

International Centre for Theoretical Physics







H4.SMR/1642-1

"IAG-IASPEI Joint Capacity Building Workshop on Deformation Measurements and Understanding Natural Hazards in Developing Countries"

17 - 23 January 2005

**Problem of Earthquake Location Accuracy** 

E.R. Engdahl

University of Colorado at Boulder Department of Physics Boulder, CO USA

## Epicenter Accuracy Based on Seismic Network Criteria

István Bondár, SAIC Stephen C. Myers, LLNL E. R. Engdahl, CU Boulder Eric A. Bergman, CU Boulder

Geophysical Journal International (2004) 156, 483-496

#### The Goal

Develop simple criteria for estimating the accuracy of catalog epicenters, based on parameters which are commonly reported or which can be derived from earthquake bulletins.

Catalog: hypocenters only Bulletin: includes station readings

## Uses of Earthquake Catalogs and Bulletins

Seismotectonic Studies
Earthquake Hazard Estimation
Inversion for Earth Structure

## Standard Catalogs

- Iterative, linearized inversion
- I-D Earth model
- Minimal effort at outlier rejection
- Simple weighting schemes
- Use only first-arriving P

#### Research Catalogs

- Non-linear inversion
- 3-D Earth model
- Outlier analysis
- Advanced weighting schemes
- Secondary phases
- Phase re-association

But for many regions and purposes, only standard earthquake catalogs are available, so how do we make the best use of them?

In particular, how do we evaluate the accuracy of the epicenters?

Most catalogs are compiled for completeness to the lowest possible magnitude, not for consistent location accuracy. Two Components of Location Accuracy
Formal Uncertainty (data variance)
Location bias (un-modeled Earth structure)

#### Formal calculations usually overestimate location accuracy

- Even with a "proper" Bayesian approach
- Violation of assumptions about the statistics of data errors causes location bias
  - Longshot: 26 km teleseismic location bias from slab effect in the Aleutians
  - California: 5 km regional location bias from lateral heterogeneity across the San Andreas fault

### So, how to proceed?

- Compile datasets of "reference" events.
- Carry out location tests with subsets of data characterized in simple ways that could be applied to any standard EQ bulletin (i.e., with station information).
- Characterize location accuracy in a probabilistic way for different situations of interest.

### Data Sets

#### Reference Events

Fiducial Explosions • Essentially error-free • Used for local network study • Well-Located Earthquakes and Explosions Special studies to reduce location bias to a minimum. • Used for regional/teleseismic study.

## Fiducial Explosions

- Dead Sea shot
- Swiss ammunition explosion
- Many nuclear shots

#### Dead Sea Shot

- In November 1999, three calibration explosions were detonated in the Dead Sea.
- The two larger events were recorded at stations in Israel, Jordan, Lebanon and Syria, to distances of 250 km.
- The combined local network provides excellent network coverage with considerable azimuthal redundancy.



#### Swiss Explosion

- On 1992 November 2, an ammunition storage site in the Swiss Alps exploded.
- While the epicenter and depth are tightly constrained, there is a small uncertainty in this origin time. However, this does not disqualify the event from fiducial status.



To our knowledge, these are the only exactly known ground truth events with the requisite station coverage at local distances.

 To leverage the data from these two rare events we use Monte Carlo location simulation to test many "realizations".

Because both explosions lie in rather complex regions where strong heterogeneity can be expected, the analysis of these events should provide conservative estimates of location

accuracy.

#### Well located earthquakes and explosions



- 671 earthquakes, 1234 nuclear shots
- HDC analysis of clusters.
- Reference events calibrate cluster location bias, allowing all events in the cluster to be made ~ bias-free
- Location errors < 5 km.
- Many nuclear shots are "fiducial", location errors typically < 2 km.

