

F. KÖRNER (Dresden) und J. LIEBE (Vetschau)

ZUR ERZEUGUNG VON DISTANZMATRIZEN BEIM RUNDREISEPROBLEM

1. Notation. Die Lösungsalgorithmen für das Rundreiseproblem sind in den letzten Jahren bedeutend verbessert worden. Das führt dazu, daß man in der Lage ist, diese Probleme bis zu Dimensionen weit über $n = 100$ zu lösen. Um diese Lösungsalgorithmen umfangreich testen zu können, werden Distanzmatrizen benötigt. Bisher gelangte man auf den folgenden zwei Wegen zu den benötigten Daten:

(a) Aus vorhandenen Unterlagen, wie zum Beispiel Landkarten, Atlanten und Entfernungstabellen, wurden die Daten ermittelt und in den Computer eingegeben. Der damit verbundene Arbeitsaufwand (zum Beispiel das Ablochen von 10000 Daten bei einer Dimension von $n = 100$) ist sehr hoch.

(b) Die Distanzmatrix wird mit im Computer erzeugten Zufallszahlen belegt. Damit ist man in der Lage, jede Matrix beliebig hoher Dimension schnell (das heißt ohne großen manuellen Aufwand) zu belegen.

Dabei ist jedoch zu beachten, daß die Daten, die man aus der Praxis gewinnt, einem realen, zufälligen Sachverhalt entstammen. Wenn man also die Algorithmen zum Beispiel hinsichtlich des Erwartungswertes ihrer Rechenzeit bezüglich praktischer Probleme testen will, dann muß man diesen realen Sachverhalt mittels Zufallszahlen simulieren.

THESE 1. Wird die Distanzmatrix beim Rundreiseproblem mit Zufallszahlen belegt, die nach einer bestimmten Methode erzeugt wurden, die jedoch nicht den realen Sachverhalt simulieren, so erhält man zwangsläufig Rechenzeiten, die nicht repräsentativ sind für praktische Probleme.

THESE 2. Es ist nicht auszuschließen, daß man zu einem gegebenen Verfahren (Algorithmus zur Lösung des Rundreiseproblems) Testbeispiele (Methode zur zufälligen Belegung der Distanzmatrix) finden kann, die die Güte des Verfahrens nicht typisch charakterisieren.

Aus diesen Gründen sollten in der Literatur nicht nur die verwendeten Verteilungsfunktionen, sondern auch die zur Realisierung verwendeten Zufallszahlengeneratoren, Startwerte etc. angegeben werden, damit man in der Lage ist, sich die gleichen Beispiele zu erzeugen, und man damit die Möglichkeit hat, die Güte des eigenen Algorithmus beurteilen zu können.

Um diese aufgetretenen Mängel zu beseitigen, wird nun versucht, ein Programm zur Simulation der Distanzmatrix zu entwickeln, welches die Realität möglichst gut approximiert und welches den Vorteil hat, daß bei gleichen Startwerten (für die Zufallszahlengeneratoren) gleiche Distanzmatrizen erzeugt werden.

2. Zur Modellbildung. Die Modellierung geht von der Tatsache aus, daß die Erde in Länder (mit Ausnahme von Antarktika) aufgeteilt ist und diese wiederum in einzelne territoriale Regionen untergliedert sind. Unter einer territorialen Region ist ein Verwaltungsdistrikt, zum Beispiel Bezirk, Bundesland, Unionsrepublik, zu verstehen. Diese Fakten bilden die Grundlage für die im Modell getroffenen Annahmen.

Um eine möglichst gute Approximation der Realität zu erhalten, wird folgendes Modell vorgeschlagen:

(I) Die Zentren der territorialen Regionen eines Landes sind in der Ebene gleichverteilt.

(II) Jeder territorialen Region sei nach einer bestimmten Verteilung eine gewisse Anzahl von Städten zugeordnet.

(III) Innerhalb einer territorialen Region werden die Städte als normalverteilt bezüglich des Zentrums angesehen.

