Improving out of network earthquake locations using prior seismicity for use in earthquake early warning

Amy Williamson [*1], Angela Lux[1], Richard Allen[1]

[1]*Berkeley Seismological Laboratory, University of California, Berkeley, Berkeley, California, U.S.A.*

[*]*Corresponding author: amy.williamson@berkeley.edu*

Abstract

Timely alerts sent through earthquake early warning (EEW) programs allow those alerted to take protective actions to mitigate their risk from potentially damaging shaking. Over the past few years, ShakeAlert, the EEW program focused on the west coast of the contiguous US has grown, alerting communities within California, Oregon, and Washington for earthquakes where damaging shaking is expected. ShakeAlert uses a set of algorithms including the point source algorithm, EPIC to determine the location, magnitude, and origin time of potential earthquakes. While EPIC produces low-latency and low error solutions for many events originating within the seismic network on land, numerous recent small earthquakes rupturing offshore of northern California have EPIC location solutions with high error (> 50 km compared to USGS locations). Because most events are occurring offshore, there is a limited number of stations that can trigger and contribute information in a timely manner for use in earthquake early warning. In order to better constrain location solutions in this region, we propose to include information about contemporary past seismicity into EPIC's grid search algorithm through a Bayesian framework. This prior information layer down-weights high error locations where EPIC's proposed event location coincides with an area of low prior seismicity in preference for locations with a similar level of

data fit that also have higher past seismicity. This addition to EPIC lowers the mean location error offshore northern California from 58 km to 14 km.

Key Points

Revised grid search technique produces lower average error earthquake locations for edge network earthquakes.

The new model uses recent seismicity as a weight, reducing high error locations.

These better location estimates also improve magnitude estimates.

Introduction

Earthquake early warning (EEW) is the rapid detection and alerting of regions that are expected to experience damaging ground motion from an earthquake (Allen and Melgar, 2019). EEW systems need to quickly detect the location and magnitude of an earthquake, estimate ground motion intensities, and send alerts to end users. An efficient system leveraging a dense seismic network can provide seconds to tens of seconds of warning. While brief, this amount of time is enough for those alerted to take simple but effective protective measures such as to drop, cover, and hold on. Automated systems also benefit greatly from EEW. With seconds of notice, high speed trains can slow, reducing the risk of derailment and allowing for an assessment of the track ahead. For example, California's Bay Area Rapid Transit system has been an early adopter of

EEW alerts for their transit system (Strauss and Allen, 2016). Japan's high speed Shinkansen trains also utilize an EEW system to mitigate risk (Nakamura and Saita, 2007b).

Through partnerships, the United States Geological Survey (USGS) and participating academic institutions have developed and implemented ShakeAlert, an EEW system spanning California, Oregon, and Washington. Since 2012, ShakeAlert has been issuing alerts to pilot users and select partnering institutions. By 2018, public alerts were made available in California and as of 2022, ShakeAlert issues public alerts across the entire US West coast. Communities can receive phone alerts through the MyShake App for earthquakes with a magnitude greater than 4.5 and a modified Mercalli intensity (MMI) ≥ III (Strauss et al. 2020). For regions with an expected shaking of MMI ≥ IV and earthquake magnitude greater than M5, users may also receive a wireless emergency alert (WEA; Kholer et al., 2020). A recent example of an event that prompted the issuance of both a WEA and MyShake App alert was the 25 October, 2022 M5.1 Alum Creek, California earthquake, rupturing just southeast of San Jose, California.

ShakeAlert operates using three operational layers to handle incoming seismic data, production, and outgoing alerts. The data layer ingests and handles ground motion data from seismic stations participating in the ShakeAlert network. These data are then sent to the production layer where EEW specific algorithms, such as EPIC (Chung et al, 2019), and FinDER (Böse et al., 2015, Böse et al., 2018) are housed. These algorithms detect earthquakes, solve for location and magnitude, and estimate the intensity and extent of the expected shaking. The alert layer analyzes the EEW solution and sends a product if the event passes set quality checks, has a large enough magnitude, *and* has high enough expected ground motion to merit an alert. All EEW processing occurs on redundant servers located in Seattle, Menlo Park, Berkeley, and Pasadena. This allows for operational continuity in the event of a data outage at any site.

