Normierte Verbesserungen – wie groß ist zu groß?

Rüdiger Lehmann

ALLGEMEINE VERMESSUNGS-NACHRICHTEN

Dieser Beitrag wurde nach Begutachtung als PEER REVIEWED PAPER angenommen.
Tag der Einreichung: 17. August 2009

Im Konzept des Data Snooping wird ein Messwert als grob fehlerbehaftet angesehen, wenn seine normierte Verbesserung einen gewissen kritischen Wert überschreitet. Die Größe des kritischen Wertes wird meist aus der Irrtumswahrscheinlichkeit für den entsprechenden Hypothesentest abgeleitet. Wir schlagen vor, diesen Wert so zu wählen, dass die Varianz der Schätzfunktion minimiert wird. Dann hängt dieser Wert vom Ausgleichungsmodell ab und muss vorher mit der Monte-Carlo-Methode bestimmt werden. Wir illustrieren die Methode am Beispiel einer Messreihe (direkte Messungen) und einer ausgleichenden Geraden.

1 Einleitung

Jeder im Vermessungswesen Tätige ist schon mit groben Abweichungen und Ausreißern in Messdaten in Berührung gekommen, sei es beim Sammeln solcher Daten in der Örtlichkeit oder bei ihrer Bearbeitung im Büro. Trotzdem gibt es in vielen Fällen keine etablierte und erst recht keine allgemein verbindlich festgelegte Vorgehensweise. Leider neigen Menschen im Umgang mit empirischen Daten intuitiv zu Fehlschlüssen, so dass die Intuition keine gute Richtschnur darstellt. So lesen Menschen aus Daten tendenziell das heraus, was ihnen plausibel erscheint oder ihren Erwartungen entspricht. Die Kognitionswissenschaft kennt viele Formen der kognitiven Voreingenommenheit im Umgang mit empirischen Daten (Haselton u.a. 2005).
Folgende Verfahren im Umgang mit groben Abweichungen und Ausreißern sind im Vermessungswesen mehr oder weniger etabliert (Jäger u.a. 2005):

1.1 Klassisches Data Snooping

Das Gauß-Markov-Modell (GMM) der klassischen Ausgleichung nach kleinsten Quadraten (L2-Norm) wird um einen Störungsparametervektor erweitert, der auf Signifikanz getestet wird. Die Berechnung folgt einem Iterationsschema:

1. Herkömmliche Ausgleichung nach kleinsten Quadraten (GMM)
2. Test auf Gültigkeit des stochastischen Modells durch den Globaltest. Bei Annahme der Nullhypothese des Globaltests endet die Prozedur.
3. Berechnung sogenannter normierter Verbesserungen NV
4. Überprüfung, ob die größte normierte Verbesserung einen gewählten kritischen Wert NVmax überschreitet. Ist das der Fall, wird der Messwert mit der größten normierten Verbesserung als grober Fehler verdächtigt und aus dem Datensatz gestrichen. Die Prozedur wird ab 1 wiederholt.
5. Die Prozedur endet, wenn keine groben Fehler mehr identifizierbar sind.

Normierte Verbesserungen werden durch

$$NV_\mathrm{i} = \frac{|v_\mathrm{i}|\sqrt{p_\mathrm{i}}}{\sigma_\mathbf{0}\sqrt{r_\mathrm{i}}} \tag{1a}$$

aus dem a-priori-Varianzfaktor σ_0^2, den Verbesserungen (Residuen) v_i, den Gewichten p_i und den Redundanzanteilen r_i des GMM berechnet.
Dieses Verfahren hat den Nachteil, dass grobe Fehler oft nicht gefunden werden, da sie durch die Ausgleichung nach kleinsten Quadraten auf mehrere Verbesserungen „verschmiert“ werden. Außerdem ist der Iterationszyklus bei mehr als einem groben Fehler theoretisch nicht gut begründet.

1.2 Ausreißertest nach Pope

Ist der a-priori-Varianzfaktor σ_0^2 nicht oder nicht zuverlässig bekannt, dann kann nach Pope (1976) alternativ mit dem empirischen (a-posteriori-) Varianzfaktor $\hat{\sigma}_\mathbf{0}^\mathbf{2}$ gearbeitet werden. Man gelangt dann zu sogenannten studentisierten Verbesserungen (Residuen):

$$SV_\mathrm{i} = \frac{|v_\mathrm{i}|\sqrt{p_\mathrm{i}}}{\hat{\sigma}_\mathbf{0}\sqrt{r_\mathrm{i}}} \tag{1b}$$

Der Ausreißertest nach Pope sollte nur verwendet werden, wenn kein Data Snooping nach Baarda möglich ist, denn er ist weniger zuverlässig (vgl. z.B. Hekimoğlu und Koch 2000).

1.3 Robuste Verfahren

Es kommen Ausgleichungsverfahren zum Einsatz, welche eine bestimmte Anzahl von groben Fehlern tolerieren, ohne das Ergebnis zu verfälschen. Am wichtigsten ist die große Klasse der M-Schätzer (HUBER 1981).
Robuste Verfahren

1. liefern grobe Fehler viel treffsicherer, da es anders als bei kleinsten Quadraten allenfalls einen geringen Verschmierungseffekt gibt (HEKIMOĞLU 2005a).
2. liefern bei Abwesenheit von groben Fehlern und auch sonst vorliegenden Voraussetzungen des klassischen Ausgleichungsmodells nicht die optimale Lösung.
3. sind an das spezielle Problem etwas anpassungsbedürftig und verlangen mehr Erfahrung bei der Wahl von Optionsparametern. Der Neuling auf dem Gebiet der Robusten Verfahren wird nicht selten zunächst völlig unplausible Ergebnisse erhalten.
4. sind nicht überall implementiert, meist nur bei hochwertigen oder wissenschaftlichen Softwareprodukten, etwa zur geodätischen Netzberechnung. Die robuste Schätzung nach Huber durch iterative Regewichtung ist aber bequem in jedes Ausgleichungsprogramm zu integrieren (siehe hierzu JÄGER u.a. 2005).

