

W. Förstner

STATISTISCHE GRUNDLAGEN FÜR DIE ZUVERLÄSSIGKEIT VON AUSGLEICHUNGSERGEBNISSEN

1. Einleitung

Im vorangegangenen Vortrag haben wir den Zusammenhang zwischen Beobachtungsfehlern und dem Ergebnis der Ausgleichung kennengelernt. Die Aufteilung der Redundanz auf die Beobachtungen ist der Schlüssel für eine Beurteilung der Geometrie. Die Anteile r_i der einzelnen Beobachtungen an der Redundanz hängen wesentlich nur vom Meßaufbau und von der Genauigkeit der Beobachtungen ab und lassen sich daher vor der eigentlichen Messung bestimmen. Sie zeigen an welchen Stellen ein Netz oder ein Block grobe Fehler absorbiert, so daß sie sich nicht oder nur sehr schwach in den Verbesserungen nach der Ausgleichung niederschlagen.

Obwohl die Redundanzanteile im Grunde ausreichen, um die Kontrollmöglichkeiten in einem Netz oder Block zu beurteilen, erfassen sie nicht den Einfluß des jeweils verwendeten Testverfahrens auf die Kontrollierbarkeit der Beobachtungen, dies ist die Sicherheit, mit der sich grobe Fehler finden lassen oder eine etwaige untere Grenze für grobe Fehler, unter der sie praktisch nicht aufdeckbar sind.

Das verwendete Prüfverfahren hängt nun eng mit dem der Auswertung zugrunde gelegten Modell von der Wirklichkeit zusammen. Es sollte zudem auf die erwarteten Fehler, das sind die Abweichungen zwischen Modell und Wirklichkeit, zugeschnitten sein.

Ich möchte daher zunächst auf einige Annahmen eingehen, die man bei der Messung, der Auswertung und der Beurteilung von Beobachtungswerten trifft, um damit die statistische Beobachtungsweise, die - bewußt oder unbewußt - bei jeder Prüfung im Hintergrund steht, zu motivieren und sie an einigen Grundbegriffen zu erläutern. Sie präzisieren einerseits den bekannten Ausgleichungsansatz nach vermittelnden Beobachtungen, der den meisten Rechenprogrammen zugrunde liegt, und enthalten andererseits den Begriff des Hypothesentests, der das Prinzip der Fehlersuche, wie es in der Praxis üblich ist, genauer beschreibt. Mit Hilfe der Macht des Tests läßt sich die Frage der Kontrollierbarkeit der Beobachtungen beantworten, die auf den von BAARDA eingeführten Begriff der inneren Zuverlässigkeit führt.

2. Grundannahmen für die Bewertung von Beobachtungen

2.1 Die Auswertung von Messungen setzt eine Vielzahl von Annahmen voraus. Sie lassen sich in drei Teilbereiche gliedern:

- Annahmen über die M e s s u n g. Die Zeigerablesungen oder die digitale Anzeige von Meßwerten wird mit einer physikalischen Bedeutung belegt, etwa einer Temperatur oder einer Strecke. Die den Meßablauf bestimmenden Faktoren (Luftdruck, Besonderheiten bei der Messung) werden soweit wie möglich oder sinnvoll registriert, in der Annahme, daß sie für die Weiterverarbeitung von

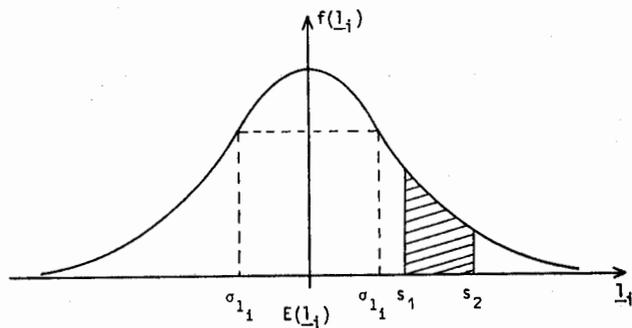
Belang sind. Die meisten derartigen Eigenschaften des Umfeldes um die Messung bleiben allerdings unbekannt. Die Auswahl der registrierten Faktoren ist eng verknüpft mit den

- Annahmen über das die Wirklichkeit beschreibende Modell, das der Weiterverarbeitung zugrunde gelegt wird, sei es physikalisch oder geometrisch. Dazu gehören etwa die Zentralperspektive oder die Normatmosphäre. Da es sich dabei immer um Näherungen handelt, müssen wir
- Annahmen über Meß- oder Modellfehler präzisieren, um nach der Auswertung Widersprüche zwischen den Meßwerten analysieren zu können.

In allen drei Fällen ist es aus operationellen Gründen notwendig stark zu vereinfachen, wodurch die Auswertung überhaupt erst durchführbar wird. Die Beurteilung ist dafür umso schwieriger. Auch für die Prüfung auf grobe Fehler müssen wir eine Reihe von zum Teil unrealistischen Annahmen treffen.

2.2 Zunächst nehmen wir an, daß die Beobachtungen \underline{l}_i normal verteilte Zufallsvariable (auch stochastische Variable) sind. Das bedeutet, daß bei Wiederholungen die Meßwerte um einen Mittelwert streuen können, die Zufallsvariable \underline{l}_i also - etwa vom Zeitpunkt der Messung abhängig - verschiedene Werte annehmen kann. Zur Unterscheidung von festen (deterministischen) Größen unterstreichen wir stochastische Variable. Die Zufallsvariable wird beschrieben durch die Dichtefunktion $f(\underline{l}_i)$ (vgl. Fig. 1), bei normal verteilten Größen eine Gauß'sche Glockenkurve.

Fig. 1
 Dichtefunktion einer Zufallsvariablen



An ihr kann man ablesen, mit welcher Wahrscheinlichkeit P (probability) der Beobachtungswert \underline{l}_i zwischen S_1 und S_2 liegt:

$$P(S_1 \leq \underline{l}_i \leq S_2) = \int_{S_1}^{S_2} f(x) dx \quad (1)$$

Ohne auf eine genaue Definition im Sinne der Statistik einzugehen, lassen sich die wichtigsten Eigenschaften einer Zufallsvariablen mit Hilfe der Figur 1 erläutern.

