UTILIZING PRIOR INFORMATION FOR DEPTH TO IMPROVE SEISMIC EVENT DISCRIMINATION

Deborah K. Carlson[1], Kelly Kane[2], Rick Schult[2], Sandy Ballard[3], Chris Young[3], and Dale Anderson[1]

Pacific Northwest National Laboratory[1], Air Force Research Laboratory[2], and Sandia National Laboratories[3]

Sponsored by National Nuclear Security Administration
Office of Nonproliferation Research and Engineering
Office of Defense Nuclear Nonproliferation
and
Air Force Research Laboratory

Contract No. DE-AC06-76RLO 1830

ABSTRACT

We have developed and tested a novel algorithm for estimating the depth of a seismic event in order to improve the discrimination of events. Information from the algorithm can be incorporated into a statistically based discrimination framework to determine the source of an event. The depth estimation approach differs from currently used algorithms, which use non-linear regression techniques, by using Bayesian techniques to incorporate constraints, or prior information about the depth of an event. We demonstrate this proof-of-concept algorithm with first arriving P-waves and their associated modeled travel times. The likelihood is constructed with Gaussian errors. Depth may be constrained with a skewed distribution if characteristics of the waveforms from an event indicate that bounds on depth are appropriate. For instance, the *Rg* phase is present in a waveform only when an event is shallow. A high confidence *Rg* phase in one or more defining waveforms can lead one to assume a shallow-skewed prior distribution for the depth parameter.

OBJECTIVE

For many seismic events, depth and origin time are the hypocentral parameters that are most poorly constrained because of the source-receiver geometry imposed by the Earth. It is not unusual for current location algorithms to return event solutions that fit the data very well and yet have event depths that are above the surface of the earth (so called "air quakes") or well below the known limits of seismicity for a given area. The solutions are statistically valid in that the confidence bounds are large enough to encompass more realistic depths the specified percent of the time, but they are unsatisfying to seismologists. The effect of repeatedly seeing such unreasonable depth estimation is to develop a mistrust of the depth determinations in general, even when depth may be well constrained. What is needed is a means to flexibly incorporate a priori information about acceptable depth distributions. This will better constrain the hypocentral depth estimates when they are poorly controlled by the data, and let the data control the depth estimate when the data have good depth control. Our research has developed an algorithm that shows promise in achieving these depth estimation properties.

The single event hypocenter location model is

$$t_i = \tau + T_i(\mathbf{s}) + \varepsilon_i \,,\ i=1,2,\ldots,n \qquad (1)$$

where t_i is the arrival time at the i^{th} station, τ is the event origin time, T_i is the travel time from the event located at $\mathbf{s} = [x, y, z]^T$ to the i^{th} station, and $\varepsilon_i \sim \text{iid } N(0, \sigma^2)$.

Regardless of the solver used to estimate **s**, it is possible with the non-linear regression formulation to estimate *z* as a negative number (an airquake), or to get an unreasonably large estimate of the depth given known seismicity. One approach to correcting an air quake is to simply set negative estimates of *z* to zero (the surface). Our efforts focused on the incorporation of constraints on *z* in the non-linear regression formulation. Fully mature research will provide a general mathematical framework that allows one to incorporate constraints (or whatever prior information exists) into the estimation problem for all of the parameters.

The research and development presented here assumes the origin time is known without error. Additionally, we have assumed no prior information on latitude or longitude, (y, x). Two different prior distributions for depth were considered: the first a uniform distribution, which assumes equal probability for all possible depths within a reasonable range; the second a shallow-skewed distribution which assigns the greatest probability to very shallow depths.

The test case for this proof-of-concept effort is an event near the coast of Japan whose hypocenter estimation has been difficult to obtain with commonly used location algorithms. The dataset consists of first arriving P waves. Figure 1 shows the 14 locations of stations which observed this event.

