COMBINING ANALYST AND WAVEFORM-CORRELATION-BASED ARRIVAL TIME MEASUREMENTS IN THE BAYESLOC MULTIPLE-EVENT LOCATION ALGORITHM

Stephen C. Myers, Gardar Johannesson, Douglas A. Dodge, and Nathan A. Simmons

Lawrence Livermore National Laboratory

Sponsored by National Nuclear Security Administration

Award No. DE-AC52-07NA27344/LL10-DM4TT-NDD02

ABSTRACT

We add relative arrival-time measurements that are derived from waveform correlation to the Bayesloc multiple-event location algorithm. Bayesloc is a formulation of joint probability over event locations, travel time corrections, phase labels, and arrival-time measurement errors. The Bayesloc formulation is hierarchical with distinct statistical models for each component of the multiple-event system, including prior constraints for any of the parameters. Bayes' Theorem allows calculation of the joint probability for hypothesized configurations of Bayesloc parameters, which facilitates using the Markov-Chain Monte Carlo (MCMC) method to draw samples from the joint probability function. The marginal posteriori distribution for each parameter or covariance between parameters is inferred from MCMC samples. Correlation-based picks are integrated into the Bayesloc formulation by including a new category of arrival time measurement that is derived from correlation of empirical waveforms. Because relative picks are derived from correlation between two waveforms and absolute-time picks are made by analysis of a single waveform – typically an analyst, error processes for relative and absolute arrival time measurements are independent. Relative pick precision is formulated as a function of correlation coefficient and the time-bandwidth product of the correlated waveforms, and absolute arrival times precision – as described in previous work – is formulated as a function phase type, the station, and the individual event.

Bayesloc functionality is unchanged for absolute arrival-time data set, and Bayesloc operate as a double-difference locator – with the added benefit of data error modeling – for correlation picks data sets. In the general case – where a data set is a combination of correlation and analyst picks – the precise relative picks provide a cross check on the characterization analyst picks errors and improves identification of outlier analyst picks, both of which reduce location errors and estimates of travel time corrections. Improved measurement precision and estimation of travel time corrections enhances the utility of Bayesloc as a locator, and improved data-set consistency and posteriori error estimation enhance the utility of Bayesloc in the development of travel time calibration (e.g., tomography) data sets.

OBJECTIVES

Introducing differential arrival time measurements into *Bayesloc* enables direct use of waveform correlation picks, which are the most precise type of arrival-time measurement available. While analyst picks will comprise the vast majority of P-wave arrival-time measurements for the foreseeable future, introduction of waveform-correlation picks into the location and calibration system is a step towards full utilization of the waveform data for location (Waldhauser and Ellsworth, 2000; Schaff et al., 2004; Richards et al., 2006). Recognizing the need to accommodate both types of arrival time measurements – analyst picks and waveform correlation – we have formulated an extension to Bayesloc that accommodates direct input of differential arrival times. Specifically, *Bayesloc* now accepts the time difference between arrivals of the same phase type from two events at a common station. The correlation coefficient (and time bandwidth in later versions) of the waveform correlation is also input, which is used as a scaling factor in the formulation of measurement precision.

Direct use of differential times generally leads to improved precision of relative locations and can improve absolute location in cases with outstanding local-network coverage (Menke and Schaff, 2004). For most regional and teleseismic networks, continuing the use of absolute-time picks provides needed control on absolute location, and the introduction of differential times can provide precise relative locations.

Combining absolute and differential arrival times in *Bayesloc* allows differential times to aid in the development of travel time calibration data sets (e.g., Myers et al., 2011). Importantly, the errors of absolute-time and differential-time data are treated independently, and both error processes are estimated in the *Bayesloc* relocation procedure. Separating error processes is justified because correlation-based picks avoid the considerable error that is introduced by the measurement of arrival onset. This approach leads to an improved characterization of measurement errors, data weighting, and identification of outliers in the absolute-time data set.

