

Appendix B: Statistische Fehlerbetrachtung zur Lokalisierung eines Hypozentrums

Im Kapitel 5 wurde die Verteilung der gefundenen Beben im Arbeitsgebiet dargestellt. Im folgenden Abschnitt wird das Lokalisierungsproblem von Hypozentren dargestellt. Dabei wird versucht, den statistischen Fehler verschiedener Beben, die für die gefundene Seismizität als repräsentativ gelten können, zu bestimmen.

B.1) Quelle und Verteilung der Fehler

Die erste Quelle für den Fehler liegt in der Bestimmung der Zeiten der S- und P-Einsätze. Diese Einsätze deuten auf die sog. beobachteten Zeiten an. Dieser Fehler kann bei der Beobachtung bis zu ± 0.03 sec betragen, ohne einen großen *misfit* der Einsätze hervorzubringen. Diese Art von Fehler verteilt sich zufällig um den Einsatz und hängt voll vom Beobachter ab. Die Differenzen zwischen den beobachteten und theoretischen Einsätzen setzen die Basis ein, um die Koordinaten des Hypozentrums zu finden.

Die zweite Quelle des Fehlers liegt an dem Geschwindigkeitsmodell. Aus der gleichzeitigen Iteration zur Bestimmung des Hypozentrums und des entsprechenden Geschwindigkeitsmodells resultiert das beste Modell für die beste Lokalisierung.

Die dritte Quelle des Fehlers liegt im Stationsuntergrund und der Stationshöhe. Einerseits hat die Existenz von Sedimenten unter der Station einen direkten Effekt auf das Geschwindigkeitsmodell und deshalb auf die theoretischen Einsätze. Verfahren wie HYPO71 berücksichtigen keine differenzielle Stationshöhe, welche eine Verschiebung insbesondere der Tiefe des lokalisierten Herdes begründet.

Die vierte Quelle des Fehlers liegt in der Lösung selbst. Das Lokalisierungsproblem gehört zum klassischen inversen Problem der Geophysik. Dieses Problem wird deutlich an der linearen Gleichung $\mathbf{d} = \mathbf{G}\mathbf{m}$ zusammengefasst. Die einfachste Methode, um dieses Problem zu lösen, gründet sich auf die Messung der Größe der Vektoren der abgeschätzten Modellparameter \mathbf{m} und vorausgesagten Daten \mathbf{d} . Dieses Problem wird häufig durch die sogenannte *least squares method* gelöst. Um eine Lösung eines umgekehrten Problems abzuschätzen, findet die *least squares method* ein Vektor zum Modellparameter (\mathbf{m}), welcher eine bestimmte Größe der Länge des Vektors der Daten (\mathbf{d}) verkleinert (*Euclidean Distance*). Die *least squares method* benutzt die sogenannte *L2 Norm*, um die Größe der Längen zu quantifizieren. Unter Verwendung der *L2 Norm* durch Angabe des Vektors der Daten (\mathbf{d}) wird das sogenannte *root square mean* oder *rms* gerechnet. Das *L2 Norm* bringt mit sich selbst, daß die Daten ein Art *Gaussian*-statistisch Verhalten besitzen. *Gaussian*-statistisch Verhalten sind meistens *short-tailed* Verteilungen. Deshalb ist es geeignet, um große vorausgesagte Fehler in der Daten hervorzuheben.

B.2) Das inverse Problem zur Lokalisierung eines Hypozentrums

Die Koordinaten eines Hypozentrums bestehen aus den 3 geographischen Koordinaten X_0, Y_0, Z_0 und einer zeitlichen Koordinate t_0 (*origin time*). Außer den geographischen Koordinaten besitzt jede Station, die als Empfänger des Hypozentrums gilt, zwei wichtige Werte: den beobachteten und theoretischen Einsatz (Abb. B.1).