#### **Epicentral Distance Ranges**

Local	Δ < 250 km
Near-Regional	2.5° < Δ < 10°
Regional	$2.5^{\circ} < \Delta < 20^{\circ}$
Teleseismic	28° < Δ < 91°





## Location Tests

# Parameters reported or easily derived from EQ bulletins

- Epicentral distance to the closest station
- Number of stations and phases used to locate the event
- Geographic station coverage (azimuthal gap and secondary azimuthal gap)

#### However...

 Distance to nearest station and number of stations and phases used *do not* correlate well with epicenter location accuracy

The dominant factor is geographic station coverage

#### Azimuth Gap



Primary Azimuth Gap

Secondary Azimuth Gap

Although the 82° primary azimuth gap is quite good, any reading error at HKC may strongly bias the location. Secondary azimuth gap of 160° reflects this weakness.

#### Local Network Location Accuracy Criteria

- The Dead Sea and Swiss ammunition explosions were used to develop and test location accuracy for local networks.
- Each event was relocated many times with I0 randomly selected stations within 250 km of the epicenter.
- I0,000 realizations were generated for each event, and the azimuthal gap, secondary azimuthal Gap and number of atations within 30 km from the epicenter were measured for each Monte Carlo realization.

• Why one station within 30 km? Depth control • Why 10 stations? Typical of local networks • Why use only stations less than 250 km? Avoid Pn/Pg cross-over

#### Histograms of Monte Carlo realizations for local network location accuracy



It is not possible to define constraints on a network geometry that would select all events located with 5 km accuracy or better and reject those with mislocation greater than 5 km.

Therefore, we specify the confidence level with which candidate GT5 events are selected.

#### Ground Truth (GT) Criteria

• We adopt a "ground truth" nomenclature GTX<sub>C%</sub> to designate location accuracy, where the "X" suffix is the accuracy in kilometers and "C%" is the percentage confidence.

For example, events that are thought to be accurate to 5 km at the 95% confidence level are designated GT5<sub>95%</sub>.

## Cumulative percentile of mislocation of local network realizations



- Based on the Monte Carlo simulation, crustal events are located with 5 km or better accuracy at the 95% confidence level if they are located:
- (I) with at least 10 stations, all within 250 km
- (2) with an azimuthal gap of less than 110°
- (3) with a secondary azimuthal gap of less than 160°
- (4) with at least one station
   within 30 km from the epicenter

#### Distributions of mislocation, origin time, and depth for local network realizations



#### Regional and Near-regional Networks



Less is more: near-regional networks outperform regional networks

Cumulative percentile of mislocations for regional and near-regional networks (secondary azimuth gap < 120°)



# Regional and near-regional location accuracy criteria

 A secondary azimuthal gap of less than 120° selects earthquakes at:

• GT20<sub>90%</sub> for near-regional networks

GT25<sub>90%</sub> for regional networks

## Location accuracy criteria for teleseismic networks



Explosions are much better located than earthquakes
#### Teleseismic location accuracy criteria

A secondary azimuthal gap of less than 120° selects:
Earthquakes at GT25<sub>90%</sub>

Explosions at known test sites at GTI5<sub>95%</sub>

# Summary of Results

## The Method

- Use fiducial explosions (GT0) to develop location accuracy criteria for local networks
- Use well-located earthquakes (GT5) to develop location accuracy criteria for regional and teleseismic earthquakes

#### Local network location accuracy criteria

•The GT5<sub>95%</sub> epicenter accuracy criteria for earthquakes observed by local networks (0-2.5°) are:

• at least 10 stations, all within 250 km

• these 10 stations should have a primary azimuthal gap of less than 110°

 these 10 stations should have a secondary azimuthal gap of less than 160°

• at least one station should be within 30 km

#### Regional and near-regional network location accuracy criteria

A secondary azimuthal gap of less than 120° selects earthquakes at:

GT20<sub>90%</sub> for near-regional networks

• GT25<sub>90%</sub> for regional networks

## Teleseismic network location accuracy criteria

• A secondary azimuthal gap of less than 120° selects:

- Earthquakes at GT25<sub>90%</sub>
- Explosions at known test sites at GTI 5<sub>95%</sub>

## Discussion

#### Continental Earthquakes

- Although there are some subduction zone events in the GT5 dataset, the location accuracy criteria derived here are most relevant to continental earthquakes.
- Location accuracy for earthquakes near subduction zones will generally be worse.