(IV) Nach einer bestimmten Verteilung treten in den einzelnen Verkehrsrichtungen Störungen (zum Beispiel, Straße verläuft nicht geradlinig, Umleitungen usw.) auf. Das heißt, es ist ein Faktor zu bestimmen, der dieses Problem „krummer Weg“ zum Ausdruck bringt.

(V) Jede Stadt ist nicht mit jeder anderen verbunden. Da nur direkte Verbindungen in die Distanzmatrix eingehen, ist es erforderlich, eine Strategie (Welche Verbindungen sind zu streichen und wieviel?) zu entwickeln, auf Grund dieser dann der Streichungsprozeß durchgeführt wird.

Völlige Freiheit herrscht noch in der Wahl der entsprechenden Verteilungsfunktionen in (II), (IV) und (V). In der weiteren Arbeit werden für diese Punkte Verteilungsfunktionen vorgeschlagen und Schätzungen für die dazugehörigen Parameter angegeben.

3. Testverfahren auf Vorliegen einer bestimmten Verteilung. Um einen realen Sachverhalt, der durch eine zufällige Maßzahl (Zufallsgröße) erfaßt wird, simulieren zu können, ist es notwendig, die Verteilungsfunk-

tion, die dieser Zufallsgröße zugrunde liegt, zu kennen. Im allgemeinen ist jedoch nur statistisches Material, wie zum Beispiel die konkrete Existenz der Städte auf unserem Planeten vorhanden. Um aus diesem statistischen Material die Verteilungsfunktion zu bekommen, kann man folgende Methoden anwenden:

(a) Es wird eine Verteilungsfunktion, bei der Parameter (Erwartungswert, Streuung, etc.) frei wählbar sind, vorgegeben. Diese Parameter werden mittels der Maximum-Likelihood-Methode bestimmt.

(b) Man gibt sich eine dem Sachverhalt angepaßte Verteilung vor und prüft mittels des χ^2 -Anpassungstestes die Hypothese: Dem statistischen Material liegt die vorgegebene Verteilungsfunktion zugrunde.

4. Zur Approximation der Verteilungsfunktionen. Ausgangspunkt für das in dieser Arbeit vorgestellte Modell ist die DDR. Unter den territorialen Regionen sind in diesem Fall die Bezirke der DDR zu verstehen. Zentren der territorialen Regionen sind die Bezirksstädte.

Alle notwendigen statistischen Untersuchungen wurden anhand einer Stichprobe durchgeführt, die sich folgendermaßen zusammensetzt:

(a) Betrachtet wurden die 14 Bezirksstädte der DDR (Berlin wurde von den Untersuchungen ausgeschlossen).

(b) Die Stichprobe umfaßt alle Städte der DDR mit mehr als 5000 Einwohner. Die Anzahl dieser Städte beträgt 449.

(c) Aus dem Spektrum der vorhandenen Verkehrswege wurde das Autobahn- und Fernstraßennetz der DDR ausgewählt.

Der Arbeit liegen die Atlanten [1] und [2] zugrunde. An dieser Stelle sei darauf hingewiesen, daß die Ansätze von Verteilungsfunktionen und die Schätzwerte für die dazugehörigen Parameter nur für DDR-spezifische Untersuchungen und Aussagen gültig sind. Zur Simulation anderer Länder müssen analoge Untersuchungen durchgeführt werden.

Es werden nun für die Punkte (I) bis (V) Verteilungsfunktionen vorgeschlagen und Schätzwerte für die entsprechenden Parameter angegeben:

(I) In (I) wurde bereits festgelegt, daß die Zentren der territorialen Regionen (die Bezirksstädte) in der Ebene gleichverteilt sind. Für die DDR ergibt sich ein Rechteck von etwa 310 km Breite und 420 km Länge in dem die Bezirksstädte liegen.

