

H. Rodemann

## Überlegungen zum Stapeln in der Geoelektrik

---

Die folgenden Überlegungen beziehen sich auf die Schlumbergerauslagen und die Halbauslagen. Bei Kurzauslagen (bis zu einigen 100 m AB/2) werden die  $\rho_s$ -Werte in der Regel durch Mittelung bei kleinen Datenmengen gewonnen. Dabei wird der mittlere Fehler des Mittelwerts mit einem Sollwert verglichen. Bei Erreichen dieses Werts wird die Messung beendet. Außerdem kann auch eine Ausreißer-Erkennung integriert werden, wie es z.B. im Meßprogramm des NLFb der Fall ist.

Ein Stapeln im Sinne der Auswertung größerer Datenmengen ist in der Geoelektrik im allgemeinen nur dann erforderlich, wenn z.B. bei langen Auslagen bis 5 km AB/2 das Signal-zu-Rausch-Verhältnis zu klein wird bzw. Störspannungen die Messung zu stark beeinflussen, u.a. weil die Leistungsfähigkeit des Netzgeräts begrenzt ist.

Andererseits wird eine Begrenzung der Meßdauer auf einen Tag pro Punkt angestrebt, z.B. weil die Kabel aus Sicherheitsgründen über Nacht eingerollt werden müssen.

Als Störspannungen kommen infrage:

- Geräterauschen
- zeitlich variables Eigenpotential
- Ausreißer, Spitzen (Spikes), evtl. exponentiell abklingend
- Tellurik (Wechselspannung, variable Amplitude)

Der Gleichspannungsanteil und langperiodische Wechselspannungsanteile werden durch periodisches Umpolen unterdrückt. Die verbleibenden Störspannungsanteile hingegen sollen durch Stapelverfahren behandelt werden.

Ein Stapelverfahren sollte den eigentlichen Meßwert mit einer genügenden Genauigkeit bestimmen und dabei gegen Störsignale möglichst resistent bzw. robust sein, und im Falle Gaußverteilter Daten außerdem möglichst wenig zusätzliche Daten benötigen, also möglichst effizient sein.

CHAVE et al. (1987) unterscheiden bei robusten Verfahren

1. L1-Norm-Methoden, z.B. den Median, mit relativ hoher Resistenz gegen Ausreißer, aber niedriger Effizienz
2. Ausreißer-Erkennung und -Eliminierung und Behandlung der restlichen Daten mit der Methode der kleinsten Quadrate.
3. M-Methoden (Maximum Likelihood), mit im Vergleich zu L1-Methoden geringerer Resistenz gegen Ausreißer, aber höherer Effizienz. Diese Methoden werden von HANSTEIN in einem anderen Beitrag in diesem Band behandelt.

Den Verfahren gemeinsam ist die Suche nach einem Schätzwert für die Größe der gesuchten Meßvariablen (location) und nach dem Fehlermaß (scale). Dabei ist es einerseits schwierig, die Fehler der verschiedenen Methoden zu vergleichen. Der Vergleich der Varianzen, die aus mehreren Schätzwertbestimmungen der verschiedenen Verfahren bei Gaußverteiltern Daten resultieren, kann zur Bestimmung der Effektivität verwendet werden. Er kann auch dazu dienen, die Schätzwerte anderer Methoden durch Skalierung auf die Standard-Abweichung zu beziehen. Bei einer Kontamination der Daten führt solch ein Vergleich aber je nach Art der Störungen zu unterschiedlichen Resultaten. Andererseits sind die Methoden jedoch m.E. nicht so universell

Diese Definition ähnelt im Falle einer unimodalen Verteilung der einer Halbwertsbreite. Bei einer bi- oder multimodalen Verteilung erlaubt sie, abhängig vom Fehlerschlauch der pdf den Scheitelwert von konkurrierenden Maxima zu unterscheiden. Falls ein Maximum relativ zu einem benachbarten kleineren lokalen Maximum nicht signifikant ist, werden beide zu einem Maximum mit einer größeren Breite zusammengezogen. Ein Problem bei der Scheitelwertbestimmung ist die Wahl der Fensterbreite  $j$ . Dazu gibt es zwei Lösungsansätze:

1. Von PRESS et al. (1987) wird eine "high-brow"-Methode nach dem Maximum- Likelihood-Prinzip zur Bestimmung von  $j$  angegeben. Die Methode besitzt zwei Nachteile:

- Sie wählt in der Regel  $j > n/2$ , oft nahe am maximal möglichen Wert. Damit wird bei Gauß-ähnlichen pdf's ein guter Schätzwert erzielt, aber z.B. bei einer sehr asymmetrischen bimodalen Verteilung löst die Methode die verschiedenen Maxima nicht auf und erzielt ähnlich schlechte Resultate wie Mittelwert und evtl. auch Median.

