

H.Rodemann

Scheitelwertbestimmung als eine Methode bei der Stapelung von  
-----  
Geoelektrik-Meßdaten  
-----

Press et al. (1986, p.462 f) geben ein Verfahren zur Bestimmung des Scheitelwerts bei "stetigen Daten" an. Dabei wird ein stetiges Histogramm durch ein Fenster wählbarer Länge erzeugt. Ein wesentlicher Vorteil besteht darin, daß zur berechneten Häufigkeitsverteilung eine 1-sigma-Streubreite angegeben werden kann. Wie diese Methode verwendet werden kann, um ein Fehlermaß für den Scheitelwert zu konstruieren (etwa analog zur Streuung), ist bei Rodemann (1990) beschrieben.

Damals erschien im Vergleich mit Mittelwert und Median der Scheitelwert aus dem "stetigen Histogramm" nach Press et al. (1986) besonders geeignet, um bei gestörten Daten das wahre  $\rho_{hos}$  zu finden.

Es zeigte sich jedoch bei weiteren Experimenten mit synthetischen, speziell mit Gauß-verteilten Daten, daß der zum Scheitelwert gehörige Fehler (analog zur Streuung) in der Regel viel größer als der wahre Fehler ist. Andererseits ergibt sich bei Division durch die Wurzel aus der Meßwert-Anzahl (analog zum Übergang von der Streuung zum mittleren Fehler des Mittelwerts) in der Regel ein viel zu kleiner Fehler im Vergleich zum wahren Fehler.

Da ein theoretischer Ansatz nicht vorhanden war, und auch begriffliche Schwierigkeiten auftraten (z.B. was ist der mittlere Fehler eines Einzelwerts bei einer Scheitelwertbestimmung), wurde ein experimenteller Ansatz gewählt; die Ergebnisse sind noch vorläufig:

Basierend auf dem in Turbo-Basic integrierten Random-Generator (näherungsweise gleichverteilte Zufallszahlen im Bereich 0 bis 1) war mit Hilfe des in Press et. al. (1986) beschriebenen Algorithmus ein Generator für Gauß-verteilte Zufallszahlen auf IBM-Kompatiblen PC's programmiert worden.

Mit diesem Programm wurden die verschiedenen Mittelwerte "getestet". Dabei wurden Soll-Streuung, wahrer Mittelwert und Anzahl der Daten festgelegt. Wahrer und geschätzter Fehler wurden für eine möglichst große Zahl von Wiederholungen gespeichert, wobei bisher die Zeit, und nicht der Speicherplatz, als Begrenzung wirkte. Anschließend wurde in einer Schleife nach Art einer Nullstellensuche der Korrekturfaktor am geschätzten Fehler gesucht, bei dem die wahren Fehler mit einer Häufigkeit von 68.2% kleiner als die modifizierten geschätzten Fehler waren. Die Fehler werden also an die 1-Sigma-Wahrscheinlichkeit der idealen Gauß-Verteilung angepaßt.

Ergebnisse:

Alle Werte schwanken beträchtlich. Offensichtlich sind viel größere Wiederholungszahlen (und mehr Zeit) erforderlich. Es muß die Streuung der Fehler in geeigneter Form gemessen werden. Die Qualität des Zufallsgenerators soll genauer überprüft werden.



Beim stetigen Histogramm ist der Korrekturfaktor von der Zahl der Daten und von der Fensterbreite abhängig. Bei festgehaltener Breite von 50% der Datenanzahl ergibt sich sehr näherungsweise (!) ein Korrekturfaktor von  $\text{Datenanzahl}^{-1/4}$ .

Beim Median schwanken die Werte ca zwischen 1.4 und 2 (theoretischer Wert:  $\pi/2$ ). Beim Huber-Verfahren ( $\psi(x_i)=x_i$  für  $\text{abs}(x_i)\leq a$  und  $\psi(x_i)=a\cdot\text{sgn}(x_i)$  sonst, mit  $a=1.5$ ) ergibt sich ein ähnlicher Schwankungsbereich, sodaß wie beim Median der Wert  $\pi/2$  vorläufig in das Meßprogramm eingefügt wurde.