RESEARCH ACCOMPLISHED

Bayesloc

Bayesloc is a formulation of the joint probability function that spans hypocenters, travel-time corrections, pick precision, and phase labels. Initial versions of *Bayesloc* were tailored for application to event clusters (e.g., aftershock sequences), with travel-time correction and pick precision formulations that were designed for robustness. By introducing a datum-specific travel time corrections to the travel-time correction model, Myers et al. (2011) extend *Bayesloc* to data sets that cover arbitrarily large geographic areas. Importantly, *Bayesloc* phase labels are probabilistic, and at no point is any one phase label chosen. Possible labels include all phases under consideration and the possibility that the label is erroneous. *Bayesloc* allows prior constraints on any aspect of the multiple-event system, enabling directly utilization of previous work that statistically characterizes the accuracy of event hypocenters and picks [e.g. Bondár et al. 2004; Bondár and McLaughlin 2009]. The use of prior information helps to mitigate regional location bias and improve outlier identification.

Using absolute arrival-time measurements alone, *Bayesloc* has proven to be a robust method to determine locations, corrections to travel time predictions, and assessments of arrival-time and phase-labeling errors. Because *Bayesloc* includes travel time corrections, location accuracy is predominantly determined by the arrival-time measurement precision. Until now, *Bayesloc* input was restricted to sets of individual arrival-time measurements, which are typically made by an analyst picking the onset of a phase arrival. However, the most precise arrival-time measurements are based on the correlation of seismic waveforms. The formulation below extends *Bayesloc* utilize direct input of differential times between phases, which is a proven strategy for improving location precision (e.g., Shearer, 1997; Waldhauser and Ellsworth, 2000; Zhang and Thurber, 2003; Schaff et al., 2004; Richards et al, 2006). Differential times factor into the slowness component of the travel time correction model, and differential times also influence identification of absolute-time outliers by imposing powerful constraints on relative locations and the direct comparison of absolute-time differences (and associated uncertainty) with correlation-based measurements.

Notation

We follow the notation of Myers et al. (2007, 2009, and 2011), which we now summarize and extend to differential arrival-time data.

Event-origin *parameters* are:

x_i = (lat$_i$, lon$_i$, depth$_i$) = the *location* of the *i*-th event.

o_i = the *origin-time* of the *i*-th event.

Seismic signals originating from the events are recorded at multiple stations, and we denote each station by

s_j = (lat$_j$, lon$_j$, elevation$_j$) = the *location* of the *j*-th station.

We consider two types of arrival time measurements, absolute arrival times (e.g. analyst picks) and differential arrival times (e.g. differences based on waveform correlation). For absolute times:

a_{ijk} = the *k*-th picked *absolute arrival-time* from the *i*-th event at the *j*-th station.

w_{ijk} = the *phase-label* assigned to the a_{ijk} arrival time, $w_{ijk} \in \Omega = \{1,2,\ldots, M\}$, where M is the number of phase names under consideration and each integer corresponds to a seismic phase {Pg, Pn, P, Lg, etc.}.

For differential arrival times:

d_{ii^*jk} = the *k*-th estimated *differential arrival-time* between the *i*-th and the *i**-th event at the *j*-th station.

v_{ii^*jk} = the *phase-label* assigned to the d_{ii^*jk} differential arrival time, $v_{ii^*jk} \in \Omega$.

The analyst-assigned phase-labels, w_{ijk} and v_{ii^*jk}, are not necessarily correct. As such, we denote

W_{ijk} = the *true phase-name* (unknown) of the arrival a_{ijk}.

V_{ii^*jk} = the *true phase-name* (unknown) of the differential arrival d_{ii^*jk}.

Phase label error may take two basic forms for a mislabeled phase: the correct phase is either in the phase set Ω or outside of it. To account for phases not in the set Ω and erroneous arrival data, we use a null phase-label, $W_{ijk} = 0$ or $V_{ii^*jk} = 0$, and define the extended phase label set $\Omega^* = \{0,1,2,\ldots, M\}$ (See Myers et al., 2009).

