besteht. Bei derartiger „stochastischer" Abhängigkeit der Regressionsresiduen verliert die aufgestellte Regressionsgleichung ihre statistische Bedeutung, vor allem aber auch praktische Anwendbarkeit. Liegt nämlich, wie dies bei empirischen Analysen häufig der Fall ist, positive Autokorrelation vor, so

- ist die Schätzung der Regressionsparameter nach der üblichen Methode der „Fehlerquadratsummen-Minimierung" (3) nicht mehr die wirksamste, d. h. sie verliert die wichtige Minimumeigenschaft bzgl. der Schätzvarianz;
- werden eben diese Schätzvarianzen für die Modellparameter systematisch unterschätzt und somit die Signifikanzen der Prädiktorvariablen zu optimistisch beurteilt;
- wird die Residualvarianz zu gering geschätzt und damit das Bestimmtheitsmaß der Regression zu hoch angenommen.

Bei negativer Autokorrelation der Residuen ergeben sich die umgekehrten Effekte. Verstärkt tritt dieses Problem dann auf, wenn die Prädiktoren selbst noch in der gleichen Richtung autokorrelieren (Martin, 1974; Haggett/Cliff/Frey, 1977, 334f.).

Die negativen Konsequenzen bei der Anwendung solcher Regressionsgleichungen können also gravierend sein: Verwendung eines ineffizienten Regressionsmodelles, obwohl evtl. ein den theoretischen wie fachlichen Anforderungen besser genügender Ansatz gefunden werden könnte; Vortäuschung „sicherer" Interpolations- oder Prognosewerte mit zu engen Konfidenzintervallen und einer bei Kreuzvalidierungen nicht reproduzierbaren hohen Korrelation zwischen gemessenen und berechneten Werten der Zielvariablen.

Das Aufdecken einer signifikanten Autokorrelation in zeitlich (allgemein: eindimensional) anordnungsfähigen Regressionsresiduen ist relativ einfach mit Hilfe des d-Tests von Durbin/Watson (1951) möglich; enthält der Regressionsansatz zusätzlich eine autoregressive Komponente (z. B. Lufttemperatur zur Zeit i in Abhängigkeit von der Globalstrahlung im gleichen Zeitintervall sowie der Lufttemperatur des vorangehenden Zeitabschnittes i-1), so sollte der auf der d-Statistik aufbauende h-Test nach Durbin (1970) verwendet werden (s. auch Bennett, 1979, 256ff.).

Der bei geographischen Untersuchungen häufiger auftretende Fall einer Regressionsanalyse raumbezogener (allgemein: zwei- und mehrdimensional lokalisierter) Variablen bereitet nicht nur von der Testtheorie her infolge der hier fehlenden „natürlichen" Ordnungsrelation größere Schwierigkeiten; er ist auch – im Gegensatz zur allgemein verbreiteten Vorsicht bei der regressionsanalytischen Bearbeitung von Zeitreihen – als Problemfall in den die Regressionsanalyse ansprechenden deutschsprachigen Publikationen bisher zumeist unbemerkt oder unbeachtet geblieben (dazu: Bahrenberg/Giese, 1975, 200f.; Nipper/Streit, 1977, 243; Streit, 1979, 64).

In den folgenden Ausführungen soll daher ein von Cliff/Ord (1972) entwickelter Residuen-Test vorgestellt und seine Wirksamkeit an einem leicht nachvollziehbaren hydrologischen Beispiel demonstriert werden.

2. *Das Konzept der räumlichen Autokorrelation*

Für das Verständnis dieses Prüfverfahrens ist eine kurze Einführung in das allgemeine Konzept der räumlichen Autokorrelation nach Cliff/Ord (1973) förderlich; eine ausführlichere Darstellung geben Nipper/Streit (1977).

In Analogie zur Zeitreihen-Terminologie spricht man von positiver (negativer) Autokorrelation k-ter Ordnung in einer räumlichen Zufallsvariablen[2] Z, wenn die zu einem betrachteten Stichprobenwert z_i über eine bestimmte Raumschrittweite k (= 1, 2 ...) benachbarten z_j tendenziell gleich große (stark unterschiedliche) Werte aufweisen.

Für Zeitreihen ist die statistische Erfassung solcher Persistenzeffekte durch Autokorrelationskoeffizienten infolge

---

1) bezogen auf die jeweils bedingten Residualverteilungen zu jedem Prädiktorenwerte-Tupel.