Der Autor würde es sehr begrüßen, wenn robuste Verfahren in den nächsten Jahren weiter an Bedeutung gewinnen würden. Leider gibt es in den Curricula der einschlägigen Hochschul-Studiengänge immer weniger Raum für die Ausbildung in Geodätischer Ausgleichungsrechnung, und die breite Masse der Anwender wird sich nicht ohne fundierte Kenntnisse auf das Gebiet der Robusten Verfahren vorwagen. Deshalb betrachten wir im Weiteren diese Verfahren nicht.
Sämtliche Strategien und Verfahren zur Behandlung von groben Abweichungen und Ausreißern müssen einen immanenten Zielkonflikt lösen: Sie müssen verhindern, dass grobe Fehler das Ergebnis verfälschen. Gleichzeitig sollen möglichst keine brauchbaren Messdaten verloren gehen, oder es soll im Fall, dass tatsächlich keine groben Fehler vorliegen, nicht zu weit von der optimalen Lösung abgewichen werden.
Beim Data Snooping wird das Optimum zwischen beiden Fällen durch den richtigen kritischen Wert NVmax für normierte Verbesserungen angenommen: Wird dieser Wert zu klein gewählt, werden brauchbaren Messdaten verloren gehen. Praktisch wird dieser Wert aber nicht ergebnisoptimiert eingestellt, sondern durch Vorgabe einer Irrtumswahrscheinlichkeit für die Nullhypothese „Es ist kein grober Fehler vorhanden“ beim Signifikanztest des Störungsvektors. Oder es wird von allgemeinen Erfahrungen ausgegangen. In der Geodätischen Literatur wird meist angegeben:

$2{,}5 < \mathrm{NV} < 4{,}0$: grober Fehler ist möglich
$\mathrm{NV} \geq 4{,}0$: grober Fehler ist wahrscheinlich

Offen bleibt meist, ob auch mögliche grobe Fehler eliminiert werden sollen, oder nur wahrscheinliche.
Im Folgenden wählen wir einen anderen Zugang: Wir versuchen, ausgehend von einem stochastischen Modell der groben Fehler, den kritischen Wert NVmax so zu wählen, dass die Varianz der geschätzten Parameter des GMM minimal wird.

2 Messabweichungen und Ausreißer

Ausreißer (engl.: outlier) sind extreme Merkmalswerte in Messdatensätzen, also Messwerte, die im Ausgleichungsmodell nicht zu den anderen Messwerten des Datensatzes zu passen scheinen. Ausreißer legen den Schluss nahe, dass sie durch grobe Messabweichungen entstanden sind. Eine Messabweichung gilt als grob, wenn sie bei ordnungsgemäßer Funktionsweise der Messgeräte und Arbeitsweise des Messenden prinzipiell vermeidbar gewesen wäre. Man kann sie als echten Fehler im Sinne von „Irrtum“ ansehen, weshalb hier die Bezeichnung „grober Fehler“ (engl.: gross error) erlaubt ist. Der Kürze wegen sprechen wir im Weiteren von „groben Fehlern“. Andere Messabweichungen sind keine Irrtümer und damit in diesem engeren Sinne auch keine Fehler, selbst wenn sie früher so bezeichnet wurden und auch heute noch teilweise werden.
Andere Autoren (z.B. JÄGER u.a. 2005) sprechen allgemein von „Datenstörungen“, worunter alle Effekte verstanden werden können, denen im stochastischen Modell der Ausgleichung nicht Rechnung getragen wurde.
Klassische Beispiele für grobe Fehler sind das falsche Notieren von Zahlen, das Verwechseln von Zahlen oder Punkten sowie die falsche Handhabung der Geräte. Bei digitalem Datenfluss entfallen zumindest einige Fehlerquellen, jedoch kommen neue hinzu:

- Modernen Messverfahren wohnen bisher nicht gekannte Quellen für grobe Fehler inne. **Beispiel:** Bei der reflektorlosen Distanzmessung in Tachymetern und Laserscannern kommt es häufig zu Fehlreflexionen. So ist ein bestimmter Prozentsatz grober Fehler in einer mit Laserscanner erfassten Punktwolke selbstverständlich. Hier ist schon fast der Grundsatz der Vermeidbarkeit strittig. Die Beseitigung dieser Fehler ist letztlich ein routinemäßiger Vorgang bei der Auswertung.
- Neben den durch dass Messverfahren verursachten Fehlern müssen bei hochautomatisierten Messgeräten Fehler der Instrumentensoftware in Betracht gezogen werden. Diese sollten als prinzipiell vermeidbar und damit grob gelten. Solche Fehler können bei der wachsenden Komplexität und den immer kürzeren Produktzyklen kaum noch ausgeschlossen werden. Man veranschauliche sich dazu folgenden Sachverhalt: Ein Messwert durchlaufe M Verarbeitungsschritte, in denen jeweils ein grober Fehler mit den sehr geringen Wahrscheinlichkeiten P_1, P_2, P_3, …, P_M auftritt. Die Wahrscheinlichkeit für einen groben Fehler im Endergebnis beträgt bei Unabhängigkeit der Ereignisse

$$P = 1 - (1 - P_1) \cdot (1 - P_2) \ldots (1 - P_M) \approx P_1 + P_2 + \ldots + P_M \geq M \cdot \min(P_i)$$

Das bedeutet, die Wahrscheinlichkeit grober Fehler wächst etwa linear mit der Anzahl der Verarbeitungsschritte M an.
Nicht jeder Ausreißer ist durch grobe Fehler verursacht worden. Auch zufällige Messabweichungen können zu extremen Merkmalswerten führen. Wenn man diese wie üblich als stetige Zufallsgrößen ansieht und eine Normal-

verteilung unterstellt, können Abweichungen vom Erwartungswert von mehr als dem Dreifachen der Standardabweichung bekanntlich mit einer Wahrscheinlichkeit von 0,0026 auftreten. Das erscheint zunächst relativ unwahrscheinlich, hat man jedoch sehr viele normalverteilte Messungen vorgenommen, so beträgt die Wahrscheinlichkeit P, dass keine zufällige Abweichung über dem Dreifachen der Standardabweichung liegt (Poisson-Prozess)

$$P = \exp(-N \cdot 0{,}0026)$$

mit folgenden Werten

N	20	50	100	200	500	1000
P	0,95	0,88	0,77	0,59	0,27	0,07

Ab 300 Messwerte sind zufällige Ausreißer von mehr als dem Dreifachen der Standardabweichung somit wahrscheinlicher als 50 %!
In empirischen Untersuchungen hat sich sogar gezeigt, dass wirkliche Fehlerverteilungen langsamer als die Normalverteilung abklingen, wonach noch häufiger mit zufälligen extremen Merkmalswerten zu rechnen ist. Diese Erkenntnis findet sich wahrscheinlich erstmalig bei Bessel (1818). Als Maß hierfür wird in der Statistik der Exzess (Wölbung, Kurtosis) γ_2 eingeführt:

$$\gamma_2 = \frac{m_4}{\sigma^4} - 3 \qquad (2)$$

Hierbei ist m_4 das vierte zentrale Moment der Verteilung und σ die zugehörige Standardabweichung. Für die Normalverteilung gilt $\gamma_2 = 0$. Verteilungen mit $\gamma_2 > 0$ heißen leptokurtisch, mit $\gamma_2 < 0$ hingegen platykurtisch. Verteilungen mit vielen extremen Merkmalswerten sind leptokurtisch.
Für allgemeine Wahrscheinlichkeitsverteilungen kann die Tschebyschow-Ungleichung zur Anwendung gebracht werden, nach der für eine Zufallsgröße X mit endlichem Erwartungswert μ und endlicher Varianz σ^2 allgemein nur gilt:

$$P(|x - \mu| \leq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Mit $\varepsilon = 3\sigma$ erhält man, dass Abweichungen $|x - \mu|$ von mehr als $3\,\sigma$ mit einer Wahrscheinlichkeit von bis zu $1/9 = 0{,}11$ auftreten können. Bei N Messwerten wird die Wahrscheinlichkeit, dass keine Abweichung über dem Dreifachen der Standardabweichung auftritt, nach unten begrenzt durch $P_{\min} = \exp(-N/9)$ mit den Werten

N	20	30	40	50
$P_{\min}$	0,108	0,036	0,012	0,004

Bereits ab 27 Messwerte können bei leptokurtischer Wahrscheinlichkeitsverteilung der Messabweichungen zufällige Ausreißer von mehr als $3\,\sigma$ somit wahrscheinlicher als 95 % sein! Andererseits verursacht nicht jeder grobe Fehler einen Ausreißer. **Beispiel:** Eine Verwechslung von Schräg- und Horizontaldistanz ist zwar ein grober Fehler, wird jedoch bei topographischen Aufnahmen im ebenen Gelände meist nicht zu extremen Merkmalswerten führen.

3 Modellierung grober Fehler

Es scheint zunächst, als könnte ein grober Fehler durch seine (in gewissem Sinne) Nicht-Zufälligkeit keiner Behandlung mit Werkzeugen der Wahrscheinlichkeitsrechnungen und Mathematischen Statistik zugeführt werden. Das Gegenteil ist der Fall: Auch das Eintreten eines groben Fehlers kann unter gewissen Voraussetzungen als zufälliges Ereignis aufgefasst werden. Genau genommen müssen zwei zufällige Ereignisse gleichzeitig eintreten:

1. Ein Messwert wird grob verfälscht. Die Wahrscheinlichkeitsverteilung des Messwertes hängt von der Natur des groben Fehlers ab und kann oft gut durch eine Gleichverteilung beschrieben werden. **Beispiel:** Eine Fehlreflexion bei elektronischer Distanzmessung zu einem Objekt in der Distanz D durch ein zufällig zum Zeitpunkt der Messung im Strahlengang befindliches Störobjekt verursacht gleichverteilte grobe Fehler im Intervall $[-D,0]$.
2. Ein unplausibler Messwert wird während der Messung oder Datenvorverarbeitung nicht sofort erkannt, z.B. durch eine automatische Plausibilitätsprüfung, verbleibt also im Datensatz. Betragsmäßig große grobe Fehler werden wahrscheinlicher erkannt als kleine. Sinnvoll sind hier die Laplace-Verteilung oder Normalverteilung mit sehr großer Standardabweichung. **Beispiel (Fortsetzung)**: Ein grober Fehler einer einzelnen Distanzmessung durch Reflexion an einem Hindernis im Strahlengang fällt dadurch auf, dass der erhaltene Messwert zu weit vom visuellen Schätzwert oder Näherungswert abweicht. Damit erhält man grobe Fehler mit einer im Intervall im Intervall $[-D,0]$ stetig anwachsenden Verteilung.

Enthält ein Messergebnis mit einer Wahrscheinlichkeit p einen groben Fehler X_g der Normalverteilung $N(0, \sigma_g^2)$, mit der Wahrscheinlichkeit 1-p trete kein grober Fehler auf (also $X_g = 0$), so ergibt für die Zufallsgröße „grober Fehler“ X_g insgesamt eine Wahrscheinlichkeitsverteilung

$$P(X_g < \varepsilon) = \begin{cases} p\Phi\left(\dfrac{\varepsilon}{\sigma_g}\right) & \text{für } \varepsilon \leq 0 \\ p\Phi\left(\dfrac{\varepsilon}{\sigma_g}\right) + (1-p) & \text{für } \varepsilon > 0 \end{cases}$$

Φ ist die Gaußsche Fehlerfunktion. Daraus erhält man die Verteilungsdichtefunktion

$$f_g(\varepsilon) = \frac{p}{\sigma_g\sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\sigma_g^2}\right) + (1-p)\delta(\varepsilon) \qquad (3)$$

Hierbei ist δ die Dirac-Distribution, die man sich als Grenzverteilung einer Normalverteilung mit $\sigma \to 0$ vorstellen kann. Wegen der Unstetigkeit der Verteilungsfunktion bei $\varepsilon = 0$ ist im klassischen Sinn keine Ableitung definiert. Die Dirac-Distribution steht also hier für den Fall, dass kein grober Fehler im Messergebnis enthalten ist. Nun summieren wir grobe Messabweichungen X_g und zufällige Messabweichungen X_z der Verteilung $N(0, \sigma_z^2)$. Die Verteilungsdichtefunktion der Summe $X_g + X_z$ zweier unabhängiger stetiger Zufallsgrößen wird als Faltung der einzelnen Verteilungsdichtefunktionen erhalten:

$$f(\varepsilon) = \int_{-\infty}^{\infty} f_g(\varepsilon') f_z(\varepsilon - \varepsilon') \mathrm{d}\varepsilon' =$$

$$\int_{-\infty}^{\infty} \left(\frac{p}{\sigma_g \sqrt{2\pi}} \exp\left(-\frac{\varepsilon'^2}{2\sigma_g^2} \right) + (1-p)\delta(\varepsilon') \right)$$

$$\frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{(\varepsilon - \varepsilon')^2}{2\sigma_g^2} \right) \mathrm{d}\varepsilon'$$