(II) Die 14 Bezirke wurden entsprechend der Anzahl a ihrer Städte den Klassen I_0, I_1, I_2 und I_3 mit der Klasseneinteilung

$$I_0: 0 \leq a < 30,$$

$$I_1: 30 \leq a < 50,$$

$$I_2: 50 \leq a < 70,$$

$$I_3: 70 \leq a < 80$$

zugeordnet. Bild 1 stellt das entsprechende Histogramm dar.

Im Bild 1 wurde auf der Abszissenachse die Klasseneinteilung und auf der Ordinatenachse die absolute Häufigkeit der Bezirke mit I_j Städten aufgetragen. In Auswertung dieser Darstellung erscheint Poisson-Verteilung als eine dem Sachverhalt angepaßte Verteilung. Für den Parameter der Poisson-Verteilung λ wurde mit Hilfe der Maximum-Likelihood-Methode der Schätzwert $\hat{\lambda} = 0,643$ ermittelt.

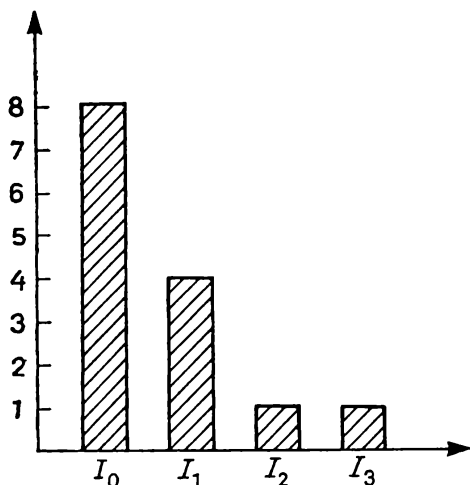


Bild 1

(III) In (III) wurde die Annahme, daß die Städte innerhalb der Bezirke normalverteilt sind, getroffen. Dies ist darin begründet, daß man den größten Teil der Städte als in der „näheren“ Umgebung der Bezirksstädte gelegen ansehen kann. Eine Ausnahme stellt der Bezirk Rostock dar.

Zur Schätzung der 2-dimensionalen Normalverteilung $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ wurde in jedem Bezirk durch die Bezirksstadt ein Koordinatensystem gelegt und mit Hilfe der Maximum-Likelihood-Methode σ_1 und σ_2 bestimmt. Unter Berücksichtigung, daß der Bezirk Rostock eine Ausnahme darstellt, wurden für die Werte σ_1 und σ_2 die folgenden Intervalle ermittelt:

$$25,50 \leq \sigma_1 \leq 51,39 \quad \text{und} \quad 17,28 \leq \sigma_2 \leq 44,46.$$

Es wird angenommen, daß diese Werte auf den Intervallen gleichverteilt sind.

(IV) Folgende Stichprobe liegt den statistischen Untersuchungen zu dieser Problematik zugrunde:

Betrachtet wurden 36 Städte der DDR. Die Entfernungen dieser Städte voneinander wurden einer Tabelle aus dem Reiseatlas der DDR [2] entnommen.

Zusätzlich wurden für die 360 Verbindungen die Euklidischen Abstände (Luftlinie) durch Messungen anhand der Karte der DDR [1] bestimmt.

Im Ergebnis des Vergleiches des Euklidischen Abstandes l^e mit dem „wahren“ Abstand l^w für jede Verbindung wurde folgender Ansatz gewählt: $l^w = l^e(1 + \varepsilon)$. Es läßt sich $\varepsilon \geq 0$ als multiplikativer Faktor interpretieren, der mit dem Euklidischen Abstand zu verknüpfen ist, um einen annähernd realen Abstand zu erhalten. Es erfolgte folgende Klasseneinteilung:

$$I_0: 0,000 \leq \varepsilon < 0,050,$$

$$I_1: 0,050 \leq \varepsilon < 0,100,$$

$$I_2: 0,100 \leq \varepsilon < 0,150,$$

$$I_3: 0,150 \leq \varepsilon < 0,200,$$

$$I_4: 0,200 \leq \varepsilon < 0,250,$$

$$I_5: 0,250 \leq \varepsilon < 0,300,$$

$$I_6: 0,300 \leq \varepsilon < 0,375,$$

$$I_7: 0,375 \leq \varepsilon < 0,500,$$

$$I_8: 0,500 \leq \varepsilon < 0,650,$$

$$I_9: 0,650 \leq \varepsilon < 0,850,$$

$$I_{10}: 0,850 \leq \varepsilon < 1,000.$$

Das Bild 2 stellt das entsprechende Histogramm dar. In Auswertung des Histogramms wurde diesem Sachverhalt eine Binomialverteilung zugeordnet. Mit der Maximum-Likelihood-Methode wurde für den Parameter p der Binomialverteilung der Schätzwert $\hat{p} = 0,4022$ ermittelt.

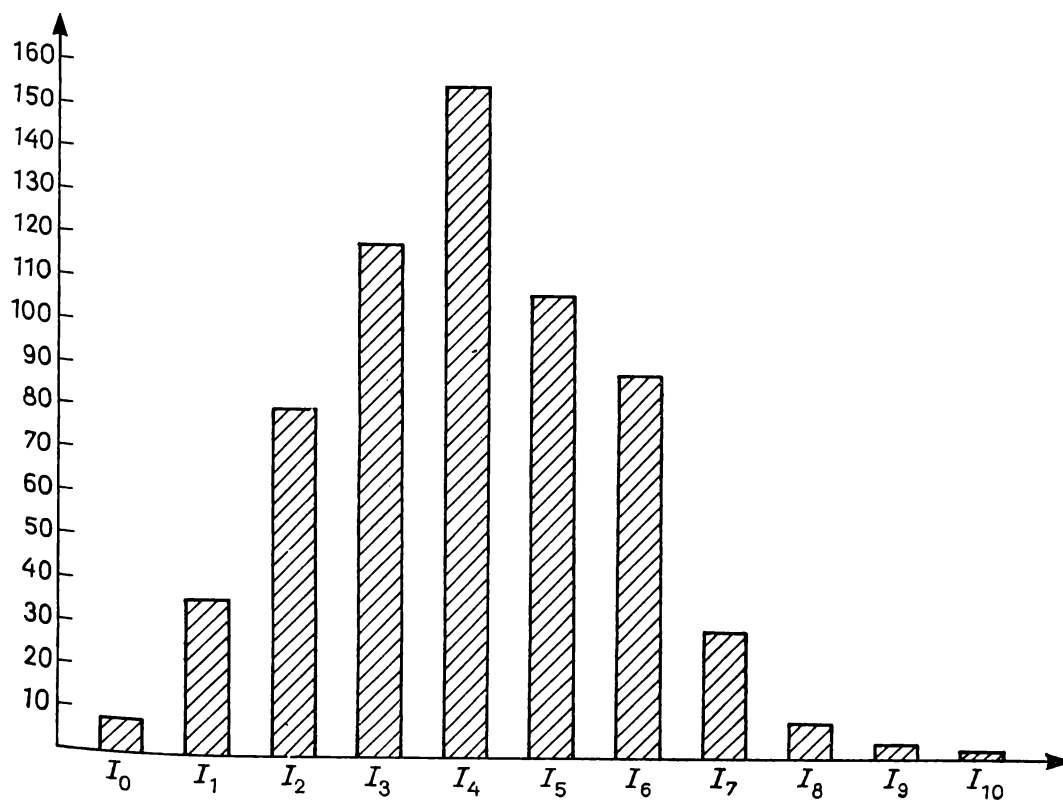


Bild 2

(V) Betrachtet wird das Problem der Ermittlung der direkten Verbindungen eines Systems von n Städten. Es wurde die gleiche Stichprobe wie in (IV) verwendet. Anhand des Reiseatlases der DDR [2] wurden außerdem die direkten Verbindungen bezüglich Autobahn- und Fernverkehrsstraßennetz dieser 36 Städte ermittelt. Die Anzahl der zu streichenden Verbindungen lag zwischen 70 und 80 %.