2) genau genommen „Prozeß"; jedoch soll hier auf den Rückgriff in die Theorie raumvarianter stochastischer Prozesse verzichtet werden. Es sei jedoch erwähnt, daß im folgenden schwache räumliche Stationarität des (metrischen) Prozesses angenommen wird, d. h. Mittelwerte und Varianzen sind räumlich invariant und die Kovarianz hängt nicht von der absoluten Lage sondern lediglich von der relativen Position der Raumeinheiten zueinander ab. Der einfacheren Darstellung halber sei ebenfalls vereinbart, daß der Prozeß auf den Mittelwert Null normalisiert ist.

der natürlichen zeitlichen Anordnung und damit eindeutigen Bestimmbarkeit zeitlich benachbarter Werte relativ einfach als normierte Kovarianz der über k Zeitschritte benachbarten Stichprobenwerte möglich. Bei räumlichen Bezugseinheiten können Nachbarschaftseffekte jedoch in beliebig vielen Richtungen und über verschiedene – nicht notwendigerweise metrische – Distanzen hinweg auftreten. Daher ist jede Definition von räumlichen Autokorrelationskoeffizienten stets an eine vorzugebende Strukturierung des Raumes gebunden, d. h. es muß für jede Raumschrittweite k ein Kriterium vereinbart werden, nach dem zwei Raumeinheiten als benachbart bezüglich dieser Raumschrittweite zu gelten haben. Das Aufdecken einer räumlichen Autokorrelation hängt damit entscheidend von der sinnvollen Definition solcher Nachbarschaftskriterien ab: Sie sollten entsprechend der jeweiligen Fragestellung inhaltlich begründet, also aus der gesicherten Kenntnis oder hypothetischen Annahme über Richtung und Stärke räumlicher Transfer- und Interaktionsvorgänge zwischen den Raumeinheiten abgeleitet werden.

Eine Formalisierung solcher Nachbarschaftsbeziehungen ist am einfachsten über „Gewichte" $w_{ij}^{(k)}$ möglich, wobei

$$w_{ij}^{(k)} = \begin{cases} >0, & \text{falls die Raumeinheiten } R_i \text{ und } R_j\ (i \neq j) \\ & \text{bzgl. der Raumschrittweite k als Nachbarn gelten sollen;} \\ =0 & \text{sonst.} \end{cases} \tag{4}$$

Im Falle der Existenz einer Nachbarschaftsbeziehung zwischen R_i und R_j kann zusätzlich eine eventuell unterschiedliche Intensität ihrer Bindung durch die Größe des Wertes $w_{ij}^{(k)}$ angegeben werden.

Als einfaches Beispiel dazu sei das im nächsten Abschnitt verwendete Nachbarschaftskriterium für die Teilflächen (= Raumeinheiten) eines regelmäßig-rechteckigen Gitternetzes angeführt: Ist der die räumlichen Daten erzeugende Prozeß annähernd isotrop, und ist außerdem anzunehmen, daß die ihn steuernden Vorgänge eine distanzabhängige Veränderlichkeit aufweisen, so kann die folgende Nachbarschaftsdefinition sinnvoll sein:

Zwei Raumeinheiten R_i und R_j heißen Nachbarn bezüglich der Raumschrittweite $k = 1$, wenn sie der räumlichen Kontingenzbedingung genügen, ihre Grenzen also mindestens einen gemeinsamen Punkt besitzen („ecken- und kantenbenachbart" im Gitternetz). (5)
Zwei Raumeinheiten R_i und R_j heißen Nachbarn bezüglich der Raumschrittweite $k > 1$, wenn R_i im eben definierten Sinne einen Nachbarn der Ordnung $k = 1$ besitzt, der zugleich ein Nachbar der Ordnung $k - 1$ zu R_j ist.

Eine in der Mitte des Gitternetzes gelegene Raumeinheit besitzt demnach 8 Nachbarn bzgl. der Raumschrittweite $k = 1$ und 16 Nachbarn bzgl. $k = 2$ usw. Verzichtet man auf eine Intensitätsdifferenzierung innerhalb jeder Raumschrittweite, so erhalten alle k-ten Nachbarn von R_i das gleiche Gewicht, also etwa $w_{ij}^{(k)} = 1$.

Weitere Beispiele zur inhaltlichen Ableitung und formalen Definition solcher Nachbarschaftsgewichte diskutieren NIPPER/STREIT (1977, 249–254).