Die Dirac-Distribution ist das neutrale Element der Faltung. Die Faltung zweier Verteilungsdichtefunktionen der Normalverteilung ist wiederum eine solche. Damit erhält man für die Gesamtmessabweichung als Summe grober und zufälliger Abweichungen folgende Verteilungsdichtefunktion:

$$f(\varepsilon) = p \int_{-\infty}^{\infty} \frac{\mathbf{1}}{\sigma_g \sqrt{\mathbf{2}\pi}} \exp\left(-\frac{\varepsilon'^2}{2\sigma_g^2} \right) \frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{(\varepsilon - \varepsilon')^2}{2\sigma_z^2} \right) \mathrm{d}\varepsilon'$$

$$+(1-p) \int_{-\infty}^{\infty} \delta(\varepsilon') \frac{1}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{(\varepsilon - \varepsilon')^2}{2\sigma_z^2} \right) \mathrm{d}\varepsilon' \qquad (4)$$

$$= \frac{p}{\sqrt{\mathbf{2}\pi(\sigma_z^2 + \sigma_g^2)}} \exp\left(-\frac{\varepsilon^2}{2(\sigma_z^2 + \sigma_g^2)} \right) + \frac{1-p}{\sigma_z \sqrt{2\pi}} \exp\left(-\frac{\varepsilon^2}{2\sigma_z^2} \right)$$

Hierbei handelt es sich um eine sogenannte skalenkontaminierte Normalverteilung mit verschwindendem Erwartungswert, die Varianz beträgt

$$\sigma^2 = p(\sigma_z^2 + \sigma_g^2) + (1-p)\sigma_z^2 = \sigma_z^2 + p\sigma_g^2$$

Der Exzess (2) beträgt

$$\gamma_2 = \frac{3p(1-p)}{\frac{\sigma_z^4}{\sigma_g^4} + 2p\frac{\sigma_z^2}{\sigma_g^2} + p^2}$$

und ist offensichtlich positiv, außer wenn keine groben Fehler auftreten können ($p = 0$) oder alle Messwerte sicher grob falsch sind ($p = 1$). In den relevanten Fällen ist die Verteilungsdichtefunktion also leptokurtisch.
Eine Alternative zu (4) ist die zusammengesetzte Dichtefunktion nach Huber (1981):

$$f(\varepsilon) = C(c,\sigma) \begin{cases} \exp\left(-\frac{\varepsilon^2}{2\sigma^2} \right) & \text{für} \quad |\varepsilon| < c \\ \exp\left(\frac{c^2 - 2c|\varepsilon|}{2\sigma^2} \right) & \text{für} \quad |\varepsilon| \geq c \end{cases} \qquad (5)$$

$C(c,\sigma)$ ist eine Normierungsfunktion und c ein Parameter, der zwischen der reinen Laplace-Verteilung c = 0 und der reinen Normalverteilung c → ∞ vermittelt. Varianz und Exzess müssen hier numerisch ermittelt werden.
Für entsprechende Wahl der Parameter unterscheiden sich (4) und (5) nur darin, dass (4) für $|\varepsilon| \to \infty$ letztendlich langsamer abklingt. Abbildung 1 veranschaulicht diese Dichtefunktionen. Die skalenkontaminierte Normalverteilung würde sich mit dem entsprechend kleinen Parameter p oft besser als die Normalverteilung als Geodätische Fehlerverteilung eignen. Bekanntlich setzt das klassische

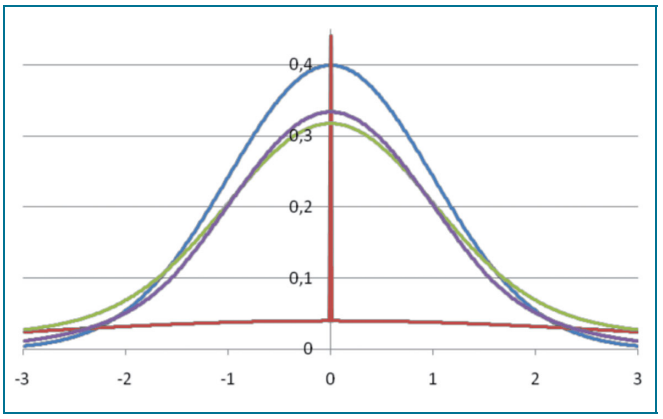

Abb. 1: Dichtefunktionen der standardisierten Normalverteilung (blau) und der Verteilung grober Fehler (3), (rot) mit p = 0,3; σg = 3, der kontaminierten Normalverteilung der Summe (4), (grün) mit p = 0,3; σz = 1; σg = 3 und der Huberschen zusammengesetzten Verteilung (5), (violett) mit c = 1,5.

Gauß-Markov-Modell (GMM) aber nicht explizit eine Fehlerverteilung voraus, wonach die Ergebnisse der klassischen Ausgleichung sich durch die Kontaminierung gar nicht ändern. Die Ursache für dieses auf den ersten Blick überraschende Verhalten ist, dass im GMM nur die beste Lösung unter allen *linearen* Schätzfunktionen gesucht wird, und die Lösung des Data Snooping und anderer robuster Verfahren *nichtlinear* von den Eingangsdaten abhängt.

4 Data Snooping bei Messreihen – Das Verfahren von Nalimov

Zunächst beschränken wir uns auf Messreihen (direkte Messungen)

$$x_1, x_2, \ldots, x_n$$

wie sie in vielen messenden Disziplinen auftreten. Zu schätzen ist der wahre Wert der zu messenden Größe x. Enthalten die Messwerte grobe Fehler, so ist für $n > 2$ das arithmetische Mittel $\bar{x}$ der Messwerte im Allgemeinen nicht die optimale Schätzung (Lehmann 2008). In vielen Ingenieur- und Naturwissenschaften haben sich Verfahren entwickelt, die den Geodätischen Verfahren z.T. verwandt sind:

Verfahren von Nalimov (1963):

1. Wähle eine Irrtumswahrscheinlichkeit α für den statistischen Test
2. Berechne das arithmetische Mittel $\bar{x}$ und die empirische Standardabweichung s_x aller Messwerte $x_1, x_2 \ldots, x_n$
3. Berechne $\Delta = \max |x_i - \bar{x}|$
4. Berechne die Testgröße $T = \Delta / s_x$
5. Falls T einen α-abhängigen kritischen Wert überschreitet, wird der Wert x_i, für den das Maximum in 3 angenommen wird, aus der Messreihe gestrichen, n: = n − 1 gesetzt und die Prozedur ab 2 wiederholt.
6. Das arithmetische Mittel der verbleibenden Elemente der Messreihe ist die gesuchte Schätzung.