Die Methode ist jedoch erfolgversprechend. Tests mit den so berechneten Fehlern für nicht-Gaußsche Daten zeigen eine gewisse Übertragbarkeit. Bei stark asymmetrischen Daten ist das stetige Histogramm meist am leistungsfähigsten, bei symmetrischen Daten dagegen schlechter. Um die Wahl der besten Verteilungen zu erleichtern, sind Schiefe und Kurtosis dargestellt, zusätzlich die aus Mittelwert und Streuung resultierende Gaußkurve im Vergleich zu einem einfachen Histogramm.

Die folgenden graphischen Darstellungen aus dem entwickelten Meßprogramm sollen einige Beispiele mit synthetischen Daten zeigen:

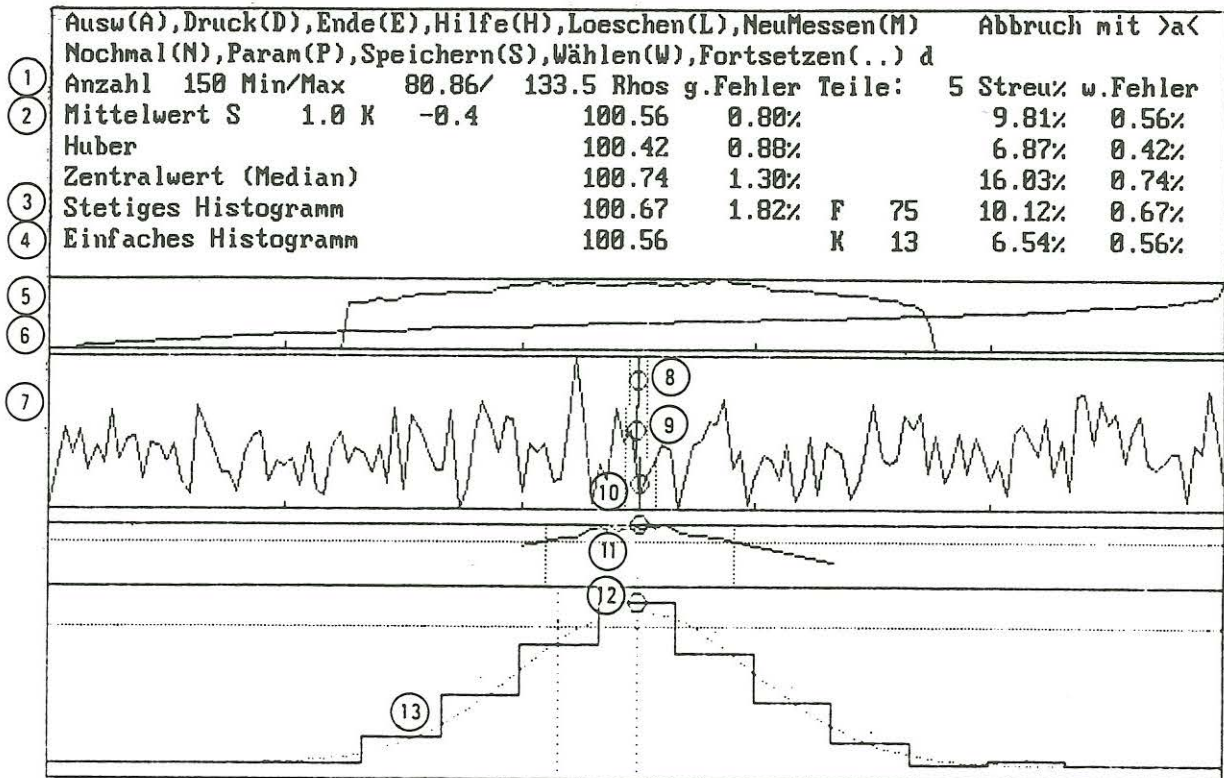
1. Gaußsche Daten
2. Asymmetrische Verteilung
3. Symmetrische, nicht Gaußsche Verteilung.