In this study, we propose bEPIC, an improvement to the methodology of the EEW algorithm EPIC, with the goal of limiting high location error solutions for offshore or edge of network earthquakes. We employ a Bayesian framework to modify EPIC's currentgrid search location algorithm, including the contemporary seismic history over the western US as prior information. We test bEPIC on a catalog of recent earthquakes within the ShakeAlert region. The revised earthquake locations through bEPIC are compared against historic EPIC solution and performance improvement is calculated with respect to the final USGS solution.

The use of a Bayesian framework for earthquake characterization Is not new. Yin et al. (2018) incorporates the recent seismic activity rate to determine the likelihood that a trigger at any individual station is related to an earthquake. This can reduce the likelihood of non-seismic triggers contaminating location algorithms, producing high error results. The GaMMA model (Zhu et al. 2022) has a component that similarly models the probability of a phase pick given the potential for multiple local earthquakes, and with this information associates the P and S-wave phases at nearby seismic stations and determines earthquake source parameters such as location and magnitude. When solving the rupture extent for large earthquakes, Minson et al. (2013) uses a Bayesian methodology to solve finite-fault inverse problems in synthetic environments, and then use the same model to focus on the 2011 Tohoku-oki earthquake (Minson et al., 2014). Lomax et al. (2009) lay out a generalized methodology for Bayesian grid search techniques, including test examples using the locations of stations and sources from past Italian earthquakes. In this study, the introduction of bEPIC fulfills a need for an algorithm that will operate under the temporal limitations of EEW without the need for S-wave or full waveform information in the initial grid search.

Algorithm Background

This study focuses on improvements affecting EPIC's location algorithm; formally known as the ElarmS algorithm (Allen et al., 2009; Chung et al., 2019). Prior to locating an event, a separate triggering algorithm uses an STA/LTA method to detect the arrival of P-waves on individual station channels. While this triggering algorithm is separate from the location algorithm, new triggers are regularly fed into EPIC when made available. As time elapses and seismic waves propagate further from the hypocenter, the number of triggers increases, especially in regions where the seismic network is dense. With one trigger, a preliminary epicentral location is set at the location of the triggered station. With two triggers from two different stations, the preliminary location moves to in between the first and second station. Once provided with triggers from at least four unique stations, EPIC uses a direct grid search to determine the best fitting event location (Chung et al., 2019; 2020). The grid space initially is centered on the preliminarily epicentral location determined from the sub-four triggers estimate. As better location estimates are made available through the inclusion of more triggers at unique stations, the grid space is updated and the grid center moves to the previous timestep's best event location. The best location from this new search is the grid node that has the lowest data misfit between observed and modeled station trigger times.

As a way of limiting the computational cost and reducing the number of parameters to be solved, EPIC pre-assigns a depth of 8.0 km for all events. This simplification is reasonable as most target events in the ShakeAlert reporting region occur on shallow crustal faults (Wurman et al. 2007; Thompson et al. 2021). In tandem with the location algorithm, the magnitude is calculated using a distance and P-wave peak displacement (P_d) scaling relation (Wurman et al., 2007; Kuyuk and Allen, 2012; Chung et al. 2020). As new stations are triggered and more waveform information

115 prior to S-wave arrival is available, EPIC updates its location and magnitude solution. Numerous
116 quality checks are included to differentiate between noise, calibration pulses, and earthquake
117 shaking, this is primarily done within the trigger algorithm before passing on to the location phase.
118 EPIC also limits false triggers by including a “filter bank” discriminator to reject picks related to
119 large teleseismic earthquakes (Chung et al. 2020).

120 As a means of limiting the total number of tested points in the location grid search, EPIC
121 also computes the activity level at each grid search node based on the number of nearby active
122 (triggered) and inactive (not trigged) stations (Sedar Kuyuk et al., 2014). At each timestep, EPIC
123 tallies the number of triggered (active) and untriggered (inactive) stations that are located inside
124 the grid search spatial domain. At each grid node in the search, the maximum distance between
125 the node and active stations is calculated. Then, a circle centered on the grid node with a radius
126 equal to distance of the furthest active station is drawn. The percent of active stations inside this
127 circle must exceed a preset threshold. Here, we set this value to be 30%. This means that at least
128 30% of stations near this grid node need to have triggered to for this node to be considered as a
129 viable earthquake location. The purpose of this exercise is to limit erroneous earthquake locations
130 on land in areas where many functioning stations did not trigger. However, this activity level filter
131 has little effect on limiting potential offshore grid node locations due to a lack of offshore stations
132 used in the ShakeAlert network.