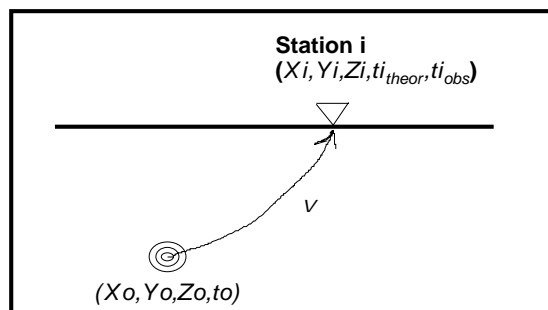


Abbildung B.1

Skizze zu der Beziehung zwischen den geographischen und Zeit-Koordinaten eines Hypozentrums und des Empfängers (Station i).

Sketch for showing the relationship between geographical and time coordinates of an hypocenter and an station (i).

Die Entfernung zwischen Quelle und Station einerseits, und das Geschwindigkeitsmodell andererseits, stehen in Verbindung durch eine Funktion, welche die Laufzeit bestimmt. Ausgehend von einer Lage des Hypozentrums und einem Startmodell für die Geschwindigkeitsverteilung wird einer theoretischer Einsatz für jede Station i gerechnet (G.1).

$$t_{i_{theoretical}} = t_0 + f(X_i - X_0, Y_i - Y_0, Z_i - Z_0, v) \quad (G.1)$$

Das Ziel des gesamten Prozess ist die Differenz zwischen beobachteter und berechneter Laufzeit zu verkleinern. Dafür wird eine Verschiebung des Herdes, welche durch die Werte dX, dY, dZ und dt ausgedrückt wird, durchgeführt (G.2).

$$t_{i_{theoretical}} - t_{i_{observed}} = dt_i \leftrightarrow dX, dY, dZ, dt \quad (G.2)$$

An diesem Schritt wäre es sinnvoll eine Änderung des Geschwindigkeitsmodells auszuführen, um die Differenzen nicht nur zu verkleinern, sondern auch die Verschiebung des Herdes stabiler zu machen. Falls die Differenz dt_i kleiner als ein bestimmten Wert ε ist, kommt das Verfahren zum Ende (G.3).

$$\begin{aligned} dt_i &\leq \varepsilon (?) \\ \text{Yes} &\Rightarrow \text{End} \quad \text{No} \Rightarrow (\text{G.B.4}) \end{aligned} \quad (G.3)$$

Ansonsten werden die entsprechenden dX, dY, dZ und dt auf die Koordinaten des Hypozentrums angewendet (G.4) und wieder mit dem ersten Schritt angefangen (zurück zur G.1).

$$\begin{aligned} Dx &\Rightarrow X_0 \pm Dx \\ Dy &\Rightarrow Y_0 \pm Dy \\ Dz &\Rightarrow Z_0 \pm Dz \\ Dt &\Rightarrow t_0 \pm Dt \end{aligned} \quad (G.4)$$

Wenn ein homogenes Medium unter Berücksichtigung wird und ein XYZ-Koordinaten-System gewählt wird, kann man die Funktion der G.1 wie in der G.5 schreiben, welche nichts anders sind als $time = distance/velocity$.

$$\begin{aligned} &\text{Homogeneous Medium} \\ f(X_i - X_0, Y_i - Y_0, Z_i - Z_0, v) &= (1/v) \cdot \sqrt{(X_i - X_0)^2 + (Y_i - Y_0)^2 + (Z_i - Z_0)^2} \end{aligned} \quad (G.5)$$

Unter dieser Berücksichtigung kann die Differenz der beobachteten und theoretischen Einsätze ausgedehnt wird (G.6).

$$dt_i = (t_i / x_0) \cdot dx + (t_i / y_0) \cdot dy + (t_i / z_0) \cdot dz + (t_i / t_0) \cdot dt \quad (G.6)$$

Hier steht ein wichtiger Hinweis: Die Differenzierung wird nur für die Koordinaten des Hypozentrums gemacht (homogenes Medium). Unter Verwendung verschiedener Stationen erscheint die bekannte Gleichung des inversen Problems (G.7), welche die Daten (\mathbf{d}), die Modellparameter (\mathbf{m}) und das Modell (Matrix $\mathbf{G}_{n \times 4}$) enthält.