Die Gesamtfläche unter der Kurve ist 1, denn irgendein Wert wird mit $P = 1 = 100\%$ -iger Sicherheit angenommen. Der Mittelwert wird auch mit Erwartungswert $E(\underline{l}_i)$ bezeichnet. Falls f symmetrisch ist, treten positive und negative Abweichungen gleich wahrscheinlich auf. Die Wahrscheinlichkeit, daß ein ganz bestimmter Wert, z.B. $E(\underline{l}_i)$, gemessen wird, ist 0. Daher auch die berechtigte Skepsis gegen identische Meßwerte!

Die Streuung um den Erwartungs- oder Mittelwert ist die Standardabweichung σ_{l_i} oder der mittlere Fehler der Messung. Wegen des Fehlerfortpflanzungsgesetzes wird oft auch mit dem Quadrat der Standardabweichung, der Varianz gerechnet. Man benutzt die Schreibweise

$$\underline{l}_i \sim N(E(\underline{l}_i), \sigma_{l_i}^2), \quad (2)$$

um \underline{l}_i als normal verteilte Zufallsvariable mit Erwartungswert $E(\underline{l}_i)$ und Varianz $\sigma_{l_i}^2$ zu bezeichnen. Entsprechend kann man mehrere Zufallsvariable \underline{l}_i zu einem Zufallsvektor \underline{l} zusammenfassen. An die Stelle der Varianz tritt dann die Varianz-Kovarianzmatrix $K_{ll} = \sigma_o^2 Q_{ll}$.

Die Kovarianzmatrix

$$K_{ll} = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \cdots & \rho_{1n}\sigma_1\sigma_n \\ \rho_{21}\sigma_2\sigma_1 & \sigma_2^2 & \cdots & \rho_{2n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1}\sigma_n\sigma_1 & \rho_{n2}\sigma_n\sigma_2 & \cdots & \sigma_n^2 \end{pmatrix}$$

enthält auf der Hauptdiagonalen die Varianzen, d.h. die Quadrate der Standardabweichungen der einzelnen Beobachtungen und auf den Nebendiagonalelementen die Kovarianzen, die vom jeweiligen Korrelations-Koeffizienten ρ_{ij} abhängen. Da wir i.a. keine Korrelationen zwischen den Beobachtungen voraussetzen, ist K_{ll} i.a. eine Diagonalmatrix. Die Gewichtsmatrix P_{ll} ist die Inverse der Gewichtskoeffizientenmatrix Q_{ll} , die sich von der Kovarianzmatrix nur um den Varianzfaktor bzw. das Quadrat des mittleren Gewichtseinheitsfehlers unterscheidet. Die gemeinsame Verteilung der Beobachtungen $\underline{l} = (\underline{l}_i)$ wird mit

$$\underline{l} \sim N(E(\underline{l}), \sigma_o^2 Q_{ll}) \quad (3)$$

bezeichnet.

2.3 Neben der Annahme über die Art der Verteilung der Beobachtungen treffen wir Annahmen über den Erwartungswert und die Genauigkeit der Beobachtungen. Bei der Genauigkeit der Beobachtungen haben wir, bedingt durch die Methode der kleinsten Quadrate, wenig Spielraum: Wir müssen die Gewichtsverhältnisse unter den Beobachtungen vorgeben, d.h. die mittleren Fehler bis auf einen gemeinsamen Faktor σ_o^2 und die Korrelationen als bekannt voraussetzen:

$$K_{ll} = \sigma_o^2 Q_{ll} \quad (4)$$

Aus Kostengründen und weil i.a. zu wenig Information vorliegt nehmen wir fast ausschließlich unkorrelierte Beobachtungen an. Die Vernachlässigungen sind beträchtlich, denn Bildkoordinaten können bis zu 70 % korreliert sein. Bei photogrammetrischen Messungen nimmt man sogar an, daß die Beobachtungen, Bild- oder Modellkoordinaten, gleich genau sind, obwohl die mittleren Fehler innerhalb der Bilder bzw. Modelle ohne weiteres um den Faktor 2 streuen können. Für die Überprüfung der Gewichtsannahmen stehen Verfahren zur Verfügung, auf die wir aber hier nicht weiter eingehen.

Annahmen über den Erwartungswert lassen demgegenüber sehr viel mehr Freiheit. Sie führen auf die bekannte Form der Fehlergleichungen:

$$E(\underline{l}) = A x \quad (5)$$

Sie formulieren eine erwartete (lineare oder linearisierte) Abhängigkeit der Beobachtungen von unbekanntem Faktoren x (Koordinaten, Transformationsparameter etc.). Sie ist wesentlich genauer bekannt. Etwa repräsentiert die Zentralperspektive den photographischen Aufnahme-prozeß so gut, daß die verbleibenden systematischen Restfehler im 10 μ m Bereich bleiben. Die tatsächlichen Beobachtungen \underline{l} erfüllen das Modell wegen der genannten Vereinfachungen nicht. Die Differenz

$$\underline{\varepsilon} = \underline{l} - Ax \quad (5a)$$

zwischen den Beobachtungen und dem Modell sind Beobachtungs- oder Modellfehler. Eine Trennung von beiden ist prinzipiell nicht möglich. Etwa kann ein 1 m-Fehler als Beobachtungsfehler aber auch als Fehler im Modell angesehen werden, da die momentane Unkonzentriertheit des Beobachters nicht berücksichtigt wird.

Das der Berechnung zugrunde gelegte Modell wird also durch das sogenannte stochastische Modell (Gl. (4)) und das Funktionalmodell (Gl. (5)) beschrieben.