133 The latest version of EPIC excels at providing low latency location and magnitude
134 solutions for earthquakes located on land, within the dense seismic network. This is illustrated in
135 Figure 1, which shows the location and magnitude estimate performance of EPIC for earthquakes
136 from October 2018 to May 2022. One area of potential improvement is the rapid location of
137 earthquakes rupturing offshore of Northern California in a seismically active area around the

Mendocino Triple Junction (MTJ). Almost all events with high error (> 50 km with respect to final USGS locations) occur within the MTJ region. Here, the average location error for EPIC's final event solution is 58 km. In contrast, the average location error for events that originate outside the MTJ is only 4 km. High location errors also can contribute to large errors in the magnitude calculation. Because the magnitude is calculated with a scaling law that depends on station-epicenter distances, a large, and erroneous increase in distance will contribute to an overestimate in magnitude; in some poorly located cases the magnitude is a full unit higher than the true solution. While some recent earthquakes used in our test dataset were too small to merit the issuance of an alert, larger magnitude earthquakes within the MTJ area are possible. For example, the 1992 M7.1 Cape Mendocino earthquake ruptured just offshore, possibly on the subduction interface (Oppenheimer et al., 1993). This was preceded by the 1994 M6.9 earthquake, further west along the Mendocino Fault Zone and numerous other M6.0+ earthquakes along the fault zone and within the Gorda deformation zone (Rollins and Stein, 2010). Improved detection and characterization of all offshore events raises confidence in the less common, larger offshore earthquakes, limiting potential future false alarms.

Data

To test the performance of the bEPIC algorithm, we compiled a replay catalog composed of past events from ShakeAlert. Each event in the replay catalog has an EPIC solution containing the earthquake location, magnitude, and origin time over the entire time history of the event, i.e. over a duration of tens of seconds as the seismic network collected real-time data. The catalog also contains the event ID and source information for the corresponding USGS event solution to allow for an easy comparison. We include earthquakes with either an EPIC or USGS determined

magnitudes greater than 3.5 and occurring between October 2018 and May 2022. For each replay event, we gather information on which stations were triggered and their trigger times. We visually inspect the triggers for each event and manually discard events where the trigger dataset is contaminated with triggers from noise, boxcar functions, the passage of teleseismic waves, and cases where the seismic waves of more than one event are visible within the seconds around the trigger (Figure S01). Of an initial catalog with 628 events, we discarded 86 events due to poor data quality, leaving us with a replay catalog of 542 earthquakes. While events with high noise and poor quality data are also prone to producing high location error solutions, we wish to first find improvements to EPIC for cases where the data quality is good, but the grid search location algorithm performs poorly. While it is an important problem for early warning, we leave the question of how to identify and handle the ingest of poor-quality data in real time to a future study. Many of the remaining earthquakes rupture within California, but a few events also occur in British Columbia, Canada, Washington, Oregon, Nevada, and Baja California, Mexico.

For each event in the replay catalog, we create a table of all the stations that triggered, the trigger times, and the timestamp when EPIC included that data into the location algorithm. This table is queried at each time step in the replay. In a typical solution, EPIC will recompute the earthquake location as more station triggers are made available due to the passage of body waves at further and further distances from the source. While EPIC will often recompute the earthquake location at sub-second intervals, the exact timing of each re-computation is dependent on the time between new station triggers. EPIC ceases computing locations when no new stations are triggered due to the P-wave's attenuation with distance to below the triggering algorithm's detection level or due to stations exceeding a set distance limit of 200 km away from the previously computed

location, whichever is closer. By having our replay model mirror EPIC's inclusion of stations over time, we ensure a better solution comparison between our replay and real-time performance.

We gather earthquake magnitude and location information from the ANSS Comprehensive Earthquake Catalog (ComCat) for use in our prior seismicity function. The base catalog contains all events from 01 January 2000 to May 2022 of magnitude 3.0 or greater in a region enveloping the ShakeAlert reporting zone. As new earthquakes occur within this zone, the catalog can be updated at regular temporal intervals to reflect the new information. All events, regardless of magnitude and age, are weighed equally in the seismicity catalog. We test our algorithm with the assumption that bEPIC will have access to the seismicity catalog, which would be updated over time to include new events. When testing bEPIC with our replay catalog, we only include in seismicity up to, but not including the time of our test event. This removes the potential that the test event's appearance in the ComCat catalog, or any potentially related aftershocks biases our solutions. When not testing past replay events, the full catalog up to present is used.