$$n \text{ stations} \Rightarrow i=1 \dots n$$

$$\text{G.6} \Rightarrow \Delta d = G \Delta m \quad (G.7)$$

wobei

$$\begin{aligned} \Delta d &= (dt_1, dt_2, \dots, dt_n) \Rightarrow \text{Data} \\ \Delta m &= (dX, dY, dZ, dt) \Rightarrow \text{Model Parameter} \\ \mathbf{G}_{n \times 4} &\Rightarrow \text{Model (Data Kernel)} \end{aligned}$$

und

Least Square Solution (L_2 Norm):

$$Error = (\sum_{i=1,n} |e_i|^2)^{1/2}, e_i = \Delta d_i - \sum_{j=1,n} (G_{ij} m_j)$$

Die Benutzung von $\Delta \mathbf{d}$ und $\Delta \mathbf{m}$ wird häufig in der Literatur ohne Δ geschrieben, aber man muss die richtige Bedeutung sicher stellen. Unter Verwendung der *least square Method (L2 Norm)* kann man beweisen, daß eine Einschätzung der Modellparameter wie in der Gleichung G8 aussieht.

$$\Delta \mathbf{m} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \Delta \mathbf{d} \quad (\text{G.8})$$

Das umgekehrte Problem kann auf die G.9 und G.10 angewendet werden. Hauptsache des Problems ist die sogenannte *generalized inverse matrix* zu finden, um den Vektor \mathbf{m} , ausgehend vom Vektor \mathbf{d} , zu berechnen. Nachdem die *generalized inverse matrix* gefunden wurde, muß die Frage beantwortet werden, wie gut die Daten die Einschätzung der Modellparameter anpassen. Hier entsteht das sogenannte *eigenvalue problem*. Der sogenannte *rank* der Matrix \mathbf{G} ist als die Anzahl von *nonzero singular value (p)* definiert und wird häufig kleiner als n . Das bedeutet, das die Matrix \mathbf{G} ermöglicht, nur einigen Aspekten des Vektors \mathbf{d} hervorzuheben.

$$\mathbf{G}^{-g} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \quad \text{:Generalized Inverse} \quad (\text{G.9})$$

$$\begin{aligned} \mathbf{m}^{\text{est}} &= (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d}^{\text{obs}} \\ \mathbf{d}^{\text{pre}} &= \mathbf{G} \mathbf{m}^{\text{est}} \end{aligned} \quad (\text{G.10})$$

Die Gleichung G.11 bildet die Beziehung zwischen den beobachteten und theoretischen Daten. Beiden Vektoren werden durch die Matrix \mathbf{N} in Verbindung gebracht. Die Matrix \mathbf{N} ist die sogenannte *data resolution matrix* oder *information density matrix*. Die beschreibt wie gut die beobachteten Daten die vorausgesagten zusammenpassen. Falls $\mathbf{N} = \mathbf{I}$ (*identity matrix*) ist der Fehler des Voraussagen null. Im Allgemeinen ist dieser Fall nur theoretisch. Die vorausgesagten Werte werden von ihren „nachbarten“ Werten beeinflusst. In dieser Sinne beschreiben die Reihen der \mathbf{N} -Matrix wie gut die Daten (*diagonal*) unabhängig von ihren „nachbarn“ Daten vorausgesagt werden können. Falls die Daten gut gelöst wurden, wird die \mathbf{N} -Matrix auf der Diagonale mit 1.0 gefüllt und mit 0 außerhalb der Diagonalen. Die Elemente der Diagonalen der \mathbf{N} -Matrix sagen aus, welche Bedeutung ein Datum (z.B. die Phase einer bestimmten Station) für die Vorhersage ist. Die *data resolution matrix* gibt eine Idee der Einfluss der Stationen bei der Lokalisierungsverfahren. Unter Berücksichtigung muß man haben: Die *data resolution matrix* hängt von der Daten nicht, sondern nur von der matrix \mathbf{G} (data kernel), welche das Modell und Geometrie der Lokalisierung besitzt.