Für jede Raumschrittweite k werden diese Gewichte $w_{ij}^{(k)}$ zweckmäßigerweise in Form einer Nachbarschaftsmatrix $W_{nn}^{(k)}$ so angeordnet, daß in der i-ten Zeile alle $w_{ij}^{(k)}$ ($j = 1, \ldots, n$) positioniert sind; zur Vereinfachung späterer Berechnungen ist es außerdem sinnvoll, diese Matrizen zeilenweise zu normieren, so daß:

$$\sum_{j=1}^{n} w_{ij}^{(k)} = 1 \quad \text{für alle k, } (i = 1, \ldots, n) \tag{6}$$

Definiert man nun einen um k Raumschritte gegenüber dem Merkmalswert z_i in der Raumeinheit R_i transponierten neuen Merkmalswert $z_i(k)$ als gewichtete Information aller derjenigen Raumeinheiten R_j, die bzgl. R_i gerade k-te Nachbarn sind (= Menge ${}_iR_j$), also

$$z_i(k) = \sum_{j \in {}_iR_j} w_{ij}^{(k)} z_j = \sum_{j=1}^{n} w_{ij}^{(k)} z_j \quad \text{für alle } i = 1, \ldots, n \text{ und jedes k,} \tag{7}$$

so kann in Analogie zum zeitlichen Autokorrelationskoeffizienten die Maßzahl

$$I(k) = \sum_{i=1}^{n} z_i z_i(k) \Big/ \sum_{i=1}^{n} z_i^2 \tag{8}$$

als räumlicher Autokorrelationskoeffizient der Ordnung k angesehen werden. In seiner normierten Form

$$R(k) = I(k) \cdot [V(z_i) / V(z_i(k))]^{1/2}, \quad V = \text{Varianz} \tag{9}$$

besitzt er die für Korrelationskoeffizienten vertraute Eigenschaft

$$-1 \leq R(k) \leq +1 \quad \text{für alle k} \tag{10}$$

(s. CLIFF/HAGGETT/ORD/BASSETT/DAVIS, 1975, 158).

Die Signifikanz von I(k) kann bei hinreichend großem Stichprobenumfang asymptotisch über die Standardnormalverteilung geprüft werden. Das Verfahren zur Schätzung der dazu erforderlichen Verteilungsparameter schildern CLIFF/ORD (1973, 29–37).

Diese räumlichen Autokorrelationskoeffizienten können auch für die (normalverteilten) Residuen e_i ($i = 1, \ldots, n$) einer räumlichen Regression berechnet werden; ein dafür spezifischer Signifikanztest berücksichtigt die jeweils verwendeten Prädiktorvariablen (CLIFF/ORD, 1973, 93 ff.). Damit ist der Nachweis räumlicher Persistenzen in den Residuen raumbezogener Regressionsmodelle im Prinzip möglich.

An einem allgemein verständlichen hydrologischen Beispiel sollen nun das aufgezeigte Problem und seine Lösungsmöglichkeiten unter Verzicht auf mathematisch-statistische Detailfragen erörtert werden.

3. *Beispiel einer räumlichen Regressionsanalyse zwischen Niederschlag und Abfluß*

Die Abschätzung der mittleren jährlichen Abflußhöhe A(mm/a) für gewisse Teilflächen auf der Grundlage bekannter mittlerer Gebietsniederschläge N(mm/a) und ggf. weiterer Prädiktoren zählt zu den Standardaufgaben der Hydrologie; aber auch andere geowissenschaftliche Teildisziplinen (z. B. Geomorphologie, Bodenkunde) sowie die wasserwirtschaftliche Fachplanung und allgemeine Raumplanung sind an solchen räumlichen Prognoseinstrumenten interessiert.

Eine derartige raumbezogene Regressionsanalyse soll hier für ein regelmäßig-rechteckiges Raster mit 36 Teilflächen (Größe ca. 4 × 5 km) im Bereich der oberen Fulda vorgenommen werden.[3] Diese 36 Stichprobenwerte für die Zielvariable A und die Prädiktorvariable N sind in der Tabelle 1 aufgeführt; die Daten entstammen dem *Wasserwirtschaftlichen Rahmenplan Fulda* (1964).