2.4 Da uns die Faktoren x und σ_0^2 und der Erwartungswert $E(\underline{l})$ der Beobachtungen nicht bekannt sind, leiten wir aus den tatsächlichen Beobachtungen \underline{l} Schätzungen \hat{x} , $\hat{\sigma}_0^2$ und \hat{l} mit Hilfe der Methode der kleinsten Quadrate ab. Wir erhalten nacheinander die Unbekannten

$$\hat{x} = (A^T P A)^{-1} A^T P \underline{l} \quad (6)$$

den Schätzwert \hat{l} für $E(\underline{l})$

$$\hat{l} = A \hat{x} \quad (7)$$

den Schätzwert $\hat{\varepsilon}$ für die Modell- bzw. Beobachtungsfehler,

$$\hat{\varepsilon} = \underline{l} - A \hat{x} \quad (8)$$

vom Betrag her identisch mit den Verbesserungen

$$\underline{v} = - \hat{\varepsilon} \quad (8a)$$

und eine Schätzung $\hat{\sigma}_0^2$ für den unbekanntem Varianzfaktor σ_0^2

$$\hat{\sigma}_0^2 = \frac{\underline{v}^T P \underline{v}}{r} \quad (9)$$

worin r die Redundanz ist.

2.5 Wir müssen nun beurteilen, ob dieses Ergebnis im Widerspruch zu unseren Annahmen steht, d.h. wir müssen die Schätzungen mit gegebenen oder zu erwartenden Werten vergleichen. Für keinen dieser Werte liegen i.a. gegebene Werte vor (außer etwa bei Wiederholungs- oder Deformationsmessungen). Bei den Schätzungen \hat{x} , \hat{l} und \underline{v} ist uns nur der Erwartungswert $E(\underline{v}) = 0$ bekannt:

$$\begin{aligned}
 E(\underline{v}) &= E(\underline{Ax} - \underline{l}) \stackrel{(6)}{=} E(A(A^T P A)^{-1} A^T P \underline{l} - \underline{l}) \\
 &= A(A^T P A)^{-1} A^T P E(\underline{l}) - E(\underline{l}) \\
 &\stackrel{(5)}{=} A(A^T P A)^{-1} A^T P A x - E(\underline{l}) \\
 &= Ax - E(\underline{l}) = 0
 \end{aligned} \tag{10}$$

Wir kennen über das Fehlerfortpflanzungsgesetz auch die Standardabweichungen σ_{v_i} der Verbesserungen \underline{v}_i , die - als lineare Funktionen der Beobachtungen - normal verteilt sind.

Daher sind allein die Verbesserungen selbst oder Funktionen der Verbesserungen für die Überprüfung der Beobachtungen bzw. unserer Annahmen geeignet. Dazu gehört die Schätzung $\hat{\sigma}_0^2$ und die normierten Verbesserungen

$$\underline{w}_i = - \frac{\underline{v}_i}{\sigma_{v_i}} = \frac{\hat{\underline{\epsilon}}_i}{\sigma_{\hat{\underline{\epsilon}}_i}} \quad (P = \text{Diagonalmatrix}) \tag{11}$$

des Tests gegen grobe Datenfehler. Für \underline{w}_i gilt

$$\underline{w}_i \sim N(0,1), \tag{12}$$

d.h. die normierten Verbesserungen sind normalverteilt mit dem Erwartungswert $E(\underline{w}_i) = 0$ und der Standardabweichung $\sigma_{w_i} = 1$.

3. Der statistische Test

3.1 Die Prüfung der Verbesserungen dient dazu, Fehler in unseren Annahmen, d.h. dem Modell, aufzudecken. Damit gleichbedeutend können wir dem gegebenen Modell ein oder mehrere Alternativmodelle gegenüberstellen, die sich um die vermuteten Fehler vom gegebenen Modell unterscheiden. In statistischer Terminologie nennt man das gegebene Modell die Nullhypothese H_0 und die um die Fehler erweiterten Modelle Alternativhypothesen H_{ap} . Der Index p bringt zum Ausdruck, daß es sich hier um mehrere Alternativen $p = 1, \dots$ handeln kann. Mit Hilfe geeigneter Tests möchte man zwischen H_0 und H_{ap} entscheiden.

In unserem Fall lautet die Nullhypothese H_0 :

H_0 : Es sind keine groben Fehler vorhanden, Gleichungen (4) und (5) beschreiben die Wirklichkeit.

$$\begin{aligned}
 \text{bzw. } E(\underline{l} \mid H_0) &= A x \\
 E(\underline{l}_i \mid H_0) &= a_i^T x, \quad i=1,2,\dots
 \end{aligned} \tag{13}$$

Für jede Beobachtung \underline{l}_i können wir eine Alternativhypothese aufstellen,

H_{a_i} : Es liegt ein grober Fehler in der Beobachtung \underline{l}_i vor.

Die erwartete Größe ist $E(\underline{v} \mid H_{a_i}) = \widetilde{\underline{v}}_i$

$$E(\underline{l}_i \mid H_{a_i}) = a_i^T x + \widetilde{\underline{v}}_i,$$

$$E(\underline{l}_j \mid H_{a_i}) = a_j^T x, \quad j \neq i \tag{14}$$

Wir betrachten hier nur die einfachste Annahme, daß nämlich nur eine Beobachtung grob falsch ist und die anderen l_j ($j \neq i$) nicht verfälscht sind. Die Gegenüberstellung der beiden Hypothesen H_0 und H_{a_i} ist in der Figur 2 veranschaulicht.

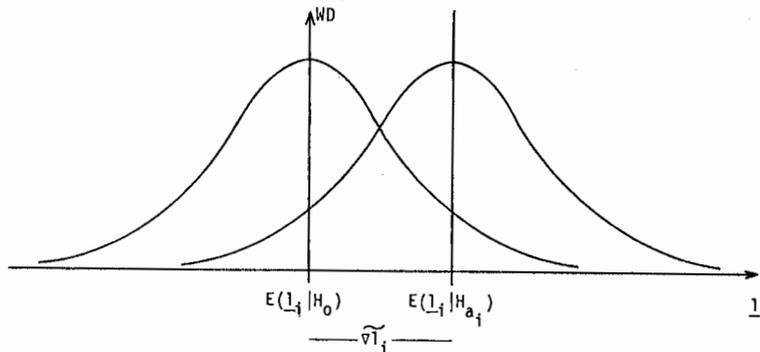


Fig. 2 Dichtefunktion für grob falsche Beobachtung

Der Fehler \widetilde{v}_l_i führe dazu, daß die Wahrscheinlichkeitsverteilung um den Betrag \widetilde{v}_l_i verschoben wird. An der Art der Verteilung ändere sich nichts.