Methods

For each step in the processing of each earthquake in our replay catalog, we calculate the earthquake location (latitude and longitude), and magnitude. We draw our preferred earthquake location from the most probable grid node using a Bayesian framework. In the following section we describe how we formulate the bEPIC algorithm, including the construction of the likelihood and prior seismicity functions.

The posterior probability density model (*pdf*) of the model parameters, $P(\boldsymbol{m}|\boldsymbol{d})$, can be obtained through Bayes' theorem:

$$P(\boldsymbol{m}|\boldsymbol{d}) = \kappa\rho(\boldsymbol{m})\alpha(\boldsymbol{m})P(\boldsymbol{d}|\boldsymbol{m})$$

where **d** is a vector containing the P-wave arrival times at *N* stations provided by EPIC's triggering algorithm: $\boldsymbol{d} = (t_1^{obs}, \ldots t_N^{obs})$. The components of the model parameter vector are the x and y coordinates of all potential hypocenter locations, $\boldsymbol{m}=(x,y)$, that are possible within the grid search domain.

The prior information is expressed by $\rho(\boldsymbol{m})$, a seismicity probability for all grid nodes in model vector **m**. The likelihood, $P(\boldsymbol{d}|\boldsymbol{m})$, indicates how well the data fit the model at each point on the grid. The constant, κ is a normalization constant that ensures that the integral of the posterior probability density function is equal to 1. An additional parameter that is currently implemented in EPIC, $\alpha(\boldsymbol{m})$, represents the station activity level at each grid node. This binary vector of length $\boldsymbol{m}$ limits the posterior solution to parts of the grid search domain where the total percent of nearby stations that triggered from the event are above a pre-set percentage. In practice, this mask removes grid nodes (by setting their value to zero) that are close to un-triggered stations using the logic that if the earthquake was close to those stations, they would have triggered. This mask is effective for events located on-land in inside the seismic network. It has limited use for offshore events because most if not all potential earthquake locations are located away from stations.

The likelihood function is a non-normalized measure of how well each grid point explains the observed data while also considering uncertainties inherent in the assumptions included in both the data and algorithm. It provides an estimate of the best event location over the spatial domain. The conditional probability from the likelihood function is:

$$P(\boldsymbol{d}|\boldsymbol{m}) \propto L(\boldsymbol{m}) = \exp\left(-1/2 \sum_{i=1} \frac{\left[d_i - d_i^{calc}(m)\right]^2}{\sigma_i^2}\right)$$

where d_i is the observed P-wave arrival time at each station, i and $d^{calc}{}_i$ is the calculated P-wave arrival time. The station calculated P-wave arrival times are derived from the station and grid node

distances, the expected travel time using a velocity model, and a calculated origin time specific to each grid node.

The arrival time uncertainty at each station is denoted by $\boldsymbol{\sigma}$, a vector of length N. The value used for the uncertainty in arrival times stems from incomplete knowledge of the velocity structure along the ray path, errors in the triggering algorithm, and uncertainties stemming from the coarseness of the grid search spatial domain. A grid node with a likelihood value that approaches the value one has a very low misfit, thus high likelihood of being the location of the earthquake. A low likelihood has a high misfit and is less likely to be the true location of the earthquake.

The prior function is a representation of our best knowledge about where earthquakes likely occur and is based on past local seismicity. The data included in the prior is pulled from the ANSS ComCat catalog. In replay mode, only the earthquakes with rupture dates prior to the test event are included in the dataset.

We construct our prior pdf, $\rho(\boldsymbol{m})$, using a two-dimension kernel density estimate (*kde*). The *kde* is a way to characterize the probability of a random variable, in this case, the location of past earthquakes. The epicenter of each past earthquake that occurred inside the gird search domain is characterized as a Gaussian kernel, K, and takes the form:

$$K(\boldsymbol{m}|\boldsymbol{\mu}, \Sigma) = \frac{1}{2\pi\sqrt{|\Sigma|}} exp(-\frac{1}{2}(\boldsymbol{m}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{m}-\boldsymbol{\mu}))$$

Where the bandwidth, Σ, is a 2x2 covariance matrix controlling the shape and extent of smoothing of each kernel. We assume the location in both the x and y direction of the past seismicity are independent variables and assign a bandwidth value following Scott's rule (Scott, 1992): A smaller bandwidth value will limit the amount of smoothing, a larger value will extend the smoothing of each Gaussian kernel. The prior, $\rho(\boldsymbol{m})$ is the summation and normalization of all kernels:

$$\rho(\boldsymbol{m}) = \frac{1}{n}\sum K(\boldsymbol{m}|\boldsymbol{\mu}, \Sigma)$$

In application, the prior *kde* is applied only over the model space, limiting the number of events from the prior seismicity catalog to those that occur inside the grid search domain. If there is a situation where no past earthquakes occurred inside the model domain, then the value of the prior will be constant and therefore have no impact on the final posterior solution.