Tabelle 1: Stichprobenwerte für die Zielvariable A (mm/a) und die Prädiktorvariable N (mm/a) in den 36 Raumeinheiten der räumlichen Stichprobe (zu lesen als A / N)

110/640	110/640	130/650	150/650	190/670	240/700
110/640	130/600	140/650	170/660	210/680	305/740
180/700	160/660	175/670	175/650	270/720	400/810
250/720	280/740	195/680	195/680	315/750	485/890
345/790	305/760	220/700	220/700	330/760	485/890
360/800	305/760	320/770	310/770	335/770	500/900

Da das Korrelogramm der (N, A)-Wertepaare deutlich auf einen linearen Zusammenhang hinweist, wird der Regressionsansatz

$$A_i = a + bN_i + e_i \; (i = 1, \ldots, n = 36) \qquad (11)$$

gewählt. Die Berechnung für die Geradenparameter ergibt $a = -773$ und $b = 1{,}42$ mit einem Bestimmtheitsmaß von $B = 0{,}97$.

Ohne Berücksichtigung einer eventuell vorhandenen räumlichen Autokorrelation in den Regressionsresiduen, die betragsmäßig zwischen 0 und 49 mm liegen, käme man auf Grund des hoch signifikanten Bestimmtheitsmaßes zu der Beurteilung, daß die räumliche Variation des Abflusses durch die Niederschlagswerte (statistisch) sehr gut erklärt wird. Eine Übertragbarkeit dieses Ansatzes als räumliches Prognosemodell etwa auf benachbarte Untersuchungsgebiete scheint daher unbedenklich zu sein.

Die Prüfung auf räumliche Autokorrelation geht in diesem Fall von der Annahme aus, daß sich die potentiellen Nachbarschaftsbeziehungen gleichmäßig in alle Richtungen der Ebene (bzw. in die 8 Hauptrichtungen des Gitternetzes) erstrecken. Diese Isotropie-Hypothese stützt sich auf die Ansicht, daß durch die starke zeitliche und – in geringerem Umfange – auch räumliche Glättung[4] Richtungsunterschiede in den steuernden Prozessen der Niederschlags- und Abflußbildung weitgehend verwischt werden. Die Berücksichtigung möglicher distanzieller Ähnlichkeitseffekte, wie sie z. B. beim Prozeß der Abflußbildung durch die relative Homogenität der naturräumlichen Steuerungselemente innerhalb kleiner Raumeinheiten auftreten können, soll durch Einbeziehung der Kontingenzbedingung für Nachbarn 1-ter Ordnung gesichert werden; sie führt auf rekursivem Wege zur Ausweisung unterschiedlicher Raumschrittweiten. Damit wird das Nachbarschaftskriterium gemäß (5) definiert.[5] Als maximale Raumschrittweite soll für die folgenden Berechnungen stets $k = 3$ verwendet werden; damit ist sichergestellt, daß bei jeder Raumschrittweite k jede Raumeinheit noch mindestens einen Nachbarn k-ter Ordnung besitzt.[6]

Die gemäß (9) errechneten räumlichen Autokorrelationskoeffizienten sowohl für die Regressionsresiduen als auch die Regressionsvariablen sind in der Tabelle 2 zusammengestellt. Die Beurteilung der Signifikanz erfolgt mit einer Irrtumswahrscheinlichkeit von 5% (zweiseitiger Test).

Tabelle 2: Räumliche Autokorrelationskoeffizienten R (k) für die Residuen des linearen Ansatzes $A = a + bN + e$ sowie für die Variablen A und N (+ = signifikant auf dem 5%-Niveau)

RSW k	Residuen e	Zielvariable A	Prädiktor N
1	0,16	0,93+	0,89+
2	0,59+	0,51+	0,48+
3	−0,16	−0,31	−0,33

Die Residuen des Regressionsmodells (11) zeigen danach in der ersten und dritten Raumschrittweite insignifikante Werte; im Raumschritt 2 dagegen tritt eine deutlich signifikante positive Abweichung von dem bei stochastischer Unabhängigkeit für alle Raumschrittweiten zu erwartenden Sollwert Null auf, so daß diese entscheidend wichtige Annahme des Regressionsmodells verletzt ist. Obwohl also das Bestimmtheitsmaß mit $B = 0{,}97$ nominell sehr hoch liegt – auf das Problem der Überschätzung wurde bereits hingewiesen – wird offensichtlich ein (zumindest statistisch) bedeutsamer Teil der räumlichen Kovariation von N und A durch das Regressionsmodell nicht erfaßt.