Der Autor möchte Prof. Weidelt, TU Braunschweig, Dr. Junge und Dr. Olsen, Univ. Göttingen für hilfreiche Hinweise danken.

#### Literatur:

Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling,  
Numerical Recipes, The Art of Scientific Computing,  
Cambridge University Press, Cambridge, 1986.

Rodemann, Überlegungen zum Stapeln in der Geoelektrik, in:  
Haak, V. und J. Homilius,  
Protokoll über das 13. Kolloquium  
"Erdmagnetische Tiefenforschung" in Hornburg v.  
19.3.-23.3. 1990.



- 1: Rhos : Rho-S (Ohm\*m)  
g. Fehler: geschätzter Fehler in %  
Streu% : rel. Streuung in %  
w. Fehler: wahrer Fehler in %
- 2: S : Schiefe  
K : Kurtosis
- 3: F : Fensterbreite
- 4: K : Klassenzahl
- 5: Datendichte nach Press et al (1986)
- 6: Daten nach Größe geordnet
- 7: Daten in zeitlicher Anordnung
- 8-10: Mittelwerte und g. Fehler graphisch dargestellt
- 8: Mittelwert 9: Huber 10: Median
- 11: stetiges Histogramm mit Streuung
- 12: einfaches Histogramm
- 13: Gaußkurve zum Vergleich mit dem einfachen Histogramm

Abb. 1: Ausdruck eines auf dem Bildschirm dargestellten Resultats des Stapel-Programms. Es wurden synthetische Daten verwendet, gaußverteilt, Mittelwert 100 (Ohm\*m), Streuung 10% .



Ausw(A), Druck(D), Ende(E), Hilfe(H), Loeschen(L), NeuMessen(M)						Abbruch mit >a<	
Nochmal(N), Param(P), Speichern(S), Wählen(W), Fortsetzen(..) d							
Anzahl	120	Min/Max	74.40/ 220.4	Rhos	g.Fehler	Teile:	4
Mittelwert	S	6.6	K	3.5	116.88	2.18%	23.84%
Huber					106.69	1.10%	7.66%
Zentralwert (Median)					101.43	0.71%	7.84%
Stetiges Histogramm					100.47	0.52%	F 60
Einfaches Histogramm					95.45		K 11