Das Verfahren ist ein Spezialfall des Ausreißertests nach Pope (1976), denn T ist proportional zur studentisierten Verbesserung SV in (1b), also mit dem Data Snooping verwandt. Nach Nalimov wird kein Globaltest durchgeführt.

5 Untersuchungen mit der Monte-Carlo-Methode – Fall 1: Messreihen

Nichtlineare Schätzfunktionen wie die des Data Snooping können selten mit analytischen Methoden untersucht werden. Noch nicht einmal kann für die Schätzfunktion (endgültiger Modellparametervektor als Funktion **aller** Messwerte) eine geschlossene analytische Formel angegeben werden, selbst nicht im einfachsten Fall der Messreihe.
Die Monte-Carlo-Methode ist ein numerisches Verfahren, das in der mathematischen Statistik eingesetzt wird, um Eigenschaften von Stichprobenfunktionen (hier die Schätzfunktionen), die sich analytisch nicht ermitteln lassen, aus einer großen Anzahl von pseudozufällig generierten Stichproben genähert bestimmen zu können. In der Geodätischen Literatur finden sich zahlreiche Anwendungen (z.B. Lehmann 1994, Koch 2007, Alkhatib u.a 2009).
Die Erzeugung skalenkontaminiert normalverteilter Pseudozufallszahlen P ist problemlos aus zwei standardnormalverteilten Pseudozufallszahlen P_{n1} und P_{n2} und einer Bernoulli-verteilten Pseudozufallszahl $P_b(p)$ möglich, die sich wiederum leicht aus einer gleichverteilten Pseudozufallszahl generieren lässt:

$$P = \sigma_z P_{n1} + \sigma_g P_{n2} P_b(p) \tag{6}$$

Diese Formel vollzieht auf der Ebene der Pseudozufallszahlen die Generierung der Messabweichung in der Praxis nach: $P_b(p) = 1/P_b(p) = 0$ bedeutet „Ein grober Fehler liegt vor/nicht vor." $\sigma_z P_{n1}$ ist der zufällige Anteil und $\sigma_g P_{n2}$ der grobe Anteil der Messabweichung. Am Beispiel einer Messreihe mit skalenkontaminiert normalverteilten Messabweichungen soll die Wirkung der einschlägigen Methoden untersucht werden. Wir wählen eine Messreihe mit folgenden Eigenschaften:

Anzahl n der Messwerte je Messreihe	10			
Verteilung der Messabweichungen	identisch skalenkontaminiert normalverteilt			
Anzahl der Monte-Carlo-Experimente	50 000 je untersuchtes Szenario			
Untersuchte Szenarien des Auftretens grober Fehler	wenige große	viele große	viele kleine	keine
Wahrscheinlichkeit p eines groben Fehlers	0,01	0,1	0,1	0
Verhältnis der Standardabweichungen $q = \sigma_g/\sigma_z$	10	10	3	

Einen Globaltest führen wir nicht durch, bzw. unterstellen, dass die Hypothese generell abgelehnt wird. Damit soll vermieden werden, dass die spezielle Wahl einer Irrtumswahrscheinlichkeit für den Globaltest die weiteren Ergebnisse beeinflusst. Wir vergleichen folgende Methoden bzw. Schätzfunktionen:

1. Das einfache arithmetische Mittel aller 10 Messwerte wird gebildet.
2. Der Messwert mit der größten Verbesserung $\max|\mathbf{x_i} - \bar{\mathbf{x}}|$ wird unabhängig von seiner Größe eliminiert, dann das einfache arithmetische Mittel der restlichen 9 Messwerte gebildet.
3. Aus den restlichen 9 Messwerten wird wieder derjenige mit der größten Verbesserung unabhängig von seiner Größe eliminiert, dann das einfache arithmetische Mittel der restlichen 8 Messwerte gebildet.
4. Aus den restlichen 8 Messwerten wird wieder derjenige mit der größten Verbesserung unabhängig von seiner Größe eliminiert, dann das einfache arithmetische Mittel der restlichen 7 Messwerte gebildet.
5. Der Median aller 10 Messwerte wird gebildet. Er entspricht der L1-Norm-Schätzung.

6. Es werden so lange Messwerte eliminiert, wie max (NV_i) größer als 2,0 ist.
7. Es werden so lange Messwerte eliminiert, wie max (NV_i) größer als 3,0 ist.
8. Es werden so lange Messwerte eliminiert, wie max (NV_i) größer als 4,0 ist.
9. Es werden so lange Messwerte eliminiert, wie max (NV_i) größer als 5,0 ist.

Anschließend wird in 6–9 jeweils das einfache arithmetische Mittel der restlichen Messwerte gebildet.

Die Qualität der Schätzfunktion $\hat{x}$ bzw. der Erfolg der Methoden bemisst sich nach der mittleren quadratischen Abweichung vom wahren Wert x^{wahr} der zu messenden Größe:

$$s_{\hat{x}} = \sqrt{\frac{1}{50000}\sum_{i=1}^{50000}[(\hat{x}_i)]_i - x_i^{wahr})^2} \qquad (7)$$

$s_{\hat{x}}$ hängt in keinem der 9 Fälle von x_i^{wahr} ab, denn würde sich x_i^{wahr} ändern, so auch alle Messwerte der Reihe i um denselben Wert und folglich genauso alle $\hat{x}_i$, so dass alle Summanden in (7) gleich bleiben (siehe hierzu auch die Diskussion in den „Schlussfolgerungen“). Einzig für die Methode 1 kann die Standardabweichung nach dem Varianzfortpflanzungsgesetz analytisch berechnet werden.

Aus Abbildung 2 kann man folgende Schlüsse ziehen:

1. Selbst bei wenigen groben Fehlern – immerhin $(1-p)^{10}$ Messreihen sind frei von groben Fehlern, das sind hier entweder 90,4 % oder 34,9 % – kann man von dem generellen (naiven) Streichen des Messwertes mit der größten normierten Verbesserung im statistischen Mittel einen Vorteil erwarten. Sind gar keine groben Fehler vorhanden, dann allerdings nicht.
2. Generell noch einen zweiten Messwert zu streichen ist nur im Szenario „viele große grobe Fehler“ von Vorteil, ein dritter auch hier nicht mehr.
3. Der Median (Methode 5) ist vor allem bei vielen großen groben Fehlern von Vorteil. Sind keine groben Fehler vorhanden, dann sollte der Median nicht angewendet werden.