2

- Eine zeitaufwendige Berechnung von ca.  $N$  unvollständigen Gamma-Funktionen ist erforderlich, sodaß der Programmablauf trotz Koprozessor sehr verlangsamt wird, auch wenn die Zahl durch vereinfachende Annahmen (nur ein lokales Maximum der Likelihood-Funktion) reduziert wird.

2. PRESS et al. (1987) erwähnen, daß zur Erzielung einer hohen Auflösung eine kleine Fensterbreite (minimal  $j=3$ ) erforderlich ist, mit dem Nachteil einer geringeren Genauigkeit und der möglichen Bevorzugung eines zufälligen Maximums der pdf. Hier eignet sich die Bestimmung von  $j$  durch Minimierung der oben definierten Fehlerbreite dazu, eine hohe Auflösung der Scheitelwertsuche zu erzielen und gleichzeitig eine Genauigkeitsschwelle nicht zu unterschreiten, denn im allgemeinen nimmt die Fehlerbreite bei sehr kleinem  $j$  zu, weil dann auch bei Gaußverteilten Daten mehrere lokale Maxima der pdf zu einem breiteren Maximum zusammengefaßt werden. Andererseits steigt auch bei großem  $j$  in der Regel die Fehlerbreite an, weil ein ausgeprägtes Maximum durch die starke Mittelung verbreitert wird. Bei der beschriebenen Minimierung werden außerdem Fälle ausgeschlossen, bei denen alle pdf-Werte zu einem Maximum zusammengefaßt werden und dadurch die Fehlerbreite nicht durch den Übergang zu geringeren pdf-Werten bestimmt werden kann.

Der hohe Grad an Robustheit muß ähnlich wie die Robustheit des Medians mit einer geringeren Effizienz bei Gaußverteilten Daten bezahlt werden. Während die Varianz von Median-Werten, die aus Gaußverteilten Daten gewonnen werden können, um den Faktor  $\pi/2$  größer ist als die von Standard- Mittelwerten (CHAVE et al., 1987, p.635), haben Tests mit dem beschriebenen Verfahren gezeigt, daß die Varianzen noch einmal um etwa den Faktor 1,5 bis 2 größer sind.

Die Robustheit soll am Beispiel einer ausgeprägt bimodalen Verteilung gemäß beigefügter Abbildung (Kopie des Bildschirms) gezeigt werden. Der synthetische Spannungsverlauf enthielt ein wahres Signal von 100 (z.B. mV) mit einem Gaußschen "Geräterauschen" mit einer Streuung von 3 mV und Gaußverteilten Spikes mit einer relativen Häufigkeit von 70%, einer Streuung von 200 mV und einer mittleren Spannung von 100 mV. Wie häufig solche oder andere ähnlich kontaminierte Daten im Gelände anzutreffen sind, soll bei Geländetests überprüft

werden.

Die erzielten Ergebnisse sind folgende:

	geschätzte Fehler	wahre Fehler
Mittelwert	11%	84%
Median	12%	18%
Stetiges Histo.	4%	2%

Im folgenden werden einige Details zum erwähnten Programm besprochen und die erwähnte Abbildung diskutiert.

Das Programm kann den Meßfortschritt fortlaufend numerisch und graphisch anzeigen und erlaubt, die Messung jederzeit zu unterbrechen und nach eigener Entscheidung zu beenden, wobei ein gewählter Meßwert, eventuell incl. Fehler, gespeichert wird.