The expression of the prior is shown across the western United States in Figure 2. Here we show all past seismicity incorporated into our reference catalog. In this example, all past data is incorporated into the kernel density estimate to illustrate areas of low and high seismicity. Inset figures show the distribution of seismicity within the MTJ (Figure 2B) and for reference, in a section of California (Figure 2C). In execution, the prior is only computed over the same spatial domain as the location grid search. Additionally, when replaying past earthquakes in our testing catalog, we take care to only include earthquakes that occurred at times prior to our replay event. While the effect of the addition of one or a handful of earthquakes to the past seismicity catalog is small, we want to limit potentially biasing effects in our results.

EPIC's triggering algorithm provides a list of stations that triggered, which is then used as inputs into the location algorithm. However, stations that are within our grid search domain but did not trigger can be used as a data set to define our activity mask. The activity mask, $\alpha(\boldsymbol{m})$ contains either a 1 or a 0 depending on if the grid node is within or outside of the network. The network is defined here as any grid node where at least 30% of the nearby stations are triggered. To determine what is inside or outside this network, we calculate the distance between the grid node and all stations inside the grid search domain. From this, we draw a circle with a radius equal to the distance of the furthest triggered station and assess how many stations inside this circle triggered and how many did not. If more than 30% of stations are triggered inside the

circle, then this grid node is considered inside the network. The value for alpha at the index of this grid node is then assigned to be 1. Because new stations have been added to the ShakeAlert network over the duration of our test catalog, we only include stations that were installed prior to the time of the replay event.

Finally, the posterior pdf is the product of the likelihood, prior, and activity mask functions, normalized by a constant κ. The best solution is drawn from the maximum value of the posterior function. This best solution is then used as the new initial location for creating a grid search in subsequent iteration of solving the event location, if more information from new seismic stations is made available. This best location is also used when calculating the earthquake magnitude.

Once the earthquake location is known, the magnitude at each station is computed using a distance and P-wave peak displacement scaling relation (Kuyuk & Allen, 2013):

$$M = a \cdot log_{10}(P_d) + b \cdot log_{10}(R) + c$$

where a, b, and c are empirical constants and set as 1.23, 1.39, and 5.39 respectively. R is the distance from the station to the hypocenter, in kilometers. P_d is the peak displacement observed on the sensor in up to the first 4 seconds following the P-wave arrival time. The final earthquake magnitude is an average of the individual station magnitudes. As the location is updated and more of the initial waveform is made available at triggered stations, the magnitude is updated.

Results

The inclusion of the prior seismicity layer to EPIC's location algorithm greatly improves location accuracy for earthquakes nucleating in the MTJ. Figure 3 shows the improvement in location for events across the ShakeAlert reporting region and specifically within the MTJ (Figure 3B). Figure 4 shows the improvement in location accuracy when using bEPIC for the entire

296 reporting area (Figure 4a) and for events within the MTJ (Figure 4b). In the original EPIC location
297 solutions, the mean and median location error for events inside the MTJ is 58 km and 34 km,
298 respectively. The solutions from bEPIC reduce the mean error to 14 km and the median error to 7
299 km. On average, bEPIC improves the location accuracy of earthquakes in the MTJ by 44 km. The
300 inclusion of prior seismicity also does not negatively affect the location quality for events that
301 occurred on-land and outside of the MTJ. With EPIC, the events that occurred outside the MTJ
302 had a mean and median location error of 4 and 2 km. bEPIC produces results with mean and
303 median location errors of 3 and 2 km for this same dataset. Two events outside of the MTJ had
304 EPIC location errors that exceeded 100 km. Both occurred near the Southern California border
305 with Mexico. bEPIC improves the location estimate for both of these cases as well.
306 EPIC and bEPIC use the same scaling laws to compute the earthquake magnitude. Because
307 this scaling law relies in part on the distance between the station and the earthquake, an
308 improvement in earthquake location also has the potential to improve the magnitude estimate.
309 Figure 4C and D show an improvement in the accuracy of the magnitude calculation when using
310 bEPIC. The average magnitude error for earthquakes within the MTJ improves from 0.4 units with
311 EPIC to 0.06 magnitude units using bEPIC. For events outside of the MTJ, the average magnitude
312 error for EPIC and bEPIC events is similar at 0.06 and 0.07 magnitude units, respectively. Overall,
313 the magnitude error is strongly related to the location error, where an increase in location error
314 also increases the magnitude error (Figure S02). A reduction in the former will likely cause an
315 improvement in the latter.
316
317 **Discussion**