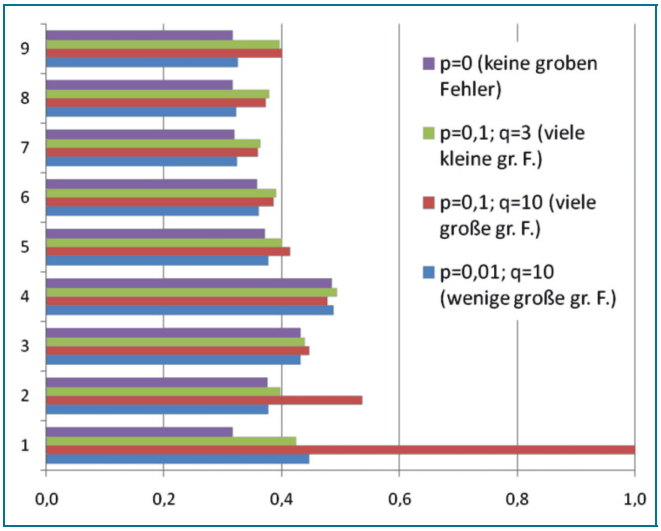

Abb. 2: Standardabweichungen $s_{\hat{x}}$ in (7) für die 9 untersuchten Schätzfunktionen/Methoden und die vier untersuchten Szenarien

4. Mit dem Data Snooping (Methode 6 bis 9) werden, falls grobe Fehler vorhanden sind, die besten Ergebnisse erreicht. Ein Optimum erscheint bei einer Schranke von ca. NVmax ≈ 3 zu liegen. Eine Interpolation liefert ein Minimum für die Standardabweichung der Schätzfunktion bei etwa 3,1 bis 3,2.
5. Das Optimum ist bei wenigen groben Fehlern nur relativ schwach ausgeprägt. Eine genauere Bestimmung des optimalen NVmax-Wertes ist also entbehrlich.
6. Das Data Snooping führt bei großen groben Fehlern zu genaueren Ergebnissen als bei kleinen, da diese besser erkannt werden können.

Die besten Ergebnisse der Methode 8 weichen nur wenig von dem theoretisch besten Ergebnis, wenn genau alle groben Fehler eliminiert würden, ab. (Die Anzahl der groben Fehler je Messreihe unterliegt einer Binomialverteilung.)

	Theoretisch bester Wert	Methode 7 (NVmax = 3,0)
Wenige große gr. F.	0,318	0,323
Viele große gr. F.	0,335	0,359
Viele kleine gr. F.	0,335	0,364
Keine groben Fehler	0,316	0,320

Schließlich geben wir noch an, wieviele falsche Entscheidungen beim Hypothesentest auf grobe Fehler die einzelnen Methoden liefern. Ein grob falscher Messwert ist durch $P_b(p) = 1$ in (6) gekennzeichnet.

Fehler erster Art:

Ein nicht grob falscher Messwert wird gestrichen. Hier ist der Anteil bei allen 4 Szenarien etwa gleich und liegt bei

NVmax	5,0	4,0	3,0	2,0
Anteil [%]	0,000	0,008	0,28	4,0

Fehler zweiter Art:

Ein grob falscher Messwert wird nicht gestrichen.

NVmax	5,0	4,0	3,0	2,0
Anteil [%]				
Wenige große gr. F.	40	32	25	17
Viele große gr. F.	41	33	25	18
Viele kleine gr. F.	90	82	69	51

Obwohl bei NVmax = 3,0 noch immer eine beträchtliche Anzahl grober Fehler im Datensatz verbleibt, sind geringere NVmax-Werte offenbar ungünstig, denn die negativen Effekte der steigenden Anzahl von Fehlern erster Art überwiegen im statistischen Mittel den Gewinn.

6 Untersuchungen mit der Monte-Carlo-Methode – Fall 2: Ausgleichende Gerade

Ein abschließendes Beispiel soll die Darlegungen weiter illustrieren. Wir betrachten den Fall einer ausgleichenden Gerade

$$y_i = at_i + b$$

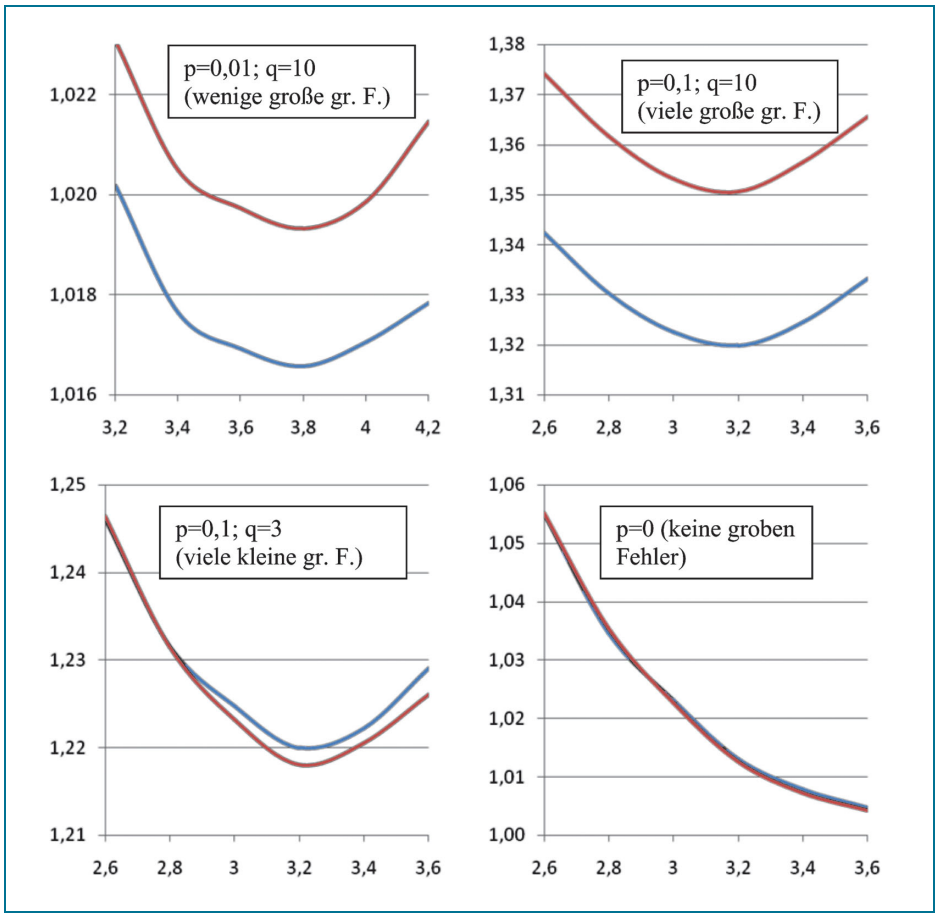

Abb. 3: relative Standardabweichungen $\frac{s_{\hat{a}}}{s_{\hat{a},0}}$ (blau) und $\frac{s_{\hat{b}}}{s_{\hat{b},0}}$ (rot) für die vier untersuchten Szenarien (Ordinaten) in Abhängigkeit von NVmax (Abszissen)

mit n fehlerfreien äquidistanten Abszissen t_i, der dem Vermessungspraktiker etwa von der Kalibrierung elektronischer Distanzmesser bekannt ist. Die Parameter a,b müssen für $n > 2$ ausgeglichen werden.
Wir wählen eine ausgleichende Gerade mit folgenden Eigenschaften:

Anzahl n der Messwerte	10: $t_1 = 1, \ldots, t_{10} = 10$			
Verteilung der Messabweichungen	identisch skalenkontaminiert normalverteilt			
Anzahl der Monte-Carlo-Experimente	50 000 je untersuchtes Szenario			
Untersuchte Szenarien des Auftretens grober Fehler	wenige große	viele große	viele kleine	keine
Wahrscheinlichkeit p eines groben Fehlers	0,01	0,1	0,1	0
Verhältnis der Standardabweichungen $q = \sigma_g/\sigma_z$	10	10	3	

Indem für jeden generierten Datensatz fortlaufend der Messwert mit der größten normierten Verbesserung NV gestrichen wird, erzeugt man eine Sequenz von Parametervektoren $(\hat{a}, \hat{b})$, die zusammen mit dem jeweils größten NV-Wert in einer Tabelle abgespeichert wird. In dieser Tabelle kann man sehr schnell für unterschiedliche NVmax-Werte die zugehörigen Lösungen aufsuchen und somit die Standardabweichungen der Parameter wie in (7), wobei x durch a bzw. b ersetzt wird, berechnen. Zur übersichtlichen Darstellung der Ergebnisse in Abbildung 3 beziehen wir diese Werte auf die entsprechenden Standardabweichungen $s_{\hat{a},\mathbf{0}}$ und $s_{\hat{b},\mathbf{0}}$, die man ohne grobe Fehler (d.h. $p = 0$) und ohne Data Snooping (d.h. NVmax $\to \infty$) erhalten hätte.
In diesen Diagrammen erkennt man, dass es für $p > 0$ ein wohl definiertes Optimum für NVmax gibt, welches für beide Modellparameter a,b etwa identisch ist. Es kann nicht ausgeschlossen werden, dass in bestimmten Modellen das Optimum für die einzelnen Parameter differiert. Hier muss ein Kompromiss gefunden werden, indem eine skalare Zielfunktion definiert wird. Das ist nicht eindeutig möglich und hängt davon ab, welcher Teil des Ausgleichungsmodells in der jeweiligen Anwendung besonders wichtig ist. Diese Zielfunktion könnte im Fall der ausgleichenden Gerade eine zu prädizierende Ordinate $y(t)$ oder die in einem interessierenden Intervall $[t_a, t_b]$ am ungenauesten prädizierbare Ordinate sein.
Weiter wird deutlich, dass das Optimum von den stochastischen Eigenschaften der Messwerte abhängt. Diese Eigenschaften müssen also mindestens genähert bekannt sein, um das richtige Optimum zu erhalten.
Sind grobe Fehler auszuschließen, so sollte kein Data Snooping durchgeführt werden, egal wie groß die NV-Werte sind (d.h. NVmax $\to \infty$).

7 Vergleich mit Hekimoğlus Verfahren

In (HEKIMOĞLU 2005) wurde vorgeschlagen, die Zuverlässigkeit des Data Snooping zu steigern, indem alle Gewichte mit einem Faktor $k > 1$ multipliziert werden. Als Maß für Zuverlässigkeit gilt die Definition nach Hekimoglu und Koch (2000). Die Teststatistik des Data Snooping wird dann um den Faktor größer. Dasselbe hätte man erreicht, wenn man die Irrtumswahrscheinlichkeit entsprechend geändert hätte, so dass NVmax um denselben Faktor kleiner wird. Wie in diesem Beitrag gezeigt wird, hat die Wahl des NVmax-Wertes natürlich einen Einfluss auf die Zuverlässigkeit des Data Snooping. Es wäre Zufall, wenn das Optimum bei $k = 1$ läge.
Die Unterschiede zwischen Hekimoglus Verfahren und der hier vorgeschlagenen Methode liegt darin, dass

1. die mehr oder weniger willkürliche Wahl einer Irrtumswahrscheinlichkeit entfällt,
2. als Erfolgskriterium nicht die Zuverlässigkeit der Aufdeckung grober Fehler gilt, diese ist nur Mittel zum Zweck, sondern die Genauigkeit der letztlich gesuchten Größen (Parameter oder Funktionen derselben), und
3. mit der Monte-Carlo-Methode ein transparentes und standardisiertes Verfahren eingeführt wird, welches auch dem Praktiker zur Bestimmung seines optimalen NVmax-Wertes empfohlen werden kann.

8 Rechenaufwand

Ein gewisser Nachteil des NVmax-Optimierungsverfahren kann im Rechenaufwand gesehen werden. Es müssen mehrere Ausgleichungsberechnungen mit Inversion der Normalgleichungsmatrix durchgeführt werden. Diese ändert sich auch ständig, je nachdem, welche Messwerte zuvor gestrichen wurden.
Beispiel: 100 Messwerte mit 30 Parametern sollen ausgeglichen werden. Maximal 5 Messwerte können gestrichen werden, ohne dass eine Neumessung erfolgen muss. Im ungünstigsten Fall ergeben sich

$$\binom{100}{0} + \binom{100}{1} + \ldots + \binom{100}{5} \approx 8 \cdot 10^7$$

zu invertierende Matrizen der Dimension 30. Mit dem Gauß-Jordan-Verfahren sind dazu je ca. 27 000 Gleitkommaoperationen notwendig. Bei einem Intel Core i7 Prozessor mit Hyperthreading kann man etwa 33 GFLOPS erwarten (VILSBECK 2008), also benötigt die Inversion etwa $8 \cdot 10^7 * 27000/(33 \cdot 10^9/\mathrm{s}) = 65\mathrm{s}$. Dazu kommen je Monte-Carlo-Experiment die Generierung von 100 Pseudozufallszahlen und eine Matrix-Vektor-Multiplikation der Matrix-Dimension (100,30), also 6000 Gleitkommaoperationen. Falls die Anzahl der Monte-Carlo-Experimente allgemein nicht höher als in den berechneten Beispielen liegen muss, so kann dieser Rechenaufwand vernachlässigt werden. Unter Ausnutzung von Rechenvorteilen lässt sich die Berechnung noch wesentlich beschleunigen.