Beispiel:

An einem Meßband wird die Strecke um 1 m falsch abgelesen: $\widetilde{v}_l = 1$ m.
 Die Genauigkeit der Ablesung und damit die Standardabweichung des Meßwerts wird von diesem Fehler nicht beeinflusst.

Gegenbeispiele:

- a) Ähnliches gilt für die Messung eines falschen Signals in einem Meßbild. Die Genauigkeit der Einstellung wird von der Fehlidentifizierung nicht verändert. Wir haben es aber mit 2 groben Fehlern \widetilde{v}_x und \widetilde{v}_y zu tun.
- b) Die Messung einer Strecke wird von den Witterungsbedingungen verfälscht. Das betrifft i.a. den Mittelwert und die Genauigkeit.
- c) Die Registrierung eines falschen Instruments (Bandmaß statt Latte) führt i.a. zu einem groben Fehler des Gewichts der Beobachtung, während der Mittelwert i.a. nicht beeinflusst wird.

Trotzdem beschränken wir uns hier auf den einfachsten Fall eines einzelnen groben Fehlers. Eine entsprechende Erweiterung auf mehrere grobe Fehler oder auf Fehler in den Gewichtsannahmen ist aber prinzipiell möglich (s. Abschn. 5).

3.2 Der Test besteht in dem Vergleich der normierten Verbesserung w_i mit dem frei wählbaren kritischen Wert k (z.B. $k = 3$, 3 σ -Regel). Die Situation ist in Fig. 3 für den Fall dargestellt, daß kein grober Fehler vorliegt. Der Bereich $(-k, +k)$ ist der Annahmehereich A , da wir die Beobachtung l_i akzeptieren, wenn w_i in diesen Bereich fällt. Wir nehmen H_0 an. Im anderen Fall, daß w_i in den Verwerfungsbereich V fällt, d.h. $|w_i| > k$ ist, lehnen wir die Nullhypothese ab und vermuten einen groben Fehler.

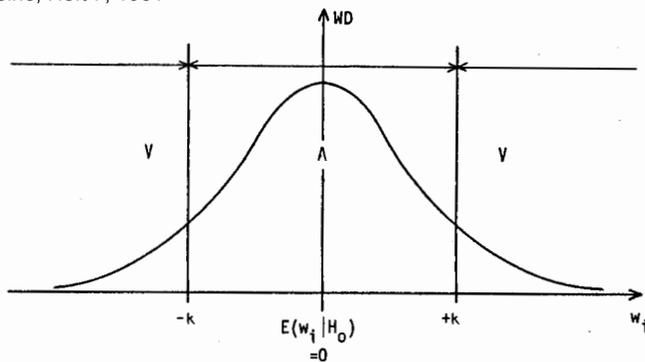


Fig. 3 Annahme- und Verwerfungsbereich

Beispiel:

a) $\underline{v}_i = 1 \text{ cm}$, $r_i = 0.3$, $\sigma_{l_i} = 1.0 \text{ cm} \rightarrow |\underline{w}_i| = 1.83 < 3 = k$
 bzw. $\underline{v}_i < 3 \cdot \sigma_{v_i} = 3 \cdot \sigma_{l_i} \sqrt{r_i} = 1.64 \text{ cm}$.

Die Beobachtung wird akzeptiert.

b) $\underline{v}_i = 5 \text{ } \mu\text{m}$, $r_i = 0.09$, $\sigma_{l_i} = 5 \text{ } \mu\text{m} \rightarrow |\underline{w}_i| = 3.33 > 3 = k$
 bzw. $\underline{v}_i > 3 \cdot \sigma_{v_i} = 4.5 \text{ } \mu\text{m}$. Die Beobachtung wird verworfen. Der kritische Wert von $4.5 \text{ } \mu\text{m}$ für die Verbesserung \underline{v}_i ist kleiner als die Standardabweichung σ_{l_i} der Beobachtung!

Der Vergleich von a) und b) zeigt den Einfluß des unterschiedlichen Redundanzanteils auf die Testgröße \underline{w}_i .

3.3 Wir wollen nun die Entscheidungen beim Testen näher betrachten. Dazu unterscheiden wir

1. ob die normierte Verbesserung in den Annahme- oder in den Verwerfungsbereich fällt;
2. ob ein grober Fehler vorliegt oder nicht.

Ein grober Fehler $\widetilde{\nabla l_i}$ führt zu einer Verfälschung $\delta_i = \widetilde{\nabla w_i}$ der Testgröße \underline{w}_i und damit zu einer Verschiebung der Verteilungsfunktion von \underline{w}_i .

Mit $\nabla v_i = -r_i \nabla l_i$ gilt für die Verschiebungsgröße (Nichtzentralitätsparameter)

$$\delta_i = \widetilde{\nabla w_i} = \frac{-\widetilde{\nabla v_i}}{\sigma_{l_i} \sqrt{r_i}} = \frac{\widetilde{\nabla l_i}}{\sigma_{l_i}} \cdot \sqrt{r_i} . \tag{15}$$

Je schwächer die Geometrie ($r_i \ll 1$) bzw. je größer die Standardabweichung σ_{l_i} der Beobachtung, desto kleiner ist der Einfluß eines Beobachtungsfehlers auf \underline{w}_i .