#### Special Studies

• It is assumed that no special effort has been made to remove location bias through the use of an optimal velocity model or travel time corrections, or through special analysis of waveforms or readings to improve the phase picks.

 Location accuracy can be improved by such studies.

#### **Distant Stations**

- For local networks, use of stations beyond 250 km may *reduce* epicenter location accuracy
- Consider the trade-off between reduced azimuthal gaps and increased bias from phase association problems and lateral heterogeneity

## Mixed Regional-Teleseismic Studies

 Use of regional+teleseismic arrivals may yield worse location accuracy than teleseismic arrivals alone

Geographic coverage vs. lateral heterogeneity

#### Final observations

- Typical local networks can achieve 5-km levels of epicenter location accuracy, even without a "custom" velocity model, if azimuthal control is good and the solution is not biased by the use of regional distance data
- Regional networks provide no better location accuracy than teleseismic networks if they do not account for lateral heterogeneity in the crust and upper mantle

#### Improved Locations and Focal Depths for Well-Constrained Teleseismic Earthquakes

 E.R. Engdahl, Van der Hilst, R.D., and Buland, R.P., 1998, Global teleseismic earthquake relocation with improved travel times and procedures for depth determination: Bulletin of the Seismological Society of America, v. 88, p. 3295-3314 (EHB Method).

#### The Problem

Although useful for seismic hazard assessment, global compilations of earthquake hypocenters and associated phase arrival times and residuals often are too inhomogeneous to be confidently applied, for example, to problems such as Earth structure determination.

The main problem is the varying level of mislocation, particularly focal depth, introduced largely by errors in the reference Earth model, unaccounted for effects of lateral heterogeneity, and phase misidentification. The result is loss of structural signal in the residuals.

#### **The Solution**

The bias in hypocenter determination can be significantly reduced and at least part of the lost structural signal recovered by

- •Using a proper reference Earth model
- •Using later arriving phases in the relocation procedure
- •Limiting the events of interest only to those that are well-constrained teleseismically

#### **Standard Location Methods**

- Standard location methods are based on Geiger's method (Geiger, 1910, 1912) that became practical with the advent of modern computers
- The basic methodology is that predicted phase arrival times for a trial hypocenter and origin time are calculated for the observing stations using the chosen reference Earth model (1-D)

The phase arrival time residuals (observed minus calculated) are then related to hypocenter (latitude  $\theta$ , longitude  $\phi$ , depth z) and origin time (t<sub>o</sub>) perturbations by a linearized equation of the form

$$\mathbf{r} = -\sin\theta\,\sin\alpha\left(\frac{\partial T}{\partial\Delta}\right)(\Delta\phi) + \cos\alpha\left(\frac{\partial T}{\partial\Delta}\right)(\Delta\theta) + \left(\frac{\partial T}{\partial z}\right)(\Delta z) + \Delta t_{o}$$

where  $\alpha$  is azimuth from the event to the station.

This can be written in matrix form as  $\mathbf{r} = \mathbf{A} \mathbf{x}$ where  $\mathbf{r}$  is the vector of residuals,  $\mathbf{A}$  is the matrix of derivatives, and  $\mathbf{x}$  is the vector of origin time and hypocenter perturbations. Because of non-linearity this system of equations is solved iteratively via matrix inversion until convergence is attained.

Options for carrying out the matrix inversion include forming the normal equations

 $\mathbf{A}^{\mathrm{T}} \mathbf{r} = \mathbf{A}^{\mathrm{T}} \mathbf{A} \mathbf{x}$ 

followed by the application of a standard solution algorithm for square symmetric matrices or using step-wise linear regression, the QR algorithm , or singular value decomposition. Weighting in the EHB method is performed based on both reported arrival-time reading precision and phase variance as a function of distance.