Given a proposed event location *x*, let

$F_w(x_i, s_j)$ = the *model-predicted travel-time* of phase *w* from event location x_i to station location s_j.

We further abbreviate the notation by letting $F_{wij} = F_w(x_i, s_j)$. The model-predicted travel-time is only an approximation to the true (unknown) travel-time of each phase. We therefore explicitly define,

$T_w(x_i, s_j) = T_{wij}$ = the *corrected travel-time* of phase *w* from event location x_i to station location s_j.

We will refer to a subset of parameters by simply dropping one or more subscripts. For example, a_{ij} denotes the collection (multiple-phases) of the n_{ij} arrival-times observed at station *j* from event *i*.

The Bayesloc Statistical Model

The framework is an extension of Myers et al. (2007,2009, 2011) in which the multiple-event location problem is decomposed into 3 components.

1) **Travel-Time Model.** The conditional distribution of the corrected travel-times (T) given travel-time predictions (F) and collection of travel-time correction parameters (τ);

$$p(T \mid F, \tau) \qquad (1)$$

2) **Arrival Data Model.** The conditional distribution of the arrival-time data (a) and the differential arrival-time data (d) given the origin times (o), the corrected travel times (T), phase configurations (W, V), and a collection of arrival data error parameters (σ, ρ);

$$p(a \mid o, T, W, \sigma)\, p(d \mid o, T, V, \rho) \qquad (2)$$

3) **Prior Model.** A prior distribution for hypocenter parameters, arrival data error parameters, travel-time correction parameters, and a prior distribution for phase configurations;

$$p(x,o)\, p(\tau)\, \mathrm{p}(\sigma)\, p(\rho)\, p(W \mid w)\, p(V \mid v) \qquad (3)$$

Note that we assume that absolute arrival-times and errors (a,σ) are independent from the differential arrival-times and errors (d,ρ). Similarly, we assume a prior independence between the phase labels for the two data sources (W and V).

Using Bayes' theorem, these three physically related probability models are brought together in a joint posterior distribution

$$p(o,x,T,W,V,\sigma,\rho,\tau \mid a,w) =$$
$$p(a \mid o,T,W,\sigma)\,p(d \mid o,T,V,\rho)\,p(T \mid F(x),\tau)\,p(W \mid w)\,p(V \mid v)\,p(x,o)\,p(\sigma)\,p(\rho)\,p(\tau)\,/\,p(a)\,p(d) \tag{4}$$

where $p(a)$ and $p(d)$ is the marginal distribution over the arrival data. Eqn 4 allows us to easily combine the 3 components of the hierarchical model to calculate the conditional probability for locations, travel-times, and phase-name configurations given a set of arrival data.

Myers et al. (2007, 2009, and 2011) describe the travel-time correction (Eqn 1) and absolute arrival-time data error precision (Eqn 2) models in detail. Summarizing, the travel-time correction is given by

$$\delta_{wij} = T_{wij} - F_{wij} = \alpha_w + \alpha_j + \alpha_{wj} + \beta_w G_{ij} \tag{5}$$

Where α_w and β_w are broad-area, phase-specific shift and scaling parameters, G_{ij} is the event-station geographic-distance, while the remaining α terms are station and station-phase specific terms that are meant to capture small-scale travel-time adjustments (see Myers et al., 2011, for details). This particular version of the travel-time correction term is well suited for a cluster of events. An extension of the travel-time correction to broader region is given in Myers et al. (2011), which adds event-specific corrections.