A common feature in EPIC solutions for earthquakes located within the MTJ is a location-overshoot. This is when there is not enough station data in the right places to constrain the grid search location, creating a large area of reasonably high likelihood extending far offshore. Because of the similar high likelihood, it is common for EPIC to draw a location far offshore and away from the true location. Then, as more data is made available, EPIC redraws its grid search domain based on this overshot location, which can perpetually move the EPIC location further and further offshore. Once a location is more than 200 km from any station, EPIC discards that event. This can lead to missed events, where potentially an alert may have been warranted, but is never issued. The use of prior seismicity in bEPIC limits this location-overshoot phenomenon.

Looking at our test catalog as a whole, the inclusion of prior seismicity greatly reduces the location error for earthquakes rupturing within the MTJ and performs as well as the non-Bayesian EPIC for earthquakes rupturing on land and within dense seismic networks. The test catalog includes earthquakes with varying arrangements of station locations and prior seismicity levels. Here, we draw attention to three earthquakes from the test catalog to use as points of discussion. First, an earthquake originating within the MTJ, where our study is primarily focused. Second, an earthquake inland in southern California, where station coverage is excellent. Third, an earthquake on the coast of northern California that occurred outside the region of contemporary seismic activity. The spatial distribution of the likelihood, prior seismicity and posteriors are mapped for each of the three examples in Figures 5-7.

The MTJ example earthquake is a M3.5 earthquake that occurred on 23 March 2020 (Figure 5). The USGS location for this event is 40.309ºN 124.672ºW, which is 33 km west of Petrolia, California. The final EPIC location for this event was 39.93ºN 125.87ºW which is 110 km to the southwest of the USGS solution. The tendency of EPIC to locate events in this region to

the southwest of their true locations is also noted for many other recent MTJ events and visible in Figure 1. This large error in location contributed to an overestimate of the magnitude by 0.7 magnitude units (M4.2). The mapped likelihood function (Figure 5a) shows a clear band of near equally high likelihood extending far offshore while the prior seismicity (Figure 5b) is generally concentrated along the Mendocino Triple Junction, Mendocino Fracture Zone, and within more diffuse seismicity on land and within the Gorda Plate. Combining the prior and the likelihood functions, creates a mapped posterior that has a lowered probability of the event being located far offshore, and a slightly increased probability of the event occurring in the locations where both the likelihood and the prior seismicity have larger values (Figure 5c). The bEPIC solution at the same final timestep is 8 km northeast of the USGS solution with a magnitude of 3.48. Many earthquakes from the test catalog that occurred in the MTJ have a very similar likelihood function and final posterior probability. This is in part because of the narrow azimuth of the coastal stations with respect to the source and limited availability of stations within a timeframe that would allow for a timely alert to be issued. While additional stations further north and south of the rupture zone would allow for better azimuthal coverage, and thus a better overall control on the event location, waiting for these stations negates the use of the location for EEW. Instead, including in the prior information acts as a useful and automatic check on the location algorithm.

The second earthquake is a M5.3 in Southern California on 05 June 2021 (Figure 6). The USGS location for this event is 33.14ºN 115.635ºW, which is just south of the Salton Sea. The earthquake occurred in an area with high seismic activity and is part of the Brawley Seismic Zone (Hauksson et al., 2022). The region is also surrounded by a dense seismic network. Because this earthquake is surrounded by local stations, the location is well constrained with just the likelihood function alone. As shown in Figure 6a, the mapped likelihood creates a bullseye pattern with a

single clear maximum point indicating where the earthquake should be located. For this event, EPIC's final solution was only 3.2 km from the final USGS solution. This is a typical average location error for earthquakes rupturing on land. The prior seismicity (Figure 6b) is high over this region, particularly on the cross faults associated with the San Andreas and Imperial Faults. Because there is only one small area with high likelihood associated with this grid search, the prior seismicity does not greatly affect the final posterior solution (Figure 6c). The final posterior solution (1.7 km from USGS solution) is nearly identical to the solution provided by EPIC. While EPIC sans the prior component produced a satisfactory solution on its own, we advocate for the inclusion of the prior anyways. Creating a system where EPIC decides when and if to use a prior seismicity can potentially generate its own errors, particularly if a solution may appear well constrained with limited data, but then move or require a prior as more station data is incorporated into the solution with time.