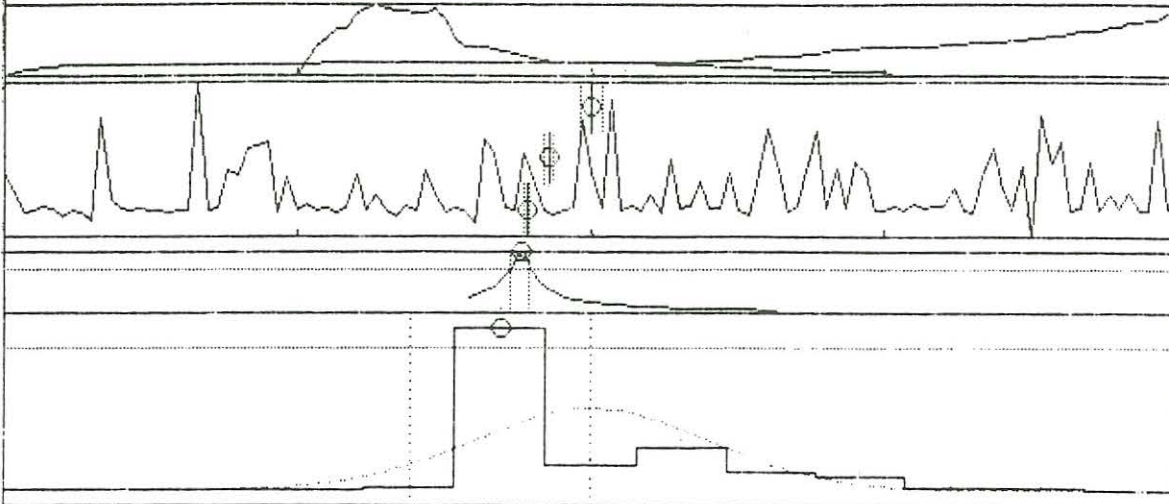


Abb. 2: Synthetische Daten, asymmetrisch verteilt.  
wahres Rho: 100 (Ohm\*m), Gaußsche (Geräte-) Streuung 2 %, Spikes gaußverteilt, mittl. Häufigkeit 50 %, mittlere Amplitude 30, zentriert um 140  
Vergl. Erläuterungen zu Abb. 1.

Ausw(A), Druck(D), Ende(E), Hilfe(H), Loeschen(L), NeuMessen(M)						Abbruch mit >a<	
Nochmal(N), Param(P), Speichern(S), Wählen(W), Fortsetzen(..) d							
Anzahl	210	Min/Max	11.35/ 182.3	Rhos	g.Fehler	Teile:	7
Mittelwert	S	-2.0	K	5.0	98.02	2.22%	32.16%
Huber					100.06	0.28%	2.57%
Zentralwert (Median)					100.08	0.41%	5.99%
Stetiges Histogramm					99.87	0.46%	F 105
Einfaches Histogramm					98.02		K 15

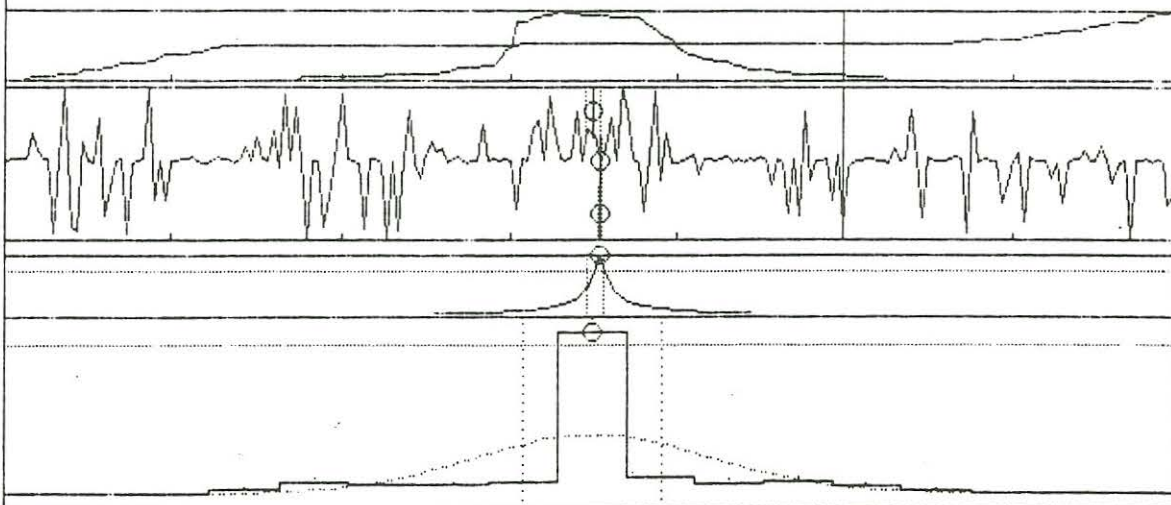


Abb. 3: Synthetische Daten, symmetrisch, nicht gauß-verteilt.  
wahres Rho: 100 (Ohm\*m), Gaußsche (Geräte-) Streuung 2 %, Spike-Amplituden gleichverteilt im Bereich 10 bis 190  
mittl. Häufigkeit 50 % .  
Vergl. Erläuterungen zu Abb. 1.