$$\begin{aligned} \mathbf{d}^{\text{pre}} &= \mathbf{G}_p (\mathbf{G}_p^{-g} \mathbf{d}^{\text{obs}}) = (\mathbf{G}_p \mathbf{G}_p^{-g}) \mathbf{d}^{\text{obs}} = \mathbf{N} \mathbf{d}^{\text{obs}} \\ \mathbf{N}: \text{Data Resolution Matrix} \end{aligned} \quad (\text{G.11})$$

B.3) Bedeutung der \mathbf{G} -Matrix

Eine Vereinfachung des Problems liegt daran, nur die geographischen Koordinaten zu betrachten. Dafür muss man die zeitliche Komponente ausschließen, d.h., die Anzahl von P- und S-Einsätzen miteinbeziehen. Das ist der Fall bei ca. 35 % der gesamten Daten der vorliegenden Arbeit. In diesem Fall können die Differenzen der beobachteten und theoretischen Einsätze wie in der Gleichung G12 dargestellt werden. Eine Erweiterung der Gleichung G.12 wird in Gleichung 5.13 dargestellt, welche die bekannte Darstellung des Umkehrproblems darstellt (G.14).

$$dT = \frac{\partial \bar{T}}{\partial r} \cdot d\bar{r} \quad (\text{G.12})$$

$$\begin{bmatrix} (dt_{s-p})_1 \\ (dt_{s-p})_2 \\ \cdot \\ \cdot \\ (dt_{s-p})_k \end{bmatrix} = \begin{bmatrix} (\partial t_{s-p}/\partial x)_1 & (\partial t_{s-p}/\partial y)_1 & (\partial t_{s-p}/\partial z)_1 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ (\partial t_{s-p}/\partial x)_k & (\partial t_{s-p}/\partial y)_k & (\partial t_{s-p}/\partial z)_k \end{bmatrix} \begin{bmatrix} dx \\ dy \\ dz \end{bmatrix} \quad (G.13)$$

$$\mathbf{d} = \mathbf{Gm} \quad (G.14)$$

Der Vektor \mathbf{d} stellt die Differenzen der beobachteten und theoretischen P- und S-Einsätze an jeder Station dar. Der Vektor \mathbf{m} stellt die Veränderung der Koordinaten des Hypozentrums dar. Die Matrix \mathbf{G} stellt sowohl die Beziehung zwischen Quelle und Empfänger an jeder Station (G.4.1) als auch die Geschwindigkeitsstruktur dar. Für jede Station stellen die Komponenten x, y, z einen Vektor dar. Je nachdem, sich diese Vektoren als eine Einheit von „linearly independent vectors“ verhalten, wird die Matrix \mathbf{G} stabiler und der gesamte Prozess des umgekehrten Problem stärker. Aus diesem Grund sind einige Stationen häufig bei dem Prozess zur Lokalisierung eines Bebens „schlechter“ als anderen gezeichnet worden. Andererseits gelten die Komponenten der Matrix \mathbf{G} als ein Gewicht für die Variation der Koordinaten des Hypozentrums. Je nachdem, diese Komponenten kleiner werden, lässt sich die Variation der Koordinaten des Hypozentrums unstabiler. Davon werden oberflächennahe Beben besonders auf ihre z -Komponenten betroffen. Aus diesem Grund ist die Lokalisierungsgenauigkeit von Beben, die oberflächennah liegen, niedrig.

B.4) Zur Fehlerbetrachtung bei der Bestimmung der Modellparameter und die „data resolution matrix“

Da das Ziel dieses Abschnitts ist, eine theoretische Abschätzung des Fehlers zu betrachten, wird die Gleichung G.10 und die data resolution matrix (G.11) für ca. 30 Ereignisse berechnet. Erstens wurde versucht, eine Einschätzung der theoretischen Verschiebung eines Hypozentrum sowohl nach verschiedenen Tiefen als auch mit verschiedenen Epizentren zu berechnen. Die Starthypozentren entsprachen den lokalisierten Beben. Die theoretische Berechnung der Laufzeiten berücksichtigte die Stationshöhe und die Geschwindigkeitsstruktur des Basismodells. Zweitens wurde der Einfluss auf die Lokalisierung sowohl der Geometrie Quelle-Station als auch der Zahl von Einsätzen (S- bzw. P-Phasen) durch die Darstellung der data resolution matrix abgeschätzt.