The third test catalog example examines how the location algorithm will perform for regions where there is little recent documented seismicity. The example event is a M3.5 earthquake that occurred on 01 October, 2019, rupturing offshore and on the San Andreas fault, but south of the MTJ in a seismically quiet part of the transform system (Figure 7). While occurring offshore, the earthquake still had adequate station coverage to produce a satisfactory solution, as shown by the mapped likelihood in Figure 7a. However, this section of the San Andreas has had only two other earthquakes of magnitude 3.0 or greater in the past 20 years: a M3.0 on 12 February 2012 and a M3.5 on 24 November 2002. In contrast, there is a higher level of recent seismicity immediately to the east and on land within California's coastal range. Given the lack of seismicity near this earthquake and much larger values on the prior seismicity near the recent on land seismicity (Figure 7b), could lead one to expect the final solution to be pulled inland and into the

zone of high seismicity. However, this is not the case here. Figure 7c maps the posterior probability, which remains highest within the area that also has a high likelihood.

The reason that the bEPIC solution does not get pulled into the zone of high seismicity is related to the difference in scale between the non-normalized values mapped in the likelihood function and the normalized values in the prior. The effect that this has is that while some areas with past seismicity have higher relative values than other areas with little to no seismicity, the overall magnitude of the prior is small. What this means is that when applied to the likelihood, the greatest modification only occurs among grid nodes with high and similar values where the small weight of the prior can have an effect. In the example shown here, the relatively high seismicity in the prior at locations on land has extremely little effect as it is being applied to an area with extremely low (close to zero) likelihood.

The second reason that the bEPIC solution does not get pulled inland to the high seismicity zone in Figure 7 is that the prior seismicity has a floor in place where every point in the prior must be greater than zero. This stops the prior from having too much power in limiting new earthquakes in low seismic zones. While there is no ceiling value set for the prior, the requirement that the prior be normalized sets a limitation on how large prior values can be.

How EPIC and bEPIC would behave for potential future events is harder to capture without real data, but still worth examining. All recent seismicity with EPIC solutions that appear in the test catalog within the MTJ are concentrated within about 50 km of the coastline. However, there have been large past earthquakes that ruptured further offshore on the Mendocino Fracture Zone but still inside the ShakeAlert reporting zone. This includes larger earthquakes such as the 1994 M7.1 Ferndale earthquake as well as a handful of M6.0+ earthquakes rupturing between roughly 127ºW - 125ºW. While this region does have past seismicity that is incorporated into local priors

(Figure 2), there are clearly more recent earthquakes closer to the coastline in the database than events further west. Would the larger density of events here have a relative pulling effect on far offshore earthquakes, moving their solutions further east? We tested a preliminary set of synthetic events located along the Mendocino Fracture Zone. Early in the location search, when only 3-4 stations are available and the mapped likelihood is high over a broad region, there is an initial pulling effect towards the eastern higher seismicity region in the prior (Figure S3). However, as even a small amount of additional station data is made available, the bEPIC solution migrates westward closer to the 'true' location. In contrast, the solutions using a non-Bayesian EPIC struggle with the initial high likelihood, leading to locations that get placed hundreds of kilometers offshore. The prior seismicity, while creating an initial pulling effect, limits this location-overshoot common in EPIC. When repeating our synthetic offshore simulations, bEPIC produced on average lower location errors than EPIC.

An improvement in earthquake location affects the rest of the ShakeAlert system. Because the station-epicenter distance is used in the magnitude calculation, a more accurate location leads to a more accurate magnitude. Additionally, the magnitude and location together are used to determine the extent and intensity of ground shaking. With large enough location errors, and in kind, an overestimate of the magnitude, false alerts could be created. Too many false alerts have the potential to lower community confidence in alerts received for future events. Furthermore, accurately locating and characterizing offshore earthquakes is important because of the additional hazards associated with offshore earthquakes, such as tsunamis. All cases handled in our test data set are too small to generate a tsunami, however the MTJ has the potential to generate larger earthquakes and tsunamis. Such was the case with the 1992 Mendocino earthquake, an interslab thrust event on the subducting Gorda microplate. This event generated a small tsunami that was

identified at regionally placed tide gauges as well as at a tide gauge in Hawaii (Oppenheimer et al., 1993; Gonzalez et al., 1995). Accurately identifying the location and size of events like this, as well as limiting false alarms is highly important.