Bei der Suche nach Ursachen und Lösungsmöglichkeiten sollte zunächst die räumliche Stichprobe auf Ausreißer untersucht werden; derartige Extremwerte der Regressionsvariablen können, falls sie sich nicht in die generelle Tendenz der übrigen Wertepaare einordnen, systematische Abweichungen der Meßwerte von den Regressionswerten und damit auch Autokorrelation der Residuen zur Folge haben. Solche extremen Ausreißer treten in diesem Beispiel jedoch nicht auf.

Daher ist zu vermuten, daß die lineare Form der Regressionsgleichung (11) nicht adäquat ist oder wichtige Prädiktorvariable noch nicht berücksichtigt sind.

Zur Überprüfung des ersten Teiles dieser Hypothese sind die Residual-Autokorrelationen für zwei alternative Regressionsansätze

$$A_i = a \cdot \exp(bN_i) + e_i \qquad (12)$$

$$A_i = a \cdot N_i^b + e_i \qquad (13)$$

---

[3] Die Anzahl der betrachteten Raumeinheiten wurde bewußt klein gehalten, um so ein Nachvollziehen der Regressionsrechnungen anhand der beigefügten Daten zu erleichtern.

[4] durch die Bildung zeitlicher Mittelwerte und Gebietsmittel für die Rasterflächen.

[5] Alle nachfolgenden Analysen wurden versuchsweise auch mit dem Nachbarschaftskriterium „gemeinsame Kanten" durchgeführt. Die danach errechneten Autokorrelationskoeffizienten unterscheiden sich im Einzelfall zwar wertemäßig und auch in ihrer Signifikanz, jedoch bleiben die getroffenen Aussagen über die Brauchbarkeit der verschiedenen Regressionsmodelle im Prinzip gültig.

[6] In der Literatur wird häufig auch der graphentheoretisch definierte Durchmesser einer räumlichen Konfiguration als maximale Raumschrittweite benutzt. Dabei erstreckt sich die Berechnung der Autokorrelationskoeffizienten für höhere Raumschrittweiten dann auf einen immer geringer werdenden Anteil an den n gegebenen Raumeinheiten, was aus statistischer Sicht (zu geringe Stichprobenumfänge und damit zu hohe Schätzfehlervarianz) bedenklich ist. Analoge Erwägungen führen zur Empfehlung einer maximalen Zeitverschiebung von etwa n/10 bei Autokorrelationsanalysen in Zeitreihen der Länge n (z. B. BOX/JENKINS, 1970, 33).

in der Tabelle 3 aufgeführt; mit ihren signifikant-positiven Autokorrelationskoeffizienten erweisen sich jedoch beide Modelle ebenfalls als weniger gut geeignet.

Tabelle 3: Räumliche Autokorrelationskoeffizienten R(k) für die Residuen alternativer Modellansätze (+ = signifikant auf dem 5%-Niveau; B = Bestimmtheitsmaß der Regression)

RSW k	$A = a\exp(bN)$	$A = aN^b$	$A = a + bX + cY$	$A = a + bX + cY + dN$
1	$0{,}35^+$	$0{,}33^+$	$0{,}82^+$	$-0{,}09$
2	0,05	0,10	$-0{,}55^+$	0,13
3	$-0{,}32$	$-0{,}34$	$-0{,}78^+$	$-0{,}23$
$B = R^2$	0,90	0,92	0,70	0,98

Untersucht man die beteiligten Regressionsvariablen A und N selbst auf räumliche Autokorrelation, so erhält man einen interessanten Hinweis (s. Tab. 2): Die hoch positiven Werte für $k = 1$ und $k = 2$ – sowie die hier nicht aufgeführten signifikant-negativen Koeffizienten bei $k = 4$ und 5 (vgl. Fußnote 6) – weisen deutlich auf einen räumlichen Trend hin, der auch in den Rohdaten der Tabelle 1 erkennbar ist. Offensichtlich ist also ein gewisser Teil der mit $r(N, A) = 0{,}99$ extrem hohen Korrelation zwischen den Regressionsvariablen A und N auf den in beiden Variablen existenten gleichsinnigen Trend zurückzuführen.[7] Die berechnete lineare Trendfläche

$$A_i = a + bX_i + cY_i + e_i, \quad (X_i, Y_i) = \text{Koordinaten der Rasterflächen-Mittelpunkte (s. Tab. 1)} \tag{14}$$