In der Abbildung B.2 ist die Abschätzung der theoretischen Verschiebung nach der Tiefe dargestellt. Die einzelnen Werte deuten auf Variationen hin, die meistens unterhalb der schwarzen Kurve liegen. Dieser Fehler passt gut mit der abgeleiteten Variation aus der Tabelle 5.1 (Kapitel 5) überein.

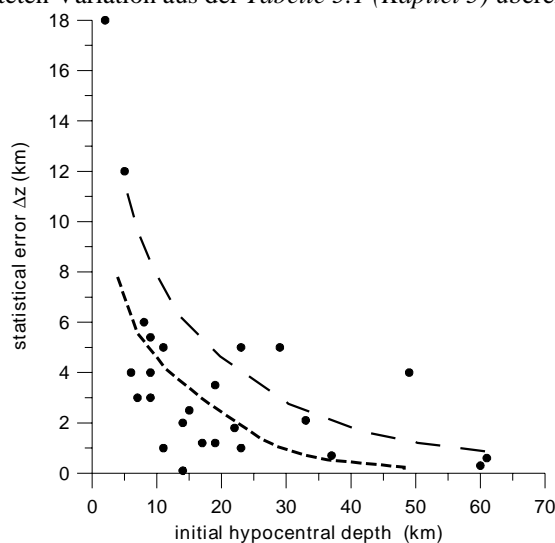


Abbildung B.2

Mögliche Werte für den Fehler in der Z-Komponente nach der Starttiefe des Hypozentrums
Possible values for the error in the Z-axis as function of initial hypocentral depth.

In der *Abbildung B.3* sind 3 Beispiele zur Bestimmung der *data resolution matrix* dargestellt. Die signifikantesten Merkmale spielen eine entscheidende Rolle, wie die S-Einsätze bei der Lokalisierungsroutine.

