Event origin depth uncertainty - estimation and mitigation using waveform similarity.

Anton Biryukov[1], Evgeny Chzen[2], Jan Dettmer[1], David Eaton[1]

[1] Department of Geoscience, University of Calgary, Calgary, AB, Canada
[2] LAMA, UMR-CNRS 8050, Université Paris Est – Marne-la-Vallée

Summary

The induced seismicity event localization using inverse kinematic algorithms is still a problem requiring special attention. Often a catalog of events detected and located using limited on-surface instrumentation lacks the precision of origin depth. Such errors can hinder the interpretation of activated zones using seismicity as a proxy.

The location constraints are affected by the inaccuracy of the velocity model, limited acquisition geometry and the assumptions inherited by the location algorithm, among other factors. The issue becomes even more aggravated when only surface arrays are employed, where the geometry strongly favors the accuracy of epicentral location as opposed to depth location. The reported depth errors are sometimes comparable to a formation thickness or even the origin depth itself. The passive seismicity is often located using the P and S first breaks only. Finding additional features of the seismic signal that could be informative of its origin can be a first step towards constraining the depth uncertainty.

Therefore, the purpose of this study was two-fold. First, we characterized the uncertainty in the event origin due to the inaccuracy in the effective velocity model using Monte-Carlo simulations. We show that a low velocity zone (LVZ) can cause the non-uniqueness and a spread of the solution over a depth range. Subsequently, seismo-grams from a set of synthetic earthquakes were simulated spanning the depth range of 2-5 km, covering LVZ. By varying the focal mechanism and event origin, we numerically generate a bank of waveforms corresponding to the events with known locations. A set of classifiers is trained on the bank to predict the event location with re-spect to LVZ based on arrival times and statistical features of the signal waveforms. We demonstrate that adding several features of the signal, descriptive of its origin can improve the location depth constraint, as opposed to using arrival times only as predictor variables.

Introduction

The earthquake localization on a regional scale using inverse kinematic algorithms is still a problem requiring special attention. The on-surface seismic networks for monitoring induced seismicity provide a sufficient location accuracy for the earthquake hypocenter, but often offer very limited resolution in depth (Eisner et al. [2009]). For computational convenience and due to the lack of information, the velocity model used for location is often parametrized as a function of depth only. In some cases such an approximation is a very strong assumption, and it is important to consider several different possible 1D structures to test the sensitivity of the location to errors in the velocity model. The sparsely distributed on-surface instrumentation and insufficient azimuthal or vertical coverage can aggravate the location uncertainty due to the limited direction of the ray-paths it is able to receive (Eisner et al. [2010]). As a result, in the absence of downhole observations, the reported uncertainty of the depth location using on-surface array only may be comparable to the origin depth value itself, such that the location of the event becomes untrustworthy to the analyst. On the other hand, in the daily industry practice the event origins are often derived based on the deterministic methods that use only 2 characteristic points (P- and S-wave arrivals) per sensor per event. The information transmitted by the event is obviously not limited to two-points only, and therefore should not be discarded when the location is performed. Improving event location can be done through finding informative features of the seismic signal that could help better constrain the location.

We start with a rigorous study of the effect of the uncertainty in the velocity model on the P- and S-wave travel time error. The ray-tracing method is employed to predict the time of the arrivals for the two phases at station location. The Monte-Carlo simulation is then performed with varying velocity model to estimate the mean and

standard deviation of the arrival time. These values are consequently used as a proxy for an event's P- and S-wave arrival time uncertainties. Using non-linear location methods (Lomax et al. [2001]), the arrival times from the Monte-Carlo simulations can be translated into spatial uncertainties and thus give an error estimate due to uncertainty in the velocity model, given an event location.

Subsequently, we simulate the synthetic seismograms from the virtual sources spanning the depth range of 2-5 km. A set of characteristic functions is applied to the resultant traces in order to extract features describing the origin. Later, a classification of the event location is carried out on two separate feature sets: (i) arrival times only and (ii) arrival times bundled with the signal features. With this comparison we highlight that the travel time data may be insufficient to constrain the origin of the event due to the velocity model accuracy while adding the waveform features has a potential to constrain the origin depth.