9 Schlussfolgerungen

In diesem Beitrag untersuchen wir die Eigenschaften des iterativen Data Snooping für die Ausgleichung grob fehlerbehafteter Messwerte. Wir empfehlen dieses Verfahren für Vermessungspraktiker, die sich nicht näher mit Robusten Ausgleichungsverfahren befassen können. Das Verfahren kann mit jeder herkömmlichen Ausgleichungssoftware durchgeführt werden, wenn diese wenigstens die Redundanzanteile r_i ausgibt, besser noch direkt die normierten Verbesserungen (1a). Insbesondere auf einen bekannten Verteilungstyp von groben Fehlern zugeschnitten liefern Robuste Ausgleichungsverfahren aber meist bessere Ergebnisse.
In der Praxis kann man so vorgehen, dass man für sein funktionales und stochastisches Ausgleichungsmodell den optimalen NVmax-Wert mit der Monte-Carlo-Methode bestimmt und dann seine Messwerte auf diese Weise ausgleicht. Die Optimierung kann sogar schon vor der Messung erfolgen, denn die Messwerte werden dazu höchstens genähert benötigt. Man erkennt dies folgendermaßen: Mit angenommenen wahren Modellparametern x^{wahr} und der Designmatrix A werden wahre Messwerte $l^{\mathrm{wahr}} = A\, x^{\mathrm{wahr}}$ erzeugt und mit wahren Messabweichungen ε verfälscht. Die Verbesserungen ergeben sich mit der Redundanzmatrix R zu

$$v = -R(l^{\mathrm{wahr}} + \varepsilon) = -RA\mathrm{x}^{\mathrm{wahr}} - R\varepsilon = -R\varepsilon$$

Der erste Summand verschwindet, denn die wahren Werte erhalten keine Verbesserung, bzw. RA ist die Nullmatrix. Somit sind die Verbesserungen und die normierten Verbesserungen nur von den (hier pseudozufällig erzeugten) wahren Messabweichungen ε abhängig, nicht von den wahren Messwerten l^{wahr} selbst.
Weiter werden keine a priori Standardabweichungen und Irrtumswahrscheinlichkeiten benötigt, sondern nur ein Größenverhältnis $\sigma_\mathrm{g}/\sigma_\mathrm{z}$ und eine Wahrscheinlichkeit grober Fehler p. Diese müssen anwendungsspezifisch, z.B. aus Erfahrungen mit dem Messverfahren, abgeleitet werden. Nach den hier vorliegenden Erkenntnissen müssen beide Werte nicht besonders genau bekannt sein.
Es ist möglich, auch andere Wahrscheinlichkeitsverteilungen wie die Hubersche Verteilung (5) zu Grunde zu legen. Die zusätzliche Schwierigkeit besteht nur darin, entsprechende Pseudozufallszahlen zu generieren. Mit Standardfunktionen numerischer Softwarepakete ist das nicht möglich.
Das Verfahren kann analog für studentisierte Verbesserungen (1b) angewendet werden. Auch andere Störparametervektoren können entsprechend angesetzt werden. Bei Transformationen wird man z.B. immer einen kompletten Anschlusspunkt eliminieren wollen. Ein entsprechendes Kriterium kann dem NV-Wert analog verwendet werden.

Literatur

[1] Alkhatib, H.; Neumann, I.; Kutterer, H.: Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques. Journal of Applied Geodesy. Band 3, Heft 2, Seiten 67–79, 2009

[2] Bessel, F. W.: Fundamenta Astronomiae. Nicolovius, Königsberg 1818

[3] Haselton, M. G.; Nettle, D.; Andrews, P. W.: The evolution of cognitive bias. In D. M. Buss (Ed.), Handbook of Evolutionary Psychology, (pp. 724–746). Wiley, Hoboken 2005

[4] Hekimoğlu, Ş.; Koch, K.-R.: „How can reliability of the test for outliers be measured?“, AVN 107: 247–252, 2000

[5] Hekimoğlu, Ş.: Do Robust Methods Identify Outliers More Reliably Than Conventional Test for Outlier?, Zeitschrift für Vermessungwesen, 2005/3, 174–180, 2005a

[6] Hekimoğlu, Ş.: Increasing Reliability of Data Snooping When Outliers are Small, Allgemeine Vermessungsnachrichten, 2005/1, 7–11, 2005b

[7] Huber, P.J. (1981): Robust Statistics. Wiley, New York 1981

[8] Jäger, R.; Müller, T.; Saler, H.; Schwäble, R.: Klassische und robuste Ausgleichungsverfahren – Ein Leitfaden für Ausbildung und Praxis von Geodäten und Geoinformatikern. Herbert-Wichmann-Verlag, Heidelberg 2005

[9] Koch, K. R.: Outlier Detection in Observations Including Leverage Points by Monte Carlo Simulations. Allgemeine Vermessungsnachrichten 10/2007

[10] Lehmann, R.: Ausgleichung in nichtlinearen Modellen mittels adaptiver Monte-Carlo-Integration. Allgemeine Vermessungsnachrichten 07/1994

[11] Lehmann, R.: Das arithmetische Mittel von Mehrfachmessungen unter Wiederholbedingungen. Allgemeine Vermessungsnachrichten 01/2008

[12] Nalimov, V. V.: The Application of Mathematical Statistics to Chemical Analysis. Pergamon, Oxford 1963

[13] Pope, A. J.: The statistics of residuals and the detection of outliers. NOAA Technical Report NOS65 NGS1, US Department of Commerce, National Geodetic Survey, Rockville, Maryland 1976

[14] Vilsbeck, Ch.: Intel Core i7 mit Nehalem-Quad-Core. http://www.tecchannel.de, 2008

Anschrift des Verfassers:
Prof. Dr.-Ing. Rüdiger Lehmann, Hochschule für Technik und Wirtschaft Dresden, Fakultät Geoinformation, Friedrich-List-Platz 1, 01069 Dresden, Tel 03 51-4 62-31 46, Fax 03 51-4 62-21 91, mailto:r.lehmann@htw-dresden.de