Weighting is easily incorporated in the inversion by constructing a weight matrix W with diagonal elements equal to the square root of the weight value and modifying  $\mathbf{r} = \mathbf{A} \mathbf{x}$  as follows:

 $\mathbf{W} \mathbf{r} = \mathbf{W} \mathbf{A} \mathbf{x}$ 

and then solving the weighted system as before.











#### ak135 (residuals normalized)







# Improving Quality and Usage of Data

- One direct method to improve seismic event locations is by improving the quality and utilization of the data.
- Standard teleseismic catalogs (ISC, NEIC) still rely almost entirely on first arriving P phases for locating events.
- Many studies have shown that the inclusion of later arriving phases can provide greater constraints on hypocenter parameters, especially focal depth.

- Epicenter constraints are improved by the inclusion of S and P-wave core phases because their travel-time derivatives differ significantly in magnitude from direct P.
- Depth to origin trade-off is avoided by the inclusion of depth phases (pP, pwP, sP) because their travel time derivatives are opposite in sign to direct P.

- A problem with the use of depth phases is that their correct identification often requires knowledge of the event depth and distance. Hence, depth phase arrivals are re-identified after each iteration using a statistically based association algorithm.
- Probability density functions (PDFs) for depth phases, centered on their theoretical relative travel times for a given hypocenter, are compared to the observed phase arrivals.
- When PDF's overlap for a particular depth phase, a phase identification is assigned in a probabilistic manner based on the relevant PDF values, making sure not to assign the same phase to two different arrivals.














#### **Model Conclusions**

The model ak135 provides a very good fit to a wide range of seismic phases.

The mantle S wave bias of iasp91 has been removed.

Most core phase times are quite well matched and a baseline problem with ISC PKP phases removed.

Thus, for global earthquake location there has been convergence on global, radially symmetric, P- and S-velocity Earth models that provide a good average fit to reported phase arrival times.

## **Station Corrections**

- Station corrections are a long-recognized mechanism for trying to compensate for upper mantle velocity heterogeneity beneath stations when 1-D velocity models are assumed.
- In the EHB method a teleseismic "patch correction" approach has been adopted, determining from P teleseismic residuals a single median correction for all stations within 5 x 5 degree regions.
- Patch medians derived separately from teleseismic P and PKP residual data agree well with each other.





## Aspherical Earth Structure

- The travel times predicted by recently developed, radially symmetric, Earth models (such as ak135) are extremely valuable for earthquake location and phase identification.
- Nevertheless, most earthquakes occur in or near subducted lithosphere where aspherical variations in upper mantle seismic wave velocities are large (i.e., on the order of 5 - 10%)
- Such lateral variations in seismic velocity, the uneven spatial distribution of seismological stations, and the specific choice of seismic data used to determine the earthquake hypocenter can still easily combine to produce bias in earthquake locations of several tens of kilometers

 Tests of location bias globally using a new archive of reference event information and the EHB location algorithm show that most explosions and earthquakes are mislocated by less than 20 km if the secondary azimuth gap to observing stations at all distances is less than 180 degrees.

# EHB vs GT5



## **Location Conclusions**

- At least in the case of events well constrained azimuthally by reporting stations, mislocation errors introduced by lateral heterogeneity can be minimized.
- For smaller and / or poorly recorded events, however, there is not much hope of significantly reducing the resulting mislocation error until we can somehow better account for aspherical Earth structure in 1-D earthquake location procedures.

# Global Seismicity: 1900-1999

 Engdahl, E.R., Villasenor, A., 2002, Global Seismicity: 1900-1999, International Handbook of Earthquake and Engineering Seismology, v. 81A, p. 665-690, Elsevier Science Ltd., Amsterdam, The Netherlands.