erreicht zwar nur ein Bestimmtheitsmaß von $B = 0{,}70$ und löst auch das Problem der Residual-Autokorrelation nicht (vgl. Tabelle 3), jedoch führt der um den Prädiktor N erweiterte Ansatz

$$A_i = a + bX_i + cY_i + dN_i + e_i \tag{15}$$

zum gewünschten Ergebnis. Mit $B = 0{,}98$ ist der statistisch erklärte Varianzanteil nominell kaum größer als beim einfachen Ansatz (11), aber die Residuen weisen jetzt die für Testzwecke notwendige und für die praktische Verwendbarkeit sinnvolle räumliche Unabhängigkeit auf.

Auf eine alternative Möglichkeit zur Berücksichtigung der in den Regressionsvariablen und Residuen des einfachen Ansatzes (11) nachweisbaren räumlichen Autokorrelation sei nur kurz hingewiesen.

Anstelle der in der Mehrfachregression (15) verwendeten absoluten Lagekordinaten X und Y, deren hydrologische Bedeutung bei der Schätzung von Abflußwerten in der Regel nur indirekter Art[8] ist, können auch die hydrologisch besser begründbaren Nachbarschaftseffekte in den Regressionsvariablen selbst in Form einer zusätzlichen autoregressiven Komponente einbezogen werden:

$$A_i = a + bN_i - cN_i(1) + e_i \tag{16}$$

oder

$$A_i = a + bA_i(1) + cN_i - dN_i(1) + e_i \quad \text{mit } N_i(1), A_i(1) \text{ gemäß Definition (7).} \tag{17}$$

Eine Beeinflussung des Abflusses A_i in der i-ten Raumeinheit durch den Niederschlag $N_i(1)$ der unmittelbar benachbarten Raumeinheiten kann hydrologisch etwa durch nicht nur lokal wirksame Grundwasserbewegungen erklärt werden; die endogene Komponente $A_i(1)$ kann z.B. Transportmechanismen im räumlich zusammenhängenden Gewässernetz widerspiegeln.

Die Kalibrierung derartiger Modelle (dazu: NIPPER/STREIT, 1978) mittels der üblichen Methode (3) der „Fehlerquadratsummen-Minimierung" liefert infolge der auftretenden Auto- und Interkorrelationen allerdings nicht mehr optimale Parameterschätzungen (CLIFF/ORD, 1973, 102–104; HAGGETT/CLIFF/FREY, 1977, 539f.).

Die folgende Überlegung zeigt, daß durch einen Modellansatz dieses Typs dem Problem autokorrelierter Regressionsresiduen im Prinzip begegnet werden kann. Betrachtet man die durch Anwendung des räumlichen linearen Filters

$$A_i^* = A_i - A_i(1), \quad N_i^* = N_i - N_i(1) \tag{18}$$

neu definierten Variablen A^* und N^*, so entspricht die lineare Einfachregression

$$A_i^* = a + cN_i^* + e_i \tag{19}$$

gerade einem Modell des Typs (17) mit spezialisierten Parametern $b = 1$ und $c = d$. Führt man – trotz der soeben geäußerten Bedenken – versuchsweise eine Parameterschätzung nach dem gängigen Kriterium (3) durch, so resultieren bei einem Bestimmtheitsmaß von $B = 0{,}72$ Residualwerte, die der Forderung nach räumlicher Unkorreliertheit (bzgl. der betrachteten Raumstruktur) voll genügen.

Tabelle 4: Räumliche Autokorrelationskoeffizienten R(k) für die Residuen des linearen Ansatzes $A^* = a + cN^* + e$ (s. Text) sowie für die Variablen A^* und N^* (+ = signifikant auf dem 5%-Niveau)

RSW k	Residuen e	Zielvariable A^*	Prädiktor N^*
1	$-0{,}23$	$0{,}45^+$	0,10
2	0,03	$-0{,}21$	0,10
3	0,01	$-0{,}27$	$-0{,}36$

Die in der Tabelle 4 ebenfalls aufgeführten räumlichen Autokorrelationskoeffizienten für die transformierten Variablen A^* und N^* lassen zudem die Wirksamkeit der angewendeten Methode (18) zum Herausfiltern eines linearen räumlichen Trends (bzw. allgemein einer hoch positiven räumlichen Autokorrelation 1. Ordnung) erkennen.