Similarly, the treatment of the absolute arrival-time data is unchanged from previous versions of *Bayesloc*. Absolute arrival errors are assumed to be Gaussian distributed with a mean of $A_{wij} = o_i + T_{wij}$, the expected (corrected) arrival time for an assumed phase $w = W_{ijk}$, and variance

$$Var(a_{ijk} - A_{wij}) = 1/\kappa_{wij}, \text{ where } \kappa_{wij} = \kappa_w \kappa_i \kappa_j \kappa_{wi} \kappa_{wj} \tag{6}$$

The differential data is treated similarly. We first note that the expected differential arrival-time, for an assumed phase w, is

$$D_{wii^*j} = A_{wij} - A_{wi^*j} = (o_i - o_{i^*}) + (F_{wij} - F_{wi^*j}) + \beta_w (G_{ij} - G_{i^*j}) \tag{7}$$

That is, the differential data does not provide any information about the broad-area phase-specific shift, α_w, nor the station-specific corrections, α_j and α_{wj}. Hence, the expected differential arrival time is less influenced by errors in the assumed travel-time model (F), particularly for two nearby events. Given the expected differential arrival-times, we assume that the differential arrival-time residuals are Gaussian distributed with mean zero and a relatively simple mode for the variance,

$$Var(d_{ii^*jk} - D_{vii^*j}) = 1/\phi_{vii^*j}, \text{ where } \phi_{vii^*j} = \phi_v \exp(\theta C_{ii^*jk}) \tag{8}$$

for an assumed phase v, where ϕ_v are phase-specific precision parameters, C_{ii^*jk} is the recorded cross-correlation associated with the differential arrival-time d_{ii^*jk}, and θ is an unknown parameter to be estimated. The statistical model for the variance can be easily extended to accommodate other information related to the precision of the differential arrival-data.

Finally, the prior distribution for the origin parameters, travel-time correction parameters, phase labels, and the precision parameters associated with the absolute arrival-time data are unchanged from Myers et al. (2007 and 2009). The only addition here is the prior model for the phase labels of differential arrival-time data and the precision parameters of the differential arrival-time residuals. The prior model for the phase-labels of differential arrival-time is taken to be of the same format as that for the arrival-time data. And similarly, the prior for the precision parameters is specified to be vague.

Markov Chain Monte Carlo (MCMC) for Posterior Inference

MCMC sampling is used to generate realizations from the joint posterior distribution of all multiple-event model parameters. MCMC sampling is well established as a method of parameter estimation and uncertainty characterization (e.g., Gelman et al., 2004). The sampler used in *Bayesloc* is described in Myers et al. (2007 and 2009), with the addition of folding in the likelihood of the differential arrival-time data where applicable. For example, when a Metrapolis random-walk sampler is used to propose a new lat-long location for a given event, the probability of acceptance reflects both how well the new location fits the absolute arrival-time data and the available differential arrival-time data. Because differential time data provide a strong constraint on relative locations, we have found it necessary to jointly sample locations of events for which correlated data are available. Likewise, differential time data provide strong constraints on the slowness (β) component of the travel time adjustment model, requiring joint sampling of that parameter. The joint (correlated) sampling of epicenters and travel time corrections is a large change to the Metropolis-Hastings and Gibbs sampling routines (respectively) that were used in previous versions of *Bayesloc*.

Correlation-Based Differential Arrival Times

We adopt widely used methodologies for computing differential times based on waveform cross correlation. We first collect all waveforms for a given station and event cluster. A user-specified, phase-specific bandpass filter is applied to each waveform and the phase window is cut from the seismogram based on either analyst picks or a theoretical arrival time. We then compute and save the Fourier transform of each phase-windowed seismogram. Complex multiplication in the frequency domain is used to compute auto correlation and cross correlation spectra. Cross correlation spectra are inverse transformed and normalized based on the average of the autocorrelation amplitudes to produce a normalized, time-domain correlation function. The correlation coefficient is the peak of the correlation function and the time shift is the offset of the peak from the center of the correlation function. The time shift and the correlation coefficient are refined by fitting a parabolic function to the sample points in the neighborhood of the peak in the time-domain correlation function (Deichmann and Garcia-Fernandez, 1992), which allows sub-sample precision for the correlation pick. The difference in time between the arrivals is computed by differencing the start time of the phase windows and adding the correlation-based time shift. This process provides input to Bayesloc of the form: eventID_1 eventID_2 station phase time_difference correlation_coefficient.