### Approach

- Combine existing global catalogs of earthquake locations and magnitudes into a single catalog
- For shallow earthquakes assign the moment magnitude Mw or the surface-wave magnitude Ms
- For earthquakes deeper than 60 km assign the moment magnitude Mw, or the body-wave magnitude mB (broadband) or mb (short-period)
- Use assigned magnitudes to determine catalog magnitude completeness thresholds and to assign magnitude cut-off values as a function of time
- Use the EHB location methodology to relocate all events within the magnitude cut-off thresholds for which digital phase arrival-time data are available

- What makes EHB hypocenters better than ISS, ISC and PDE hypocenters?
- Use of an Improved 1-D Global Travel Time Model (ak135)
- Iterative Relocation With Dynamic Phase Identification
- Use of First Arriving P, S and PKP Phases
- Use of Teleseismic Depth Phases pP, pwP and sP (with PDF's and bounce point corrections)
- Ellipticity Corrections for ak135 Model
- Empirical Teleseismic "Station" Patch Corrections (5 x 5° patches)
- Weighting by Phase Variance as a Function of Distance
- At Least 10 Teleseismic Observations
- Teleseismic Secondary Azimuth Gap < 180°















#### Focal depth distribution in earthquake catalogs





**FIGURE 1** Comparison between magnitudes reported by different catalogs relative to surface wave magnitudes ( $M_S$ ) reported in Abe's catalog (Abe, 1981, 1984; Abe and Noguchi, 1983a,b). The bin width for all histograms is 0.1 magnitude units, and the number of events in each bin is shown as a percentage of the total number of events. N, total number of events; Avg, average residual; Std, standard deviation of the residuals. Catalogs compared: (a) Pacheco and Sykes (1992); (b) Gutenberg and Richter (1954); (c) Båth and Duda (1979); (d) Rothé (1969); (e) Pasadena single-station magnitude for events before 1960; (f) Pasadena magnitudes after 1959.



**FIGURE 1** Comparison between magnitudes reported by different catalogs relative to surface wave magnitudes ( $M_S$ ) reported in Abe's catalog (Abe, 1981, 1984; Abe and Noguchi, 1983a,b). The bin width for all histograms is 0.1 magnitude units, and the number of events in each bin is shown as a percentage of the total number of events. N, total number of events; Avg, average residual; Std, standard deviation of the residuals. Catalogs compared: (a) Pacheco and Sykes (1992); (b) Gutenberg and Richter (1954); (c) Båth and Duda (1979); (d) Rothé (1969); (e) Pasadena single-station magnitude for events before 1960; (f) Pasadena magnitudes after 1959.



**FIGURE 1** Comparison between magnitudes reported by different catalogs relative to surface wave magnitudes ( $M_S$ ) reported in Abe's catalog (Abe, 1981, 1984; Abe and Noguchi, 1983a,b). The bin width for all histograms is 0.1 magnitude units, and the number of events in each bin is shown as a percentage of the total number of events. N, total number of events; Avg, average residual; Std, standard deviation of the residuals. Catalogs compared: (a) Pacheco and Sykes (1992); (b) Gutenberg and Richter (1954); (c) Båth and Duda (1979); (d) Rothé (1969); (e) Pasadena single-station magnitude for events before 1960; (f) Pasadena magnitudes after 1959.



**FIGURE 2** Frequency-magnitude (Gutenberg and Richter) relations for the centennial catalog. Open circles represent single frequencies (incremental number of earthquakes with magnitudes in  $M \pm \delta M/2$ ) and filled circles represent cumulative frequencies (total number of earthquakes with magnitudes  $\geq M$ ). The width of the magnitude interval  $\delta M$  is 0.1 magnitude units. The single and cumulative frequencies are normalized to events per year, and the magnitudes have been adjusted to  $M_S$  (see text): (a) historical seismicity (1900–1963), and (b) recent seismicity (1964–1999).



**FIGURE 2** Frequency-magnitude (Gutenberg and Richter) relations for the centennial catalog. Open circles represent single frequencies (incremental number of earthquakes with magnitudes in  $M \pm \delta M/2$ ) and filled circles represent cumulative frequencies (total number of earthquakes with magnitudes  $\geq M$ ). The width of the magnitude interval  $\delta M$  is 0.1 magnitude units. The single and cumulative frequencies are normalized to events per year, and the magnitudes have been adjusted to  $M_S$  (see text): (a) historical seismicity (1900–1963), and (b) recent seismicity (1964–1999).