Methods

Generalized workflow

This report is particularly focused on estimating the depth error due to the velocity model uncertainty. That is, we want to evaluate the jitter in the hypocenter location due to inaccuracy in the approximation of the true velocity model as a 1D layered cake model derived from V_p and V_s logs. Specifically, we would like to simulate the travel times from many virtual events with a fixed origin but variable velocity model. These travel times are then used for relocating the events, that should potentially result in a cloud centered at the origin. The spatial dimensions and the shape of the cloud can give a quantitative insight into how sufficient the 1D velocity model assumption might be for a particular region and acquisition geometry. Therefore, the general workflow can be summarized in several key steps as follows: (i) determine the background (expected) velocity model from the sonic logs, (ii) carry out Monte-Carlo (MC) simulations for the travel time sets in a perturbed velocity model for an event with a fixed origin, (iii) locate the event corresponding to the set using background velocity model, and (iv) analyze the PDF of the obtained distribution.

Figure 1. An example of the distribution of velocity profile used in the MC simulations. The color represents the values of PDF of the velocity distribution. Note the limited variation of the layer above LVZ due to the creation of the shadow zone.

Monte - Carlo travel time simulations

As no anisotropy or 2D elastic profile information was available, the baseline velocity model was approximated as a 1D laterally homogeneous layered medium. The elastic parameters for the layers were derived from the sonic wellbore logs. The velocity uncertainty in this study is modeled through applying a random normal perturbation of V_p and V_s within a layer, while keeping the thickness of the layers constant. The standard deviation of the perturbation is a fraction of the velocity value within the layer. An example of the velocity profile distribution with the 6% standard deviation is shown in Figure 1. In the case of a 1D layered medium the classic seismic ray theory can produce an exact solution at a very low computational cost. Therefore, a Python-based implementation of a ray tracing method was chosen as a best fit-for-purpose method.

Event origin location and synthetic waveform modeling

In the case of a limited recording geometry and a velocity model with sharp horizontal interfaces, a probabilistic direct global-search procedure was shown to outperform the linearized location methods and adequately determine the complete location PDF of the event, converging to a global maximum (Lomax et al. [2009]). Therefore, we opted for employing the probabilistic, non-linear, global-search earthquake location algorithm implemented in the NonLinLoc software Lomax et al. [2001].

The synthetic waveforms used for event origin classification were produced using the wavenumber integration method implemented in CPS suite (Herrmann [2013]). For the particular case of 1D media, CPS allows one to obtain the full waveforms much quicker than FD methods commonly used for similar purposes. For each of the source-receiver pairs simulated in the model we generate a database of Green's functions. The Green's functions are then convolved with the source wavelet of the given dominant frequency $f = 2.5$ Hz and a given moment tensor $M_{i,j}$ corresponding to a pure explosive source (identity matrix). To achieve variability in the signal
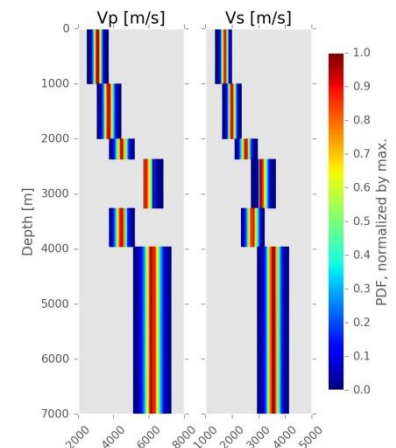
waveforms and mimic realistic amplitude uncertainty, the moment tensor was perturbed with a Gaussian noise:

$$M_{i,j} = 1 + N_{i,j}(0, \sigma_m), \tag{1}$$

where $\sigma_m = 0.1$ is the standard deviation of the perturbation. Subsequently, a set of features is extracted from the waveforms in order to bring more information about the origin of the signal. Prior to the feature extraction, the signals are all normalized to a unit amplitude. We then obtained the envelope of the signal on the vertical channel using Hilbert transform and chose the following features as the proxy:
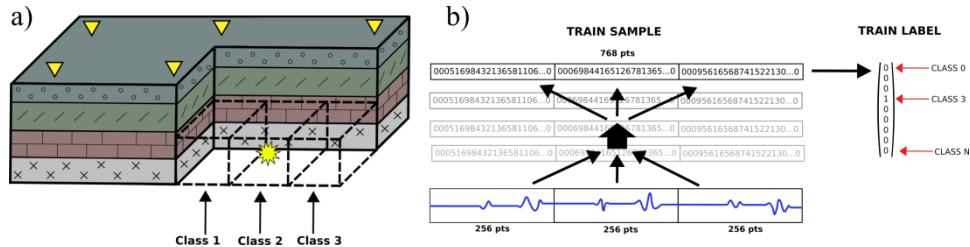

Figure 2. An illustration of the signal classification concept. The signals originating from the locations on the grid (a) are then transformed into a feature space and labeled by the class containing the event origin.

- kurtosis of the signal envelope - to show the peaked-ness of the signal and discern between signals with strong and weak multiple reflections,

- mean value of the envelope,

- standard deviation of the envelope - to show the variability in the envelope,

- area under the envelope,

- the ratio of the signal energy contained between t_p and t_s with respect to the whole trace

- number of zero-crossings of the envelope derivative - to count the number of local extrema of the envelope

The procedure of modeling and feature extraction is then repeated over the set of values for the moment tensor $M_{i,j}$ and locations covering the range of 2000-5000 meters.

Signal classification

The concept of the classification is briefly illustrated in Figure 2. The zone of interest is first meshed into nodes that can host virtual sources. The signal from an event originating at a node is transformed into the feature space and is labeled according to the location of the grid. The signal database has been divided into 3 classes by the origin depth of the event: (i) above LVZ, (ii) within LVZ, and (iii) below LVZ. The positions were marked according to their relative location with respect to the LVZ as demonstrated in Figure 3. For each position of the virtual source 150 simulations with different values of $M_{i,j}$ were carried out. The database was later split into the training (70% of the total data) and the test datasets (30% of the total). A set of classification algorithms was trained on the training data in order to predict the focal depth of the event (its class) given a set of features of an event of interest from the test data. Therefore, a comparison of prediction accuracy was done between a non-generalizing algorithm - K-nearest-neighbors (KNN,Cover and Hart [1967]), and linear classifiers - logistic regression (Jr and Lemeshow [2004]) and Support Vector Classifier (SVC) with a linear kernel (Guyon et al. [1993]). The regularization for the logistic regression and SVC, and the number of neighbors for KNN were tuned through exhaustive grid search over hyperparameter space using 5-fold cross-validation.
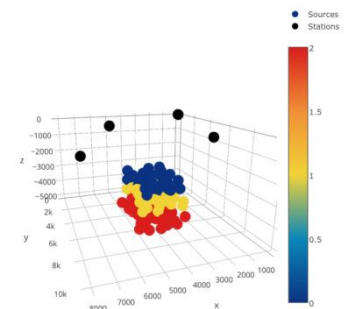

Figure 3. The relative virtual source-receiver locations for the classification experiment. The colorscale indicates which class an event belongs to.

Our interpretation here is limited to the analysis of the confusion matrix. The element of the confusion matrix $c_{i,j}$ shows the number of observations from class i predicted into class j.