RESEARCH ACCOMPLISHED

Mathematical Formulation

The general Bayesian equation is $p(\mathbf{s}|\mathbf{t},\theta) \propto \pi(\mathbf{s}|\theta) f(\mathbf{t}|\mathbf{s})$. Here f is the likelihood, which we assume to be Gaussian, and π is the prior distribution; **s** represents the vector of parameters, **t** is the vector of observed arrival times, and θ is the vector of hyperparameters in the prior distribution defined by physical-basis constraints. Ultimately we are interested in the distribution of the depth given the data, or $p(z|\mathbf{t},\theta)$. For the model given in Equation (1), the parameters are $\mathbf{s} = [x, y, z]^T$, and σ^2. Because we assume that origin time is known without error, τ is not considered a parameter, but a known value. Future work will incorporate τ as an unknown parameter. We can rearrange Equation (1) so that

$$\varepsilon_i = t_i - \tau - T_i(\mathbf{s}),\ i = 1, 2, ..., n.$$

Now the likelihood can be written

$$f(x, y, z, \tau, \sigma) = \prod_{i=1}^{n} \Phi\left(\frac{t_i - \tau - T_i(x, y, z)}{w_i \sigma}\right),$$

where $\Phi(\cdot)$ is the Gaussian cumulative distribution function (CDF). The w_i are weights associated with a data quality measure, such as *deltim* in SeaLoc or signal-to-noise.

The prior distribution, $\pi(\mathbf{s}|\theta)$, is specified as the product of individual priors $\pi(x|\theta)$, $\pi(y|\theta)$, $\pi(z|\theta)$ so that $\pi(\mathbf{s}|\theta)= \pi(x|\theta)\ \pi(y|\theta)\ \pi(z|\theta)$. While this independence-of-priors formulation might be a simplification, we have used it in our proof-of-concept effort.

We have considered two prior distributions for depth: the first a uniform distribution; the second a beta distribution with parameters selected to give the highest probability to shallow depths. The hyper parameters for the beta distribution define a density with the mode at approximately 3 km deep (Figure 2 shows this distribution). Longitude and latitude are constrained with a uniform distribution over a reasonable range of coordinates.

Results

Results are presented graphically for both prior distributions on z and using the uniform priors for x and y (Figures 3 and 4). The results are "standardized". This standardization is not that of proper probability density construction; however it results in a common scale for the plots. Depths between 0 and 110 km, by 5 km are presented. The figures show only the interesting subset of depths for each z prior distribution.

The least-squares solution is shown as a reference point in each plot as a black dot. The least-squares solution for depth is 100 km. The posterior mode solution is shown by a black asterisk on the depth plot where it occurs.

Under a uniform prior for depth, the maximum posterior density is at 50 km (Figure 3). Not surprisingly, when a shallow-skewed prior is used, the maximum density is at a much shallower depth, 10 km (Figure 4). The mode travels from the northeast to the southwest as the depth increases; if a fixed (x, y) location were selected, the non-linear behavior of the travel times would be evident as curved contours across depth. With both priors, the mode of the posterior is achieved at a location southeast of the least-squares location estimate.

CONCLUSIONS AND RECOMMENDATIONS

Our proof-of-concept work has demonstrated that a Bayesian formulation of depth estimation can provide physical-basis constraints on the estimate. Next steps include

- seismicity studies to asses the implications and veracity of the new hypocenter locations derived from the developed Bayesian algorithm
- the calculation of the marginal distribution (integrating constant) in order to estimate highest posterior density (HPD) regions, which are analogous to confidence regions, for the location parameters,
- the incorporation of the origin time as an unknown parameter,
- research to determine the influence of the prior distribution on depth when a deep-skewed prior is utilized,
- characterization of the hypocenters for a simulated set of data where the true hypocenter is known, and
- sensitivity analysis with regard to the number of data points needed to overcome the prior distribution.

REFERENCES

Ballard, S. (2004), Seismic event location: dealing with multi-dimensional uncertainty, model non-linearity, and local minima, in *Proceedings of the 25th Seismic Research Review—Nuclear Explosion Monitoring: Building the Knowledge Base,* LA-UR-03-6029, Vol. 1, pp. 193-202.

Carlin, B. .P and T. A. Louis (2000), *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman and Hall, New York.

Kennett, B. L. N. and E. R. Engdahl (1991), Traveltimes for global earthquake location and identification, *Geophy. J. Int.* 105, 405-465.

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

Lay, T. and T. C. Wallace (1995), *Modern Global Seismology,* Academic Press, San Diego.