Anschließend erfolgte eine Untersuchung des Zusammenhanges zwischen der Länge einer Verbindung und ihrer Wahl als direkte Verbindung. Dazu wurden alle 35 Verbindungen einer Stadt der Größe nach geordnet. Die Häufigkeit des Streichens der i -ten Verbindung, $i = 1, \dots, 35$, bei alle 36 Städten ist in Bild 3 dargestellt.

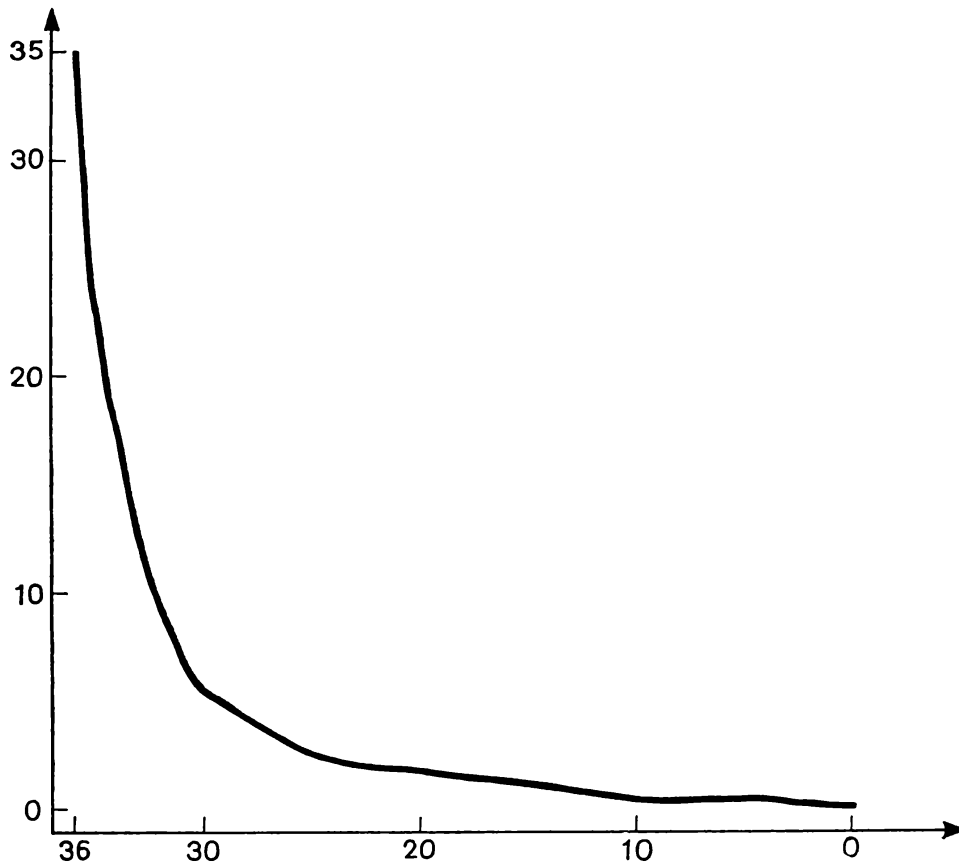


Bild 3

Im Ergebnis der Auswertung dieses Bildes wird eine Exponentialverteilung zugrunde gelegt. Der Parameter der Exponentialverteilung wird so gewählt, daß die kleinste (n minus erste) Verbindung mit der Wahrscheinlichkeit p ($0 < p < 1$) gestrichen wird. Es ergibt sich damit für den Parameter der Exponentialverteilung λ folgende Berechnungsvorschrift:

$$\lambda = -\frac{1}{n} \ln p.$$

Mit Hilfe der Maximum-Likelihood-Methode wurde für p der Schätzwert $\hat{p} = 0,025$ ermittelt.