In Figur 4 ist die Verteilungsfunktion für drei Annahmen, d.h. Hypothesen dar-

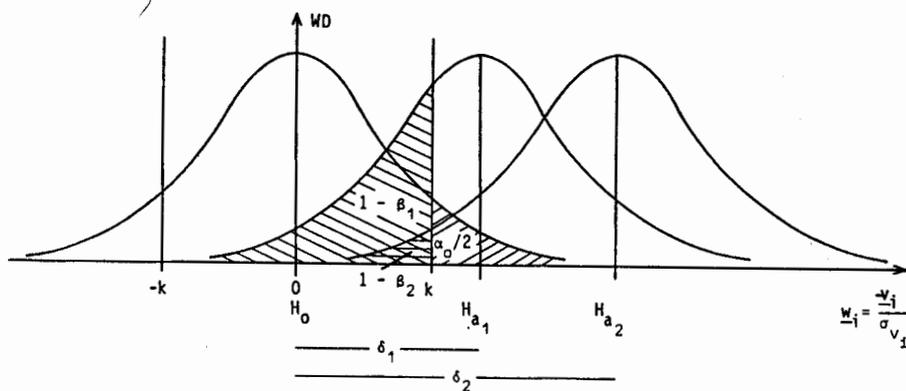


Fig. 4 Zur Macht eines Tests

H_0 (Nullhypothese): Die Beobachtung \underline{l}_i ist fehlerfrei. Der Erwartungswert der Testgröße ist

$$E(\underline{w}_i | H_0) = 0$$

H_{a1} (1. Alternativhypothese): Die Beobachtung \underline{l}_i ist um den Betrag $\widetilde{v}_1 \underline{l}_i$ grob falsch, der auf $\widetilde{v}_1 \underline{w}_i = \delta_1$ führt. Damit verschiebt sich die Dichtefunktion von \underline{w}_i um δ_1 . Also gilt:

$$E(\underline{w}_i | H_{a1}) = \delta_1.$$

H_{a2} (2. Alternativhypothese): Die Beobachtung \underline{l}_i ist um den größeren Betrag $\widetilde{v}_2 \underline{l}_i$ grob falsch, der auf $\widetilde{v}_2 \underline{w}_i$ führt. Analog gilt

$$E(\underline{w}_i | H_{a2}) = \delta_2.$$

Die Wahrscheinlichkeit, daß w zwischen $-k$ und k liegt, - die Fläche unter der linken Kurve zwischen $-k$ und k - ist das Signifikanzniveau S_0 des Tests:

$$S_0 = P(|\underline{w}_i| < k | H_0)$$

Aus k kann man S_0 ableiten und umgekehrt. Die Ergänzung $\alpha_0 = 1 - S_0$ wird die Signifikanzzahl genannt.

Ist $|w| < k$, so wird die Hypothese H_0 nicht verworfen, d.h. es wird angenommen, daß kein Fehler vorliegt. Ist dagegen $|w| > k$, so wird H_0 verworfen, und ein Fehler vermutet. Über dessen Größe ist nichts gesagt. Die Wahrscheinlichkeit dafür, daß H_0 abgelehnt wird, und H_{ap} zutrifft, ist dargestellt durch die Fläche rechts von k unter der mittleren bzw. rechten Kurve.

$$\beta_p = P(|w| > k | H_{ap}).$$

Die Wahrscheinlichkeit β_p , einen Fehler der Größe $\widetilde{v}_p \underline{l}_i$ zu finden, ist die Macht des Tests bezüglich der Alternativhypothese H_{ap} .

Da die Größe der Verbesserungen und damit auch \underline{w} sowohl von groben als auch von zufälligen Fehlern bestimmt wird, sind 4 Entscheidungssituationen möglich, die in Tabelle 1 zusammen mit ihrer Wahrscheinlichkeit dargestellt sind.

	$ \underline{w} < k$	$ \underline{w} > k$
H_0 richtig	H_0 annehmen Signifikanzniveau $1 - \alpha_0$ richtige Entscheidung	H_0 ablehnen Signifikanzzahl α_0 Fehler I. Art
H_{ap} richtig	H_0 annehmen $1 - \beta_p$ Fehler II. Art	H_0 ablehnen Macht β_p richtige Entscheidung

Tab. 1 Entscheidungen beim Testen

Neben den genannten - richtigen Entscheidungen, sind zwei Arten von Fehlentscheidungen möglich:

Fehler I. Art: Obwohl H_0 richtig ist, d.h. kein Beobachtungsfehler vorliegt, wird wegen $|\underline{w}| > k$ auf einen Beobachtungsfehler geschlossen. Die Wahrscheinlichkeit für einen Fehler I. Art, ist die Signifikanzzahl α_0 . Da mit der Ablehnung von H_0 die hier überflüssige Suche nach groben Fehlern beginnt, wird α_0 in der Wirtschaftsstatistik auch als Produzentenrisiko bezeichnet und üblicherweise klein gewählt, etwa 0.01 oder 0.001.

Fehler II. Art: Obwohl H_{ap} richtig ist, d.h. die Beobachtung falsch ist, wird wegen $|\underline{w}| < k$ kein Fehler vermutet. Die Wahrscheinlichkeit $1 - \beta_p$ für diesen Fehler II. Art hängt bei gegebenem kritischem Wert auch von der Größe des groben Fehlers d.h. auch von δ_p ab; je größer der Fehler, desto geringer das Risiko, daß er unentdeckt bleibt (vgl. rechte Kurve in der Figur). $1 - \beta_p$ wird in der Wirtschaftsstatistik auch als Konsumentenrisiko bezeichnet. Es ist die Ergänzung zur Macht β_p des Tests.

Drei Beispiele für die Abhängigkeit der Größen k , α , β und δ untereinander zeigen die Übereinstimmung dieser Betrachtungsweise mit der Erfahrung.

- a) Die Wahrscheinlichkeit α für einen Fehler I. Art, d.h. eine richtige Beobachtung zu verwerfen, ist umso kleiner je größer der kritische Wert ist (vgl. Tab. 2).

k	α
1.96	5 %
2.56	1 %
3.0	0.27 %
3.29	0.1 %

Tab. 2
Signifikanzzahl α und
kritischer Wert k

- b) Die Macht des Tests, d.h. die Wahrscheinlichkeit einen groben Fehler zu finden, nimmt mit seiner Größe zu (vgl. Tab. 3). Umgekehrt nimmt die Wahrscheinlichkeit $1 - \beta_i$ ab, daß dieser Fehler unerkannt bleibt.

δ_i	β_i	$1 - \beta_i$
3	50 %	50 %
4	84 %	16 %
5	97.7 %	2.3 %
6	99.9 %	0.1 %

Tab. 3
Verschiebungsgröße δ_i und Macht β_i
des Tests ($k = 3$)

Insbesondere lassen sich Fehler, die gleich groß wie der kritische Wert sind, nur mit 50 %-iger Sicherheit finden!

c) Die Sicherheit einen Fehler zu finden wird größer, wenn der kritische Wert kleiner wird (vgl. Tab. 4).