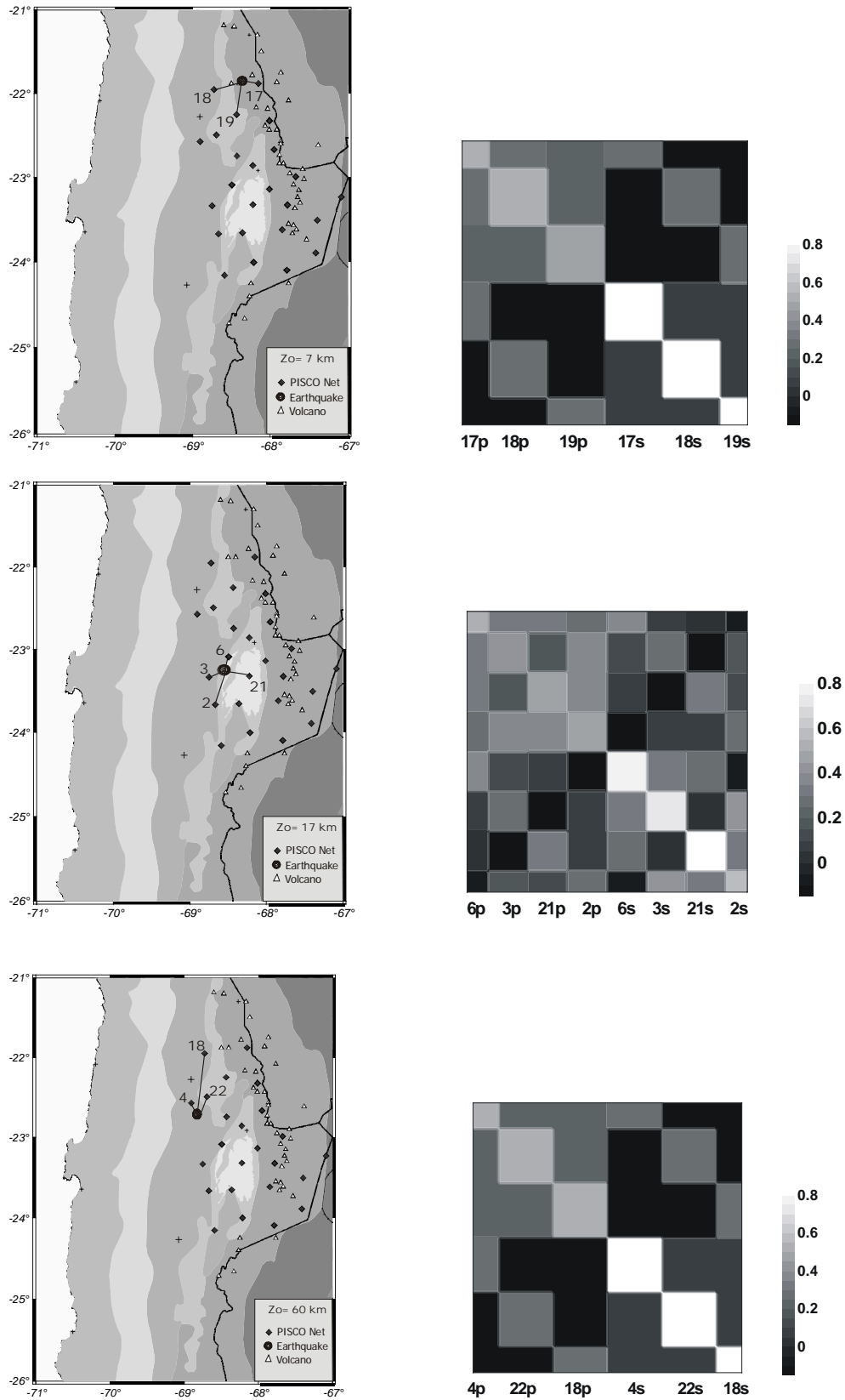


Abbildung B.3

Auf der linken Seite stellt wird die Geometrie zwischen dem selektierten Beben und den Stationen dargestellt.

Die Tiefe des Bebens wird als Z_0 angedeutet. Rechts wird die *data resolution matrix* dargestellt. Die Nummer der Stationen begleitet von einem p (P-Einsatz) oder einem s (S-Einsatz) entsprechen zu der Ordnung der Matrix. Von oben nach unten gilt dieselbe Ordnung wie von links nach rechts.

On the left hand the geometry between earthquakes and stations are presented. At the bottom the depth of the event (Z_0) is to be pointed out. On the right hand the *data resolution matrix* is showed. The number of stations accompanied by one p (P-onset) or one s (S-onset) correspond to matrix order. From above to bottom this sequence is valid as well as from left to the right.

Der Unterschied zwischen P- und S-Einsätze in der *data resolution matrix* könnte auf die unterschiedliche Größe der Komponenten bezüglich auf P- und S-Wellen in der G-Matrix wegen ihrer Differenz in der Geschwindigkeiten ($V_p > V_s$).

B.5) Zusammenfassung

- 1.- Eine Einschätzung des Vektors der Verschiebung des Hypozentrums, hier auch als Vektor der Modellparameter (\mathbf{m}) bezeichnet, und der *data resolution matrix* wurde durchgeführt..
- 2.- Aus den abgeleiteten Fehlern ist die Variation in der Tiefe eines Beben im Untersuchungsgebiet deutlich größer als die Variation auf der horizontale Ebene. Diese Ergebnisse können wie folgt zusammengefasst werden:

Hypocentral Depth (km)	Maximum Variation (km)
5-15	6-12
15-30	3-5
30 -70	1-3

- 3.- Die *data resolution matrix* enthält die Bedeutung der S-Einsätze. Sie variiert kaum mit der Anzahl und Verteilung der Stationen (*die gefundenen Beben werden durch ihre kleine Magnitude charakterisiert*), da eine geringe Zahl von Empfängern immer extrem kleine Fehler besitzt.
- 4.- Das sog. *gap* scheint keine signifikanten Differenzen in der *data resolution matrix* zu erzeugen.