4. *Nachwort*

Die mit der räumlichen wie auch zeitlichen Autokorrelation von Variablen verbundenen Probleme beschränken sich keinesfalls auf die Regressionsanalyse. Sie treten im Prinzip bei all den statistischen Methoden auf, die stochastische Unabhängigkeit der betrachteten Zufallsvariablen voraussetzen

[7] Der gemeinsame Varianzanteil zwischen den vom regionalen linearen Trend durch Anwendung des räumlichen Filters (18) bereinigten Variablen A^* und N^* sinkt demgemäß auf $B = 0{,}72$. Die Regressionsgleichung (11) zeigt also auch unter kausalanalytischem Aspekt infolge der Korrelation über den gemeinsamen Trend eine deutliche Überschätzung des hydrologisch bedeutsamen Zusammenhanges zwischen N und A.

[8] In diesem Beispiel wirkt als Hintergrundvariable die von NW nach SE (bzgl. der Rasterflächen in Tabelle 1) zunehmende Höhenlage des Reliefs.

zen. HAGGETT/CLIFF/FREY (1977, 329–336 und 374–377) weisen in diesem Zusammenhang auf die häufig bedenkenlose Anwendung z. B. des t- und F-Tests etwa bei einfachen Mittelwertvergleichen sowie Varianzanalysen, des $\varkappa^2$-Anpassungstests sowie der Faktorenanalyse hin.

Literatur

BAHRENBERG, G., GIESE, E.: Statistische Methoden und ihre Anwendung in der Geographie. Teubner Studienbücher Geographie, Stuttgart (1975), 308 S.

BENNETT, R.J.: Spatial Time Series. Analysis-Forecasting-Control. Pion, London (1979), 674 S.

BOX, G.E.P., JENKINS, G.M.: Time Series Analysis. Forecasting and Control. Holden Day, San Francisco (1970), 553 S.

CLIFF, A. D., HAGGETT, P., ORD, J. K., BASSETT, K., DAVIS, R.: Elements of Spatial Structure. A Quantitative Approach. Cambridge Univ. Press, Cambridge (1975), 258 S.

CLIFF, A.D., ORD, J.K.: Testing for Spatial Autocorrelation among Regression Residuals. Geographical Analysis, Vol. 4, (1972), S. 267ff.

–, : Spatial Autocorrelation. Monographs in Spatial and Environmental Systems Analysis, No. 5, Pion, London (1973), 178 S.

DURBIN, J.: Testing for Serial Correlation in Least Squares Regression when some of the Regressions are Lagged dependent Variables. Econometrica Vol. 38 (1970), S. 410–421.

DURBIN, J., WATSON, G. S.: Testing for Serial Correlation in Least Squares Regression. Biometrika, Vol. 38 (1951), S. 159–177.

GAENSSLEN, H., SCHUBÖ, W.: Einfache und komplexe statistische Analyse. UTB, Reinhardt, München (1973), 326 S.

HAGGETT, P., CLIFF, A.D., FREY, A.: Locational Analysis in Human Geography. Vol. II = Locational Methods. Arnold, London (1977), 605 S.

MARTIN, R. L.: On Spatial Dependence, Bias and the Use of First Spatial Differences in Regression Analysis. AREA, Vol. 6 (1974), S. 185–194.

NIPPER, J., STREIT, U.: Zum Problem der räumlichen Erhaltensneigung in räumlichen Strukturen und raumvarianten Prozessen. Geographische Zeitschrift, Vol. 65/4 (1977), S. 241–263.

NIPPER, J., STREIT, U.: Modellkonzepte zur Analyse, Simulation und Prognose raum-zeit-varianter stochastischer Prozesse. In: BAHRENBERG, G., TAUBMANN, W. (Hrsg.): Anwendung quantitativer Modelle in der Geographie und Raumplanung. = Bremer Beiträge zur Geographie und Raumplanung. H. 1 (1978), S. 1–17.

STREIT, U.: Raumvariante Erweiterung von Zeitreihenmodellen. Ein Konzept zur Synthetisierung monatlicher Abflußdaten von Fließgewässern unter Berücksichtigung von Erfordernissen der wasserwirtschaftlichen Planung. = Gießener Geographische Schriften H 46 (1979), 105 S.

Wasserwirtschaftlicher Rahmenplan Fulda, herausgeg. vom Hessischen Ministerium für Ernährung, Landwirtschaft und Forsten, Wiesbaden (1964).