α	k	β
99.9 %	3.29	76 %
99.7 %	3.0	84 %
99.0 %	2.56	93 %
95.0 %	1.96	98 %

Tab. 4
 Signifikanzzahl α und Macht
 des Tests ($\delta = 4$)

4. Der Begriff der inneren Zuverlässigkeit

Zwei Fragestellungen lassen sich mit Hilfe der Macht des Tests beantworten:

1. Mit welcher Sicherheit läßt sich ein grober Fehler $\nabla_p l_i$ ($p=1,2,\dots$) aufdecken, wenn der Test mit einem Signifikanzniveau S_0 durchgeführt wird?

Die Macht β_p hängt von α_0 und von δ_p ab:

$$\beta_p = \beta_p(\alpha_0, \delta_p)$$

Da man i.a. über die Größe der groben Fehler nichts aussagen kann, ist mehr die folgende Fragestellung von Bedeutung:

2. Welche Größe $\nabla_0 l_i$ (Index 0 für untere Grenze) muß ein Fehler mindestens aufweisen, damit man ihn mit einer gegebenen Mindestsicherheit β_0 durch den Test mit dem Signifikanzniveau S_0 finden kann? Der Mindestabstand $\delta_0 = \nabla_0 w_i$ der Alternative H_a von H_0 hängt von α_0 und β_0 ab:

$$\nabla_0 w_i = \delta_0 = \delta_0(\alpha_0, \beta_0).$$

Aus diesem unteren Abstand $\nabla_0 w_i$ kann man über Gl. 15 auf den unteren Grenzwert $\nabla_0 l_i$ für einen groben Fehler ∇l_i schließen, der bei gegebenem Signifikanzniveau S_0 gerade mit der Wahrscheinlichkeit β_0 vom Test aufgedeckt wird:

$$\nabla_0 l_i = \sigma l_i \frac{\delta_0}{\sqrt{r_i}}$$

Der untere Grenzwert enthält:

- die Genauigkeit der Beobachtung, beschrieben durch den mittleren Fehler σl_i .
- die Netzgeometrie in der Umgebung der Beobachtung, beschrieben durch den Redundanzanteil r_i und
- das Signifikanzniveau $1 - \alpha_0$ und die Mindestsicherheit β_0 (Macht des Tests, beschrieben durch den Nichtzentralitätsparameter $\delta_0 = \delta_0(\alpha_0, \beta_0)$).

Der Grenzwert ist nicht identisch mit dem kritischen Wert k, der für den Test verwendet wird, sondern immer größer (falls $\beta_0 > 50 \%$).

Für verschiedene α_0 und β_0 sind in Tabelle 5 die zugehörigen Werte für δ_0 angegeben.

$\beta_0 \backslash \alpha_0$	k	3.72	3.29	2,58	1.96
		0.01 %	0.1 %	1 %	5 %
70		4.41	3.82	3.10	2.48
80		4.73	<u>4.13</u>	3.42	2.80
90		5.17	4.57	3.86	3.24
95		5.54	4.94	4.22	3.61
99		6.22	5.62	4.90	4.29
99.9		6.98	6.38	5.67	5.05

Tab. 5 Nichtzentralitätsparameter δ_0

Je kleiner der Redundanzanteil r_i der Beobachtung, umso größer muß ein grober Fehler sein, um erkannt zu werden, umso kleiner ist ja auch die Verbesserung. Die Grenzwerte geben ein Bild über die innere Zuverlässigkeit des Netzes.

Der Fall $r_i = 1$ tritt nur auf, wenn der wahre Wert der Beobachtung bekannt ist, etwa bei einer Strecke zwischen Festpunkten, wenn keine Maßstabsunbekannte angesetzt ist. Dann können grobe Fehler $\nabla l_i < \delta_0 \cdot \sigma l_i$ nur mit einer Wahrscheinlichkeit kleiner als β_0 erkannt werden, entsprechend kann man grobe Fehler $\nabla l_i > \delta_0 \cdot \sigma l_i$ mit einer Wahrscheinlichkeit größer als β_0 erkennen. Z.B. lassen sich bei einem Signifikanzniveau von $S_0 = 1 - \alpha_0 = 99.9 \%$ erst grobe Fehler $\nabla l_i > \delta_0 \cdot \sigma l_i = 4.13 \sigma l_i$ mit größerer als 80 %-iger Sicherheit erkennen.

Wird im anderen Extrem eine Beobachtung durch die Netzgeometrie überhaupt nicht kontrolliert, so ist $r_i = v_i = 0$ und der Test nicht möglich.

Beispiel:

- a) Strecke $\sigma l_i = 1.0 \text{ cm}, r_i = 0.25, \nabla_0 l_i = 8.0 \text{ cm}$
- b) " $\sigma l_i = 1.5 \text{ cm}, r_i = 0.7, \nabla_0 l_i = 7.2 \text{ cm}$
- c) Bildkoordinate $\sigma l_i = 5 \text{ } \mu\text{m}, r_i = 1/6, \nabla_0 l_i = 49 \text{ } \mu\text{m}$
- d) " $\sigma l_i = 3 \text{ } \mu\text{m}, r_i = 1/12, \nabla_0 l_i = 42 \text{ } \mu\text{m}$

Der Vergleich der Grenzwerte ist unmittelbar nicht möglich, zumindest muß man Strecken und Bildpunkte getrennt betrachten. Aber auch dann bleibt die Frage offen, ob ein Grenzwert von 42 μm besser ist als ein Grenzwert von 49 μm . Denn bezogen auf die Genauigkeit der Beobachtungen ist die zweite Bildkoordinate mit σl_i schlechter kontrollierbar, ihr Redundanzanteil ist kleiner als der der ersten Bildkoordinate.