Robert, C. P. and G. Casella (1999), *Monte Carlo Statistical Methods*, Springer, New York.

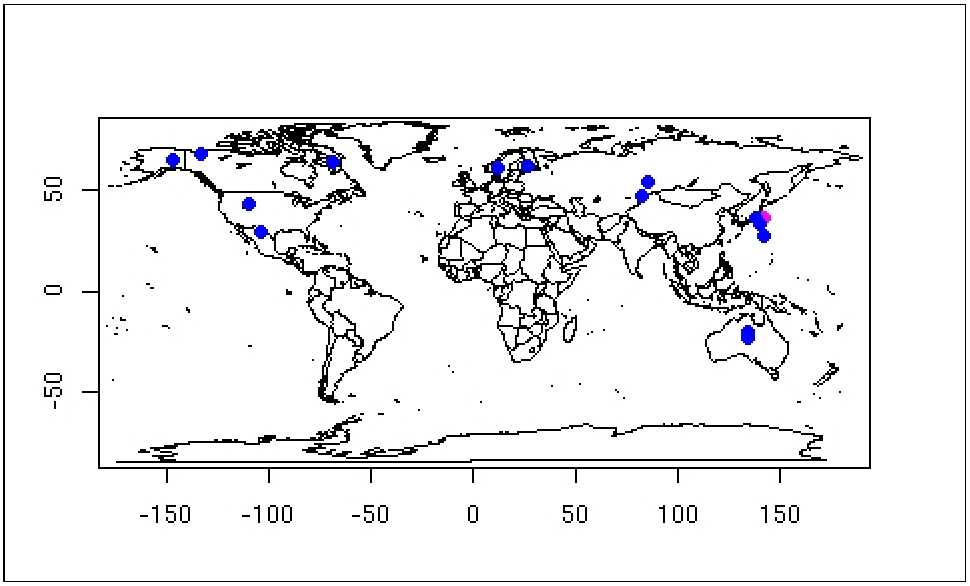

Figure 1. Map showing the estimated location of the event via (pink) least-squares and the stations which observed it (blue).

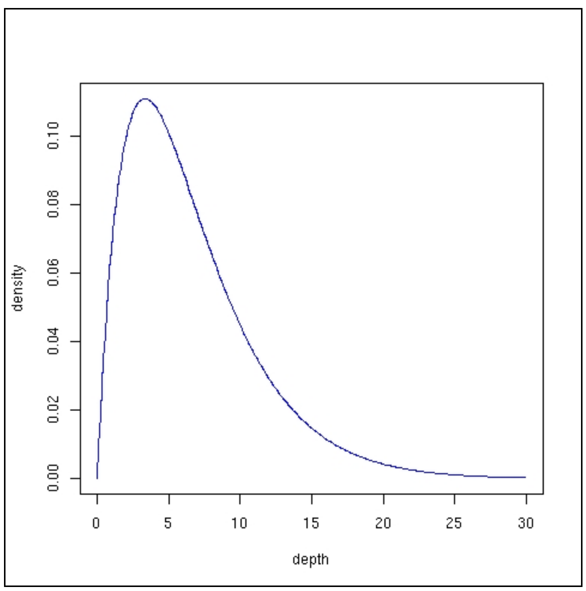

Figure 2. Shallow-skewed beta distribution for the depth parameter.