BUCHBESPRECHUNGEN

Atlas der Donauländer. Herausgeber: Österreichisches Ost- und Südosteuropa-Institut Wien. Redakteur: JOSEF BREU. Kommissionsverlag Franz Deuticke Wien 1970ff. Bisher erschienen 5 Lieferungen mit 25 Karten. Je Karte DM 17,20

Mit 5 Lieferungen liegt ein bedeutender Teil des von J. BREU redaktionell betreuten großen Atlas-Werkes vor, das zu den eindrucksvollsten Leistungen jüngerer Atlaskartographie gezählt werden muß. Die Überschreitung nationaler Grenzen – in diesem Falle von Ländern ganz unterschiedlicher Gesellschafts- und Wirtschaftsordnung – hat die Herausgeber vor Aufgaben gestellt, deren Bewältigung von vornherein nur durch eine engere internationale Zusammenarbeit denkbar erschien. Grundlegende Schwierigkeiten, für die in allen Fällen nach angemessenen Lösungen gesucht wurde, ergaben sich nicht nur aus dem Fehlen einer einheitlichen und aktuellen topographischen Basis, sondern ebenso aus der Notwendigkeit, sich auf oft sehr heterogene Daten zu stützen, aus der von Land zu Land andersartigen Verwaltungsgliederung oder auch aus der Aufgabe, einen geeigneten Weg für die Schreibweise von Ortsnamen und anderen topographischen Bezeichnungen zu finden. Wenn die kartographisch durchweg überzeugend gelöste Darstellung einzelner Sachverhalte nicht in allen Fällen dem Informationsbedürfnis voll gerecht werden kann, dann läßt sich daraus wegen der vorstehend angedeuteten Probleme gewiß kein Vorwurf erheben. Der vorliegende Kartenbestand macht den Atlas zu einer für die künftige landeskundliche Erforschung Südosteuropas wesentlichen und entscheidend weiterführenden Grundlage, wobei er vorhandene Nationalatlanten weder ersetzen kann noch ersetzen soll.

Die Kartenausschnitte umfassen neben den Staatsgebieten der Tschechoslowakei, Ungarns, Rumäniens, Jugoslawiens, Albaniens und Bulgariens Teile der angrenzenden Nachbarländer, u. a. Teile der Sowjetunion, Polens, Österreichs und Griechenlands. Für die Hauptkarten wurde der Maßstab 1:2 Mill. gewählt, verschiedene Themen sind in kleinerem oder größerem Maßstab dargestellt. Der jeweils ausführliche Erläuterungstext in deutscher, englischer, französischer und russischer Sprache gibt Auskunft über die verwendeten Quellen, die der Darstellungsweise zugrunde liegenden Überlegungen und enthält eine mehr oder weniger umfassende Interpretation des Karteninhalts. Es ist verständlicherweise nicht möglich, auf alle der inzwischen vorliegenden 25 Karten einzugehen, vielmehr kann nur auf einige von ihnen als besonders bedeutsame Bestandteile des Gesamtwerkes hingewiesen werden.

Neben zwei verschiedenen Fassungen der topographischen Karte, einer hypsographischen Karte und einer Reliefkarte enthält der erste Teil eine von P. BECK-MANNAGETTA und W. MEDWENITSCH bearbeitete geologische Übersichtskarte, deren Grundkonzept die Betonung tektonischer Baueinheiten ist. Besondere Aufmerksamkeit gebührt der von M. PÉCSI erstellten geomorphologischen Karte. In ihr werden Relieftypen und Einzelformen unterschieden und zugleich Angaben über das Alter der Oberflächenentstehung und -formen gemacht. Trotz der Fülle der in dieser Karte enthaltenen Details bietet sie eine klare Gesamtübersicht, wobei gelungene Farbgebung und zweckmäßige Signaturenwahl besonders hervorzuheben sind. Neben den Klimakarten mit ihren am Rande zu findenden Stationsdiagrammen muß aus dem ersten Teil noch die von H. NIKLFELD entworfene Karte der natürlichen Vegetation – vom Autor als „theoretische natürliche Vegetation" bezeichnet – Erwähnung finden. Auch sie ist weit mehr als eine Zusammenfassung bislang aus dem bearbeiteten Gebiet vorliegender Teilkarten unterschiedlichen Maßstabes und sehr verschiedenartiger Konzeption.