Results

Origin uncertainty estimation

The origin depth uncertainty was estimated for 6% velocity perturbation. For both simulations the true location of the event was set to $(X, Y, Z) = (4000, 6000, 3910)$. The location centroids of MC simulated events for the model are shown Figure 4. The X and Y coordinates of the event are centered at their true values, whereas the depth of the event shows a bi-modal distribution, with stronger mode centered 240 meters above the depth. The weaker mode, however tends to center around the true depth of the event.
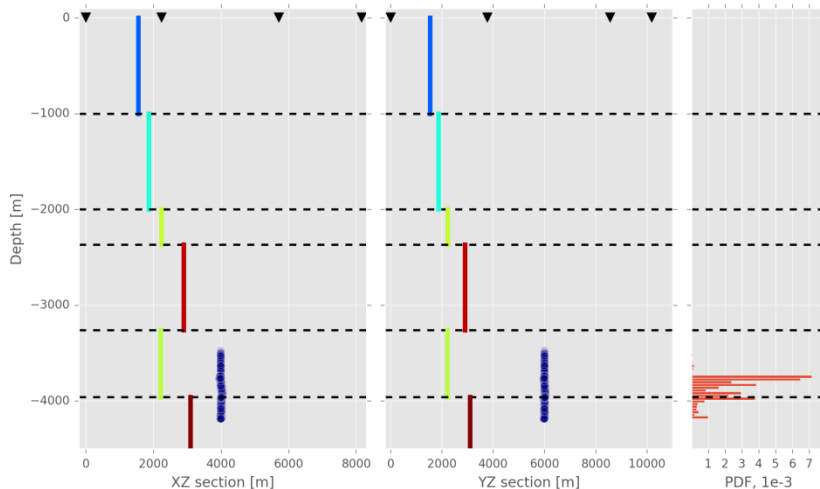

Figure 4. The results of the location of 6% velocity perturbation Monte-Carlo simulated virtual events describing the true event at $(X, Y, Z) = (4000, 6000, 3910)$. The receivers are shown in black triangles, the virtual event centroids - in blue circles.

Signal origin classification results

The classification has been carried out on the database of 90 different earthquake origins distributed uniformly along the depth range of 2-5km and covering the area around the central Fox Creek seismicity cluster as shown in Figure 5. Here we limit ourselves to a three-class classification based on the proximity of the origin to the LVZ as indicated by the color of the event origin in Figure 3. The classification has been carried out for two feature sets: (A) the travel times for P- and S-phases at each station (a total of 8 features) and (B) the features from A bundled with the signal features extracted from the vertical channels as described in Methods section. The confusion matrices for the optimally set-up classifiers are shown in Figure 6. One may see that KNN being
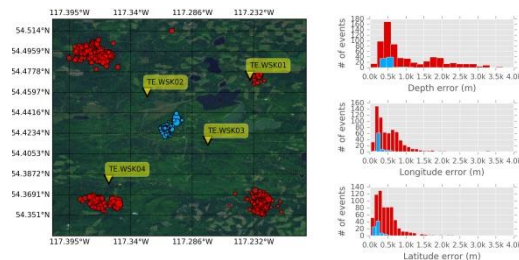

Figure 5. The distribution of induced seismicity in Fox Creek area for the events of winter 2015. The central cluster modeled in this study is shown in cyan.

References

T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1): 21–27, January 1967.

Leo Eisner, Peter M. Duncan, Werner M. Heigl, and William R. Keller. Uncertainties in passive seismic monitoring. *The Leading Edge*, 28(6):648–655, 2009.

Leo Eisner, B. J. Hulsey, Peter Duncan, Dana Jurick, Heigl Werner, and William Keller. Comparison of surface and borehole locations of induced seismicity: Comparison of surface and borehole locations. *Geophysical Prospecting*, 58(5):809–820, September 2010.

I. Guyon, B. Boser, and V. Vapnik. Automatic capacity tuning of very large vc-dimension classifiers. In *Advances in Neural Information Processing Systems*, pages 147–155. Morgan Kaufmann, 1993.

R. B. Herrmann. Computer Programs in Seismology: An Evolving Tool for Instruction and Research. *Seismolog- ical Research Letters*, 84(6):1081–1088, November 2013.

David W. Hosmer Jr and Stanley Lemeshow. *Applied Logistic Regression*. John Wiley & Sons, October 2004. Google-Books-ID: Po0RLQ7USIMC.

Anthony Lomax, Aldo Zollo, Paolo Capuano, and Jean Virieux. Precise, absolute earthquake location under Somma–Vesuvius volcano using a new three-dimensional velocity model. *Geophysical Journal International*, 146(2):313–331, August 2001.

Anthony Lomax, Alberto Michelini, and Andrew Curtis. Earthquake Location , Direct, Global-Search Methods. In *Encyclopedia of Complexity and Systems Science*, pages 2449–2473. Springer, 2009.