5. Zur rechentechnischen Realisierung. Werden für Testrechnungen Distanzmatrizen zufällig erzeugt, so ist man im allgemeinen nicht in der Lage, sich die gleichen Beispiele wieder zu erzeugen. Einerseits liegt es daran, daß die verwendeten Zufallszahlengeneratoren und entsprechende Startwerte nicht angegeben werden, andererseits arbeiten die meisten Zufallszahlengeneratoren mit der real-Arithmetik; da unterschiedliche Computer unterschiedliche Wortlängen, unterschiedliche Rundungsregeln usw. besitzen, ist es selbst bei Kenntnis des verwendeten Zufallszahlengenerators und dessen Startwerte nicht möglich, sich die gleiche Folge von Zufallszahlen und damit die gleiche Distanzmatrix auf einem anderen Computer zu erzeugen. Die Computer besitzen aber die gleiche integer-Arithmetik. Es ist daher von Vorteil, die integer-Arithmetik zur Erzeugung einer Folge von Zufallszahlen zu benutzen.

Von grundlegender Bedeutung sind auf dem Intervall $[0, 1)$ gleichverteilte Zufallszahlen. Zufallszahlen, die anderen Verteilungen genügen sollen, werden durch eine Transformation von auf dem Intervall $[0, 1)$ gleichverteilten Zufallszahlen gewonnen. Der interessierte Leser vergleiche dazu die Monographie von Zielinski [3].

Es wird nun ein Generator angegeben, der die integer-Arithmetik verwendet und der eine Folge (x_n) von auf der Menge $\{1, \dots, M-1\}$ gleichverteilter Zufallszahlen liefert. Ein solcher Generator ist zum Beispiel durch

$$x_{n+1} := ax_n + b \pmod{M}$$

gegeben. Für diesen Generator gilt folgende Aussage aus [3]:

Dieser Generator mit der Konstanten $M = 2^m$ ist ein Generator mit voller Periode $P = M$, wenn b ungerade und $a = 1 \pmod{4}$ ist.

Für die Werte $a = 5$ und $b = 1$, die Größen $M = 2^{16}$ und $M = 2^{32}$, und die Startwerte $x_0 = 1$ und $x_0 = 10$ wurden jeweils n^2 ($n = 10(10)500$) Zufallszahlen erzeugt und mittels des Anpassungstestes die Hypothese überprüft; ob die Zufallszahlenfolgen der Länge n^2 gleichmäßig auf $[0, M)$ verteilt sind. Es erfolgte zum Signifikanzniveau $\alpha = 0,01$ bei allen Tests keine Ablehnung.

Diese so erzeugte „grund“ Zufallszahlenfolge wird nun auf die im Abschnitt 4 ermittelten Verteilungen transformiert. Diese Transformationen werden mit der real-Arithmetik durchgeführt. Die dementsprechend auftretenden computerbedingten Abweichungen sind, da nur eine geringe Anzahl von real-Operationen nötig sind und die „grund“ Zufallszahlenfolge auf jedem Computer gleich ist, minimal.

Die nun angegebenen Transformationen wurden aus [3] entnommen.

(a) *Erzeugung von auf dem Intervall $[0, 1)$ gleichverteilter Zufallszahlen.* Ist (x_n) eine Folge von auf dem Intervall $[0, M)$ gleichverteilter Zufalls-

zahlen, so ist die Folge (y_n) mit $y_i := x_i/M$ auf dem Intervall $[0, 1)$ gleichverteilt.

(b) *Erzeugung einer Poisson-verteilten Zufallszahlenfolge.* Es sei λ der Parameter der Poisson-Verteilung, und es sei (x_n) eine Folge von auf dem Intervall $[0, 1)$ gleichverteilter Zufallszahlen. Damit ergibt sich folgender Algorithmus:

Man berechne die Produkte $x_1, x_1 x_2, x_1 x_2 x_3, \dots$ so lange, bis erstmals $x_1 x_2 \dots x_{i+1} > e^{-\lambda}$ gilt. Der Wert der Poisson-verteilten Zufallsgröße ist dann die Zahl i .