In addition to direct input of time differences, we have also implemented the method of optimal adjustments to absolute times based on multi-channel cross correlation (Vandecar and Crosson, 1990). Although this method does not mitigate pick bias, it does improve the overall pick precision and aids outlier removal prior to the *Bayesloc* inversion. It is our practice to input both the absolute-time and differential-time data into *Bayesloc* because of both the differing constraints on the location system that are provided by each type of data and the increased error for the absolute-time data.

CONCLUSIONS AND RECOMMENDATIONS

We have extended the statistical formulation of *Bayesloc* to include differential times between phases. Extension to differential time data is expected to improve relative locations in most cases, and it may improve absolute location by helping to identify outliers and by improving error characterization for the absolute-time portion of the data set.

A beta version of the code is close to completion and we expect to have example locations in the near future. The beta version features correlated MCMC sampling of event locations and travel time corrections, both of which are necessitated by the unique constraints imposed by differential-time data. To generate differential time data sets, we have implemented a code to determine differential times based on waveform correlation. The code is integral with the LLNL database and allows users to easily compute differential arrival times and correlation coefficients by entering a list of event identifiers, stations, and phases (with corresponding bandpass).

REFERENCES

Bondár, I., S.C. Myers, E.R. Engdahl, E.A. Bergman, Epicenter accuracy based on seismic network criteria, *Geophys. Jour. Int*. 156**:** 483–496, 2004.

Bondár, I., K.L. McLaughlin. A New Ground Truth Data Set For Seismic Studies, Seismological Research Letters Volume 80, Number 3, May/June 2009.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin, D.B. (2004). Bayesian Data Analysis (2nd ed.). Boca Raton, Florida: Hapman and Hall/CRC.

Kennett, B.J.N., E.R. Engdahl and R. Buland (1995). Constraints on seismic velocities in the Earth from traveltimes, *Geophys. J. Int* 122: 108–124.

Richards, P.G., F. Waldhauser, D. Schaff and W-Y Kim (2006). The applicability of modern methods of earthquake location, *Pure appl. geophys*., 163: 2-3, 351–372.

Ruppert, S., D., Dodge, A. Elliott, M. Ganzberger, T. Hauk and E. Matzel (2005). Enhancing seismic calibration research through software automation and scientific information management, in *Proceedings of the 27th Seismic Research Review: Ground-Based Nuclear Explosion Monitoring Technologies,* LA-UR-05-6407, Vol. 2, pp. 937–945.

Menke, W., and D. Schaff (2004), Absolute earthquake locations with differential data, *Bull. Seismol. Soc. Am*., 94: 2254–2264, doi:10.1785/0120040033.

Myers, S.C., G. Johannesson, and W. Hanley, Incorporation of probabilistic seismic phase labels into a Bayesian multiple-event seismic locator, *Geophys. J. Int*. 177: 193–204, 2009.

Myers, S.C., G. Johannesson, and W. Hanleyv (2007). A Bayesian hierarchical method for multiple-event seismic location, *Geophys. J. Int*. 171: 1049–1063.

Schaff, D., G.H.R. Bokelmann, W.L. Ellsworth, E. Zanzerkia, F. Waldhauser and G.C. Beroza (2004). Optimizing correlation techniques for improved earthquake location, *Bull. Seismol. Soc. Am.* 94: 705–721.

Shearer, P. (1997). Improving local earthquake locations using the L1 norm and waveform cross correlation: Application to the Whittier Narrows, California, aftershock sequence, *J. Geophys. Res.* 102: 8269–8283.

Zhang, H. and C.H. Thurber (2003). Double-difference tomography: The method and its application to the Hayward fault, California, *Bull. Seismol. Soc. Am.* 93: 1875–1889.