Wir teilen daher den Grenzwert $\nabla_0 l_i$ auf in einen Teil, der nur die Genauigkeit, und einen Teil, der nur die Zuverlässigkeit enthält. Wir definieren das auf die Genauigkeit bezogene Maß für die Kontrollierbarkeit der Beobachtungen $\delta'_{0,i}$:

$$\delta'_{0,i} = \frac{\delta_0}{\sqrt{r_i}}$$

Damit wird der Grenzwert $\nabla_0 l_i$ zu:

$$\nabla_0 l_i = \delta'_{0,i} \cdot \sigma l_i$$

Der Kontrollierbarkeitswert $\delta'_{0,i}$ gibt also den Faktor an, um den ein grober Fehler $\nabla_0 l_i$ mindestens größer als die Standardabweichung sein muß, um mit dem Test mit einer Mindestsicherheit von β_0 erkannt werden zu können.

Die Kontrollierbarkeitswerte $\delta'_{0,i}$ für die 4 Beobachtungen sind also a) 8., b) 4.8, c) 9.8, d) 14.. Damit ist die Strecke b) mit $\nabla_0 l_i = 4.8 \sigma l_i$ am besten und die Bildkoordinate d) mit $\nabla_0 l_i = 14 \sigma l_i$ am schlechtesten kontrollierbar.

Bei einer einmaligen Festlegung der Signifikanzzahl α_0 und der Sicherheit β_0 hängt die Kontrollierbarkeit einer Beobachtung l_i in sehr einfacher Weise von ihrem Redundanzanteil r_i ab. Beurteilung der Kontrollierbarkeitsmaße ist nichts anderes als Beurteilung der Geometrie. Allerdings enthalten die Grenzwerte $\nabla_0 l_i$ auch Informationen über den Test, und können etwa zum Vergleich verschiedener Testverfahren dienen.

Z.B. erhält man bei dem üblichen Testverfahren, das die Geometrie nicht berücksichtigt, also mit der Testgröße

$$\underline{w}_i^* = \frac{-v_i}{\sigma l_i}$$

arbeitet, die Grenzwerte

$$\nabla_0^* l_i = \sigma l_i \frac{\delta_0}{r_i}$$

Sie sind i.a. größer als $\nabla_0 l_i$. Vor allem bei schwacher Überbestimmung ($r_i \ll 1$) ist der Unterschied beträchtlich. Für $r_i = 1/9$ ist $\nabla_0^* l_i = 3 \nabla_0 l_i$! Etwa lassen sich mit der Testgröße \underline{w}_i^* im Fall d) erst Fehler ab 144 μm erkennen, gegenüber 42 μm mit der Testgröße \underline{w}_i .

5. Diskussion

Mit den Grenzwerten läßt sich die Kontrollierbarkeit der Beobachtungen beurteilen. Sie hängt ab vom Ausgleichungsansatz und vom Testverfahren. Wir haben einige Einschränkungen in Kauf nehmen müssen, die Zweifel lassen, ob das Verfahren für die praktische Anwendung geeignet ist.

5.1 Dies betrifft zunächst die i.a. unrealistische Voraussetzung, daß nur ein grober Fehler im Datenmaterial ist. Nur unter dieser Bedingung ist der Test optimal, d.h. es gibt keinen anderen Test, der bei gleichen Annahmen einen einzelnen groben Fehler sicherer findet. Ein Test auf mehrere grobe Fehler, die jeweils nur eine Beobachtung betreffen, wurde von STEFANOVIC (1978) vorgestellt. Es ist eine Verallgemeinerung des data-snooping. Das Problem mehrerer grober Fehler ist eng mit der Strategie für die Fehlersuche verknüpft. Wir werden darüber morgen im Vortrag von Herrn KLEIN mehr hören.

5.2 Ganz stillschweigend haben wir als zweites vorausgesetzt, daß eine einzelne Fehlerursache nur eine einzelne Beobachtung verfälscht. In photogrammetrischen Blöcken sind Fehlerursachen, die mehrere Beobachtungen betreffen, etwa Punktübertragungsfehler (leider) häufiger. Aber auch die Koordinaten einer Paßpunktgruppe können, verursacht durch einen groben Fehler etwa im Nivellement, um einen gleich großen Betrag verfälscht sein. Gerade für den letzten Fall hat BAARDA einen allgemeineren Test angegeben, mit dem auch ein systematischer Fehler, der alle Beobachtungen verfälscht, geprüft werden kann. Eine Verallgemeinerung des Tests auf mehrere grobe Fehler, die jeweils mehrere Beobachtungen betreffen können, ist möglich (FÖRSTNER, 1980). Damit läßt sich der größte Teil der bei der photogrammetrischen Punktbestimmung vorkommenden Fehler behandeln, also auch Punktübertragungsfehler oder Punktgruppenverwechslungen.

5.3 In allen Fällen kann man die Größe gerade noch mit dem Test erkennbarer grober Fehler angeben. Für den Fall zweier grober Fehler erhält man statt der Grenzwerte, ellipsenförmige Bereiche in denen der Punkt liegen kann, ohne vom Test als grob falsch erkannt zu werden. Ein Beispiel für diese "Grenzwert-Ellipsen" enthält Fig. 5 für einen kleinen Bündelblock mit 3 Bildern. Es wurde angenommen, daß jeweils ein Punkt verfälscht ist, d.h. x und y grob falsch sind. Beim linken Bildpunkt in der Mitte des Streifens entartet die Grenzwert-Ellipse zu einem Band mit der Breite $\nabla_0 y_i$. Die x -Koordinate dieses Punktes ist nicht kontrollierbar, da die Kernebene parallel zur x -Achse verläuft. In einem anderen Bildkoordinatensystem (leicht verschwenkt oder gar durch die Bildecken) wären x - und y -Koordinaten dieses Punktes kontrollierbar, obwohl sich an dem Grenzwertbereich nichts ändert. Dies ist ein Hinweis, daß das einfache data-snooping und die mit ihm verbundenen Aussagen über die Kontrollierbarkeit nicht immer ganz hinreichen.

5.4 Schließlich bleibt die Frage nach der Genauigkeit der Beobachtungen. Das betrifft einerseits den Varianzfaktor σ_0 , der als bekannt vorausgesetzt wurde. Verwendet man statt σ_0 die Schätzung aus der Ausgleichung $\hat{\sigma}_0$, so führt das auf den τ -Test von POPE (1975), wenn man bei der Berechnung von $\hat{\sigma}_0$ die verdächtigste Beobachtung mitnimmt. Wenn man dagegen $\hat{\sigma}_0$ aus einer Ausgleichung ohne die betroffene Beobachtung nimmt (die Ausgleichung ist dafür nicht zu wiederholen!), erhält man den t -Test (vgl. KRÜGER, 1976), für mehrere Fehler einen Fisher-Test (vgl. FÖRSTNER, 1980).