(c) *Erzeugung von normalverteilten Zufallszahlen.* Es sei (x_n) eine Folge von auf dem Intervall $[0, 1)$ gleichverteilter Zufallszahlen. Die Folge (y_n) mit

$$y_i := x_{12i+1} + x_{12i+2} + \dots + x_{12i+12} - 6$$

ist $N(0, 1)$ verteilt. Hat die Folge (y_n) die Verteilung $N(0, 1)$, so hat die Folge (z_n) mit $z_i := \sigma y_i + \mu$ die Verteilung $N(\mu, \sigma^2)$.

(d) *Erzeugung von binomialverteilten Zufallszahlen.* Die Verteilungsfunktion der Binomialverteilung lautet

$$F(k+1) = \sum_{i=0}^k \binom{n}{i} p^i (1-p)^{n-i} \quad \text{mit } k = 0, \dots, n-1.$$

Aus der Klasseneinteilung folgt $n = 10$. Zur Erzeugung der binomialverteilten Folge von Zufallszahlen wird das Intervall $[0, 1)$ nach der Berechnung der Verteilungsfunktion F in folgender Weise unterteilt:

$$J_k := [F(k), F(k+1)).$$

Jedem Teilintervall J_k wird ein Faktor R_k (Intervallmitte von I_{k+1} und I_k aus der Klasseneinteilung) zugeordnet. Fällt die gleichmäßig auf $[0, 1)$ erzeugte Zufallszahl in das Intervall J_k , so wird $\varepsilon = R_k$ gesetzt.

(e) *Erzeugung von exponentialverteilten Zufallszahlen.* Es sei λ der Parameter der Exponentialverteilung, und es sei (x_n) eine Folge von auf dem Intervall $[0, 1)$ gleichverteilter Zufallszahlen. Dann ist die Folge (y_n) mit

$$y_i := -\frac{1}{\lambda} \ln x_i$$

exponentialverteilt.

6. Schlußbemerkungen. Das hier vorgestellte Modell zur Simulation von Distanzmatrizen für das Rundreiseproblem verfolgte das Ziel, mit möglichst geringem Aufwand Distanzmatrizen beliebig großer Dimension zu generieren, die die Realität möglichst genau widerspiegeln. Zu diesem

Simulationsmodell wurde ein FORTRAN-Programm erstellt. Eine Auswertung von umfangreichen Testrechnungen zeigte, daß die generierten Matrizen die Realität hinreichend gut widerspiegeln.

Es wurde angestrebt, daß die zur Generation verwendeten Programme so beschaffen sind, daß sie auf unterschiedlichen Computern mit den gleichen Startwerten gleiche Matrizen liefern. Im Punkt 5 wurde eine Möglichkeit aufgezeigt, die Abweichungen (die von unterschiedlichen Rundungsregeln usw. herrühren) der auf unterschiedlichen Computern erzeugten Matrizen gering zu halten. Wünschenswert wäre es, eine Methode anzugeben, welche nur die integer-Arithmetik zur Generierung verwendet und damit garantiert, daß auf unterschiedlichen Computern exakt die gleichen Matrizen generiert werden.

Ein FORTRAN-Programm zur Generierung von Distanzmatrizen kann von den Autoren erhalten werden.

Literaturnachweis

- [1] *Großer Weltatlas*, VEB Hermann Haak, Geographisch-Kartographische Anstalt. Gotha/Leipzig, 1965.
- [2] *Reiseatlas der DDR*, VEB Landkartenverlag Berlin, 1972.
- [3] R. Zielinski, *Erzeugung von Zufallszahlen*, VEB Fachbuchverlag Leipzig, 1978

SEKTION MATHEMATIK
TECHNISCHE UNIVERSITÄT DRESDEN
DDR-8027 DRESDEN

ORGREB
INSTITUT FÜR KRAFTWERKE
DDR-7544 VETSCHAU

Eingegangen am 28. 10. 1981