FIGURE 3 Number of events in the centennial catalog as a function of time for the three magnitude levels specified in the legend: (a) number of events per year; (b) number of events in 10 y intervals. The total number of events in each interval is divided by the interval width to allow direct comparison between the two histograms.





#### TABLE1 List of Earthquakes (Magnitude >= 7) for 1900-1999.

Year	М	d	h:min	) Sec	Lat.	Long.	Dep.	Mag.	sc	icat	mdo	Region:Earthquake Name
1921 1921 1921 1921 1921 1922	11 11 12 12 1	11 15 8 18 6	18:36 20:36 12:31 15:29 14:10	26.2 33.8 0.0 28.8 43.8	7.90 36.12 36.00 -4.04 -20.41	127.26 70.72 140.20 -71.22 -76.39	35 152 35 545 25	7.3 7.6 7.0 7.5 7.1	Mw mB Mj mB Mw	EHB EHB UTSU EHB EHB	P&S ABE1 UTSU ABE1 P&S	Philippines
1922 1922 1922 1922 1922	1 1 1 3 3	9 17 31 4 28	5:09 3:50 13:17 13:07 3:58	33.8 1.5 28.7 44.7 1.3	23.22 -6.48 40.70 52.92 -21.45	-45.93 -71.86 -125.55 157.18 -68.13	15 359 15 241 136	7.0 7.4 7.2 7.1 7.1	Ms mB Mw mB mB	EHB EHB EHB EHB EHB	ABE1 ABE1 P&S ABE1 ABE1	Calif.: Cape Mendocino
1922 1922 1922 1922 1922 1922	9 9 10 10 11	1 14 11 24 7	19:16 19:31 14:50 21:21 23:00	9.2 42.5 6.1 3.4 15.5	24.51 24.38 -16.12 47.27 -28.44	122.04 122.64 -72.39 152.19 -72.19	35 35 160 35 25	7.5 7.1 7.6 7.3 7.1	Mw Mw mB mB Ms	EHB EHB EHB EHB EHB	P&S P&S ABE1 ABE1 ABE1	
1922 1922 1922 1923 1923	11 12 12 1 2	11 6 31 22 2	4:32 13:55 7:20 9:04 5:07	45.2 41.0 11.4 19.5 42.7	-28.55 36.44 45.74 40.49 54.02	-70.75 70.94 150.80 -125.32 161.52	35 240 35 15 35	8.7 7.3 7.0 7.1 7.1	Mw mB Ms Mw Mw	EHB EHB EHB EHB EHB	P&S ABE1 ABE1 P&S P&S	Chile
1923 1923 1923 1923 1923 1923	22333	3 24 2 16 24	16:01 7:34 16:48 22:01 12:40	48.8 44.2 44.6 43.7 19.9	53.85 55.94 7.49 6.49 30.55	160.76 162.62 124.93 127.06 101.26	35 35 87 35 25	8.5 7.2 7.1 7.0 7.2	Mw Mw Mw Ms Mw	EHB EHB EHB EHB EHB	P&S P&S P&S ABE1 P&S	Kamchatka China: Luhuo-Dawu

### TABLE 2 Frequency–Magnitude Distribution for 1900–1999

Incre	mental		Cumulative					
$\leq l$	M<	Events y <sup>-1</sup>	$M \ge$	Events y <sup>-1</sup>				
5.5	6.0	164 <sup>a</sup>	5.5	264 <sup>a</sup>				
6.0	6.5	62 <sup>a</sup>	6.0	$100^{a}$				
6.5	7.0	22	6.5	38				
7.0	7.5	12	7.0	16				
7.5	8.0	3	7.5	4				
8.0		0.7	8.0	0.7				

<sup>a</sup>For magnitudes smaller than 6.5 the number of events is based on the period 1964–1999. Centennial Earthquake Catalog (1900-1999)