Auf der anderen Seite geht das Gewicht p_i der Beobachtung unmittelbar in die Testgröße ein. Grobe Fehler in den Gewichten lassen sich von groben Fehlern der Beobachtungswerte prinzipiell nicht trennen, weshalb man auf die Festlegung des Gewichts besondere Sorgfalt verwenden sollte.

Trotz aller dieser Einschränkungen kann man das "data-snooping" nur empfehlen. Denn schon allein die Berücksichtigung der Geometrie über die Redundanzanteile macht es einfacheren Testverfahren überlegen und begründet die guten Erfahrungen.

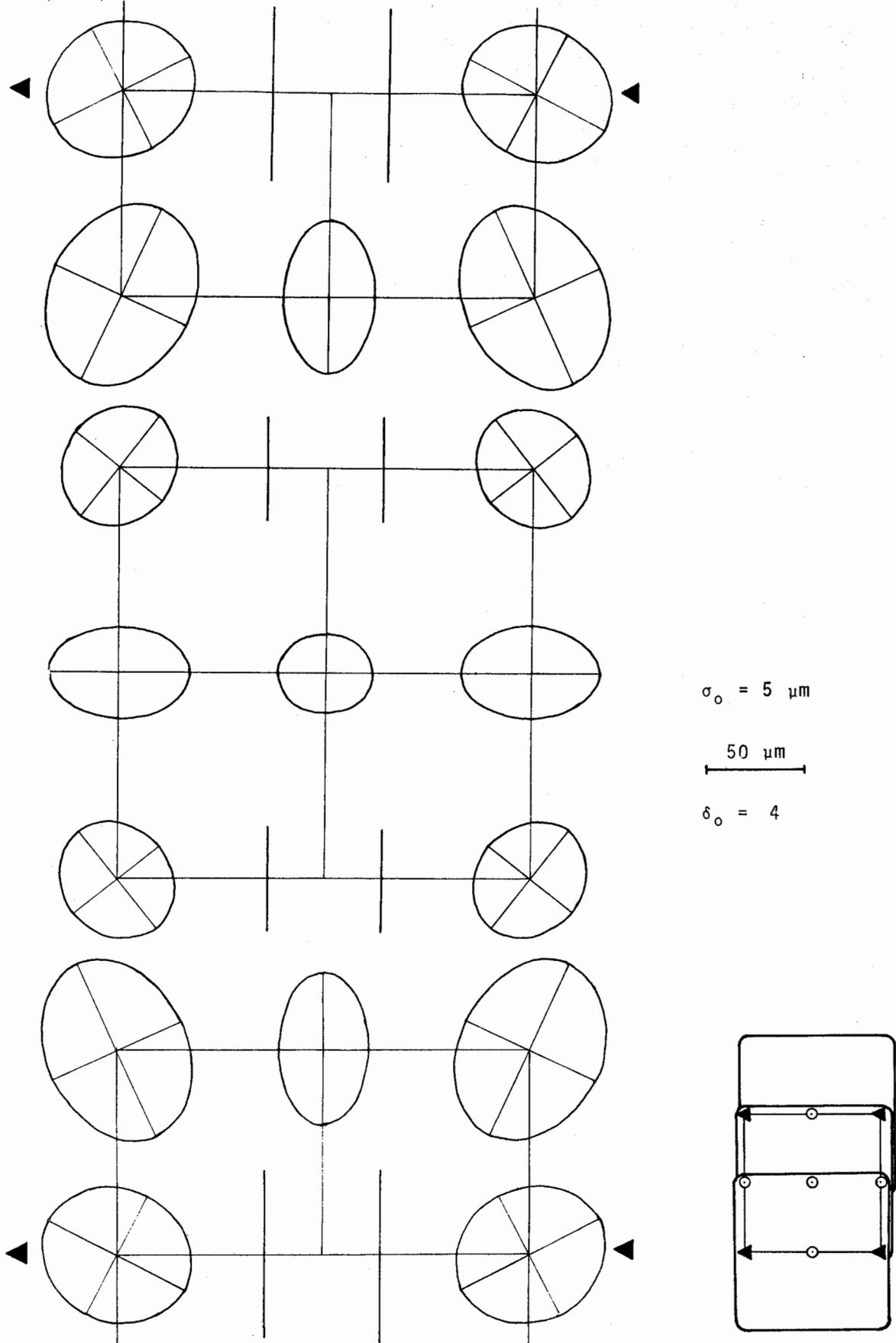


Fig. 5 Grenzwert- bzw. Kontrollierbarkeitsellipsen in einem Bündelblock

Literatur

- BAARDA, W.: Statistical Concepts in Geodesy, Netherlands Geodetic Commission, Vol. 2, No. 4, 1967
- BAARDA, W.: A Testing Procedure for Use in Geodetic Networks, Netherlands Geodetic Commission, Vol. 2, No. 5, 1968
- FÜRSTNER, W.: Das Rechenprogramm TRINA für geodätische Netze in der Landesvermessung, Nachrichten aus dem öffentlichen Vermessungswesen Nordrhein-Westfalen, 1979, Heft 2
- FÜRSTNER, W.: Evaluation of Block Adjustment Results, Presented Paper to International Congress for Photogrammetry, Hamburg, 1980
- KREYSZIG, E.: Statistische Methoden und ihre Anwendungen, Göttingen, 1968
- KRÜGER, J.: Statistisches Verfahren zur Lokalisierung grober Beobachtungsfehler in geodätischen Netzen, dargestellt an Streckennetzen, Dissertation Hannover, 1976
- POPE, A.: The Statistics of Residuals and the Detection of Outliers, Presented Paper to the XVIth General Assembly of the IAG, Grenoble, 1975
- STEFANOVIC, P.: Blunders and Least Squares, ITC-Journal 1978-1