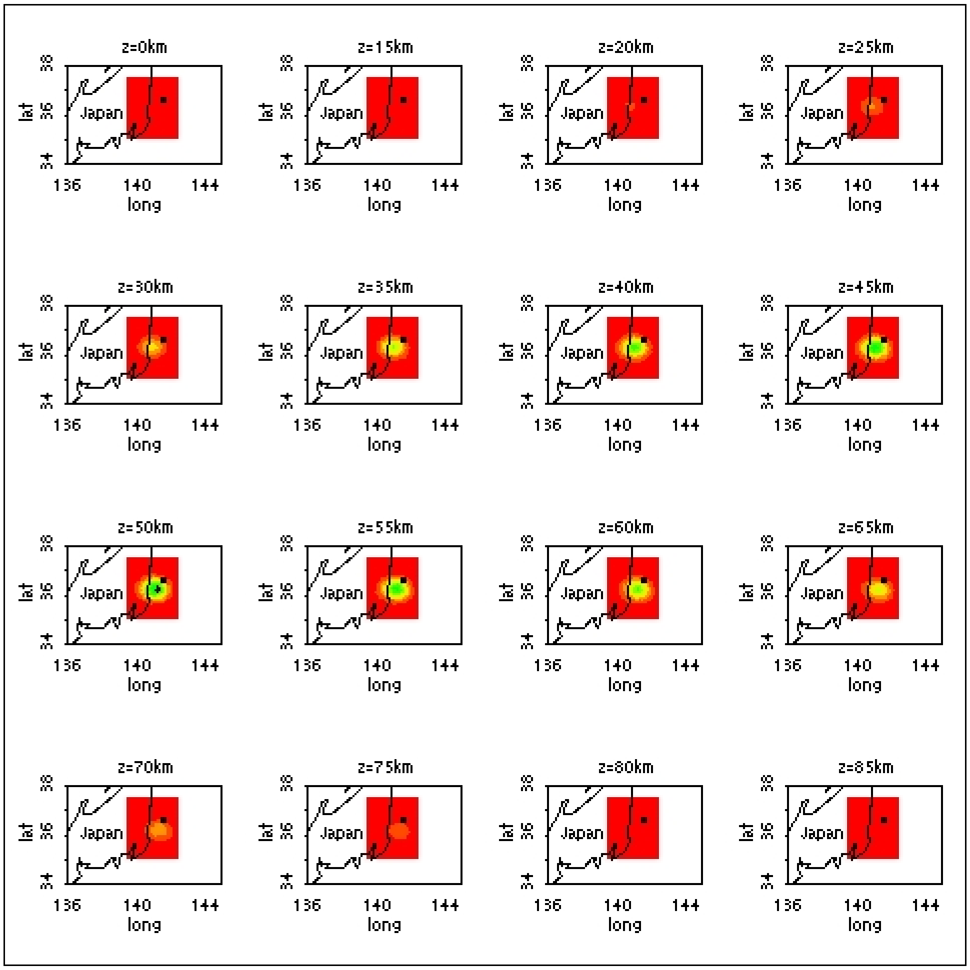

Figure 3. Standardized posterior probabilities for a subset of depths when a uniform prior distribution is used. Yellow to green colors indicate high probabilities.

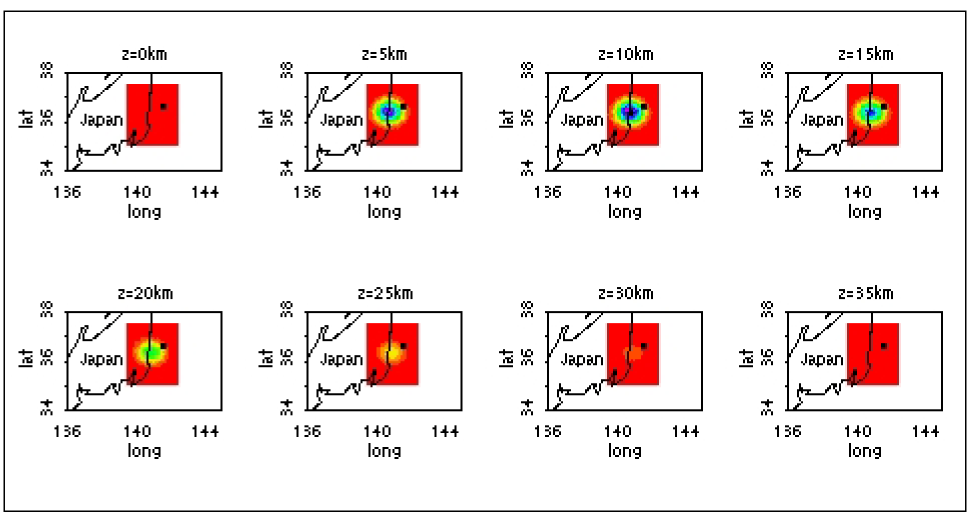

Figure 4. Standardized posterior probabilities for a subset of depths when a shallow-skewed prior distribution is used. Yellow to green to blue indicate high probabilities.