Dabei werden z. Zt. die Resultate von vier Verfahren angezeigt:

1. einfacher Mittelwert + mittl. Fehler des Mittelwerts
2. Median + mittl. Fehler (Median average deviation)
3. "stetiges Histogramm"+ Fehler des Scheitelwerts n. PRESS et al. (1987)
4. einfaches Histogramm + Fehler des Scheitelwerts

zusätzlich: Vergleich des Histogramms mit der aus 1. resultierenden Gauß-Fehlerkurve. Hierdurch können Abweichungen von der Gaußverteilung besser erkannt werden als durch Angabe von Schiefe und Exzeß.

Bei Bedarf kann eine Fourier-Analyse zur Beurteilung und ggf. Beeinflussung des Frequenzgehalts durchgeführt werden.

Für die stetige Histogramm-Methode wird die Fensterbreite mit minimalem Scheitelwertfehler oder nach der Maximum-Likelihood-Methode ermittelt, analog beim einfachen Histogramm die günstigste Klassenbreite mit minimalem Fehler.

Das verwendete Fehlermaß ist dabei analog zum stetigen Histogramm-Verfahren definiert, beruht jedoch nicht auf einer gültigen Formel für den Fehlerschlauch, ist also ein "heuristisches" Maß, das allerdings sehr oft dem Fehlermaß der stetigen Histogramm-Methode entspricht. Das Histogramm ist jedoch schon deshalb sinnvoll, weil es die Häufigkeitsverteilung im Vergleich zur pdf relativ "unverfälscht" darstellt, speziell im Vergleich mit der aus Mittelwert und Streuung resultierenden Gauß-Verteilung. Fensterbreite und Klassenzahl sind auch von Hand veränderbar.

Der Text-Teil zeigt außer Zahl der Messungen und maximaler und minimaler Spannung (oder analog  $\mathcal{E}_s$ ) die Resultate der vier Verfahren mit Fehlermaß und wahren Fehler (wF), der (leider) nur bei synthetischen Daten bekannt ist, hier aber ein Maß für die Fehler der Verfahren liefert.

Der Graphik-Teil ist in drei Abschnitte gegliedert. Im oberen Abschnitt ist der Meßwerteverlauf dargestellt. Die Werte sind so normiert, daß der Abschnitt von den Werten in vertikaler (Amplituden) und horizontaler (Zeitrichtung) ausgefüllt ist. In der Mitte dieses Abschnitts sind Mittelwert mit Fehler und darunter Median mit Fehler zu sehen. Darunter schließt sich die Darstellung der pdf nach

PRESS et al (1987) "stetiges Histogramm" an. Die horizontale Linie entspricht dem pdf-Niveau

$$p_{\max} \cdot \frac{1-1/j^{1/2}}{1+1/j^{1/2}} .$$

Der untere Abschnitt zeigt das einfache Histogramm mit Wahrscheinlichkeitsniveaus. Die Klassenzahl ist relativ hoch. Dies scheint jedoch bei der Suche nach einem scharfen Maximum gerechtfertigt. Die überlagerte Gaußkurve ist in der Abbildung leider nur schlecht erkennbar.

Die vier Resultate mit Fehlermaß sind dadurch noch einmal in graphischer Form übereinander dargestellt.

Danksagung: Hiermit möchte ich Herrn Prof. Weidelt und Herrn Hanstein für wertvolle Hinweise danken.

#### Literatur:

Chave, A.D., D.J. Thomson & M.A. Ander, On the Estimation of Power Spectra, Coherences, and Transfer Functions, JGR, Vol 92, p.633f., 1987.

Hoaglin, D.C., F. Mosteller & J.W. Tukey, Understanding Robust and Exploratory Data Analysis, John Wiley & Sons., New York, 1983.

Press, W.H., B.P. Flannery, S.A. Teukolsky & W.T. Vetterling, Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, Cambridge, 1987.

Velleman, P.F. & D.C. Hoaglin, Applications, Basics, and Computing of Exploratory Data Analysis, Duxbury Press, Boston, 1981.

pause

Messwerte	90	Umin	-356.33	Umax	727.90	
Mittelwert	184.37		MittF Mittel	10.53% wF	84.37%	
Zw. (Median)	118.20		MittF Median	12.46% wF	18.20%	
Stetig Histo	101.77	Fensterbreite	7	F Stet Histo	4.39% wF	1.77%
EinfachHisto	103.06	Klassenzahl	383	F Einf Histo	4.64% wF	3.06%