An area of needed improvement within EPIC that can affect bEPIC solutions is the inclusion of poor trigger picks from the associator algorithm. Both EPIC and bEPIC receive station trigger data from a separate algorithm that both collects triggered stations and associates nearby triggers. While many checks are in place to limit triggers from telesiemic events, which have in the past generated numerous false alarms (Chung et al., 2019) as well as quality control metrics to weed out triggers from calibration pulses, boxcar shaped features, and errant noise, some poor data still can be sent to the location algorithms. When this happens, there is no current ability for EPIC/bEPIC to identify and throw out poor picks. This means that a single station pick with a trigger time error of multiple seconds can cause an overall high misfit across all grid search locations, creating a near zero likelihood everywhere. This in turn means that the added prior information does little to correct the final location solution. While uncommon, this poor data problem will generate poor location solutions. Future additions to EPIC should include additional metrics in the location algorithm to allow for the ability to reject poor station trigger information.

An interesting future direction for EEW is the incorporation of real-time offshore data into location algorithms by means of fiber optic distributed acoustic sensing (DAS) arrays (Farghal et al., 2022). The incorporation of offshore data can, if positioned close to the source, provide initial trigger information much sooner than systems that wait for on-land stations. Using both offshore DAS array data and traditional onshore seismometers in EEW also has the benefit of greatly reducing the level of initial uncertainty when solving for the locations of offshore earthquakes, as discussed at length here. If made available, data from currently existing transoceanic fiber optic

cables would allow for greater sensitivity in detecting smaller offshore events as well as other geohazards such as sending the propagation of tsunami waves in the open ocean (Salaree et al., 2022).

Conclusion

Earthquake early warning through programs like ShakeAlert build the tools communities need to take proactive steps to mitigate risk in the seconds prior to damaging shaking. While ShakeAlert can provide timely alerts for many of the earthquakes it encounters, the lack of seismic network coverage extending offshore makes accurate event locations for earthquakes nucleating near the Mendocino Triple Junction challenging. In this study we modified ShakeAlert's point source earthquake characterization code, EPIC to include prior information about the seismicity of a region as an additional component affecting the grid search algorithm. While we include this layer to specifically target poorly located events offshore, we also apply and assess the performance of the modified point source algorithm for events across the entire ShakeAlert reporting area. Our improved code, bEPIC, produced solutions with average location errors in the MTJ of 14 km. This is an average reduction in error for events in this region of 44 km. The inclusion of the prior seismicity layer does not negatively affect in network solutions where EPIC already excelled. In these cases, the prior information has little impact on the overall posterior result and no new sources of error are introduced. With better location estimates for offshore events, the bEPIC calculated magnitude also improves, limiting the potential of a false alert being issued for a large, offshore earthquake

The improvement to the solved earthquake location by bEPIC limits the likelihood of false alerts for communities along the northern California coast. This is a region that is particularly

sensitive to false alerts, as large offshore earthquakes in this region also carry the possibility of other seismically related hazards such as tsunami.

Data and Resources

bEPIC replay information including station location and trigger times for earthquakes used in this study, as well as catalog information for all test events are available at: 10.5281/zenodo.6929789. The prior seismicity information was created by querying the ANSS Comprehensive Earthquake Catalog (ComCat): https://earthquake.usgs.gov/data/comcat/ (last accessed July 26, 2022). Some plots were made using PyGMT (https://zenodo.org/record/6702566; Wessel et al. 2019; Uieda et al. 2022). The source code for bEPIC is available through GitHub: https://github.com/amy-l-williamson/bEPIC. The source code for bEPIC is available through GitHub: https://github.com/amy-l-williamson/bEPIC. The supplemental material includes three supplemental figures. Figure S1 shows an example of a poor P-wave trigger pick. Figure S2 shows the relationship between earthquake location error and magnitude error for the replay catalog. Figure S3 shows the spread in EPIC and bEPIC locations a potential synthetic far offshore earthquake as referenced in the Discussion section.

Acknowledgements

The authors would like to thank Anthony Lomax and one anonymous reviewer for their helpful comments. The authors would also like to thank Tom Heaton for their constructive comments. This research was supported by the U.S. Geological Survey under Grant Agreement Number G21AC10525 awarded to UC Berkeley.

Declaration of Competing Interests

The authors declare no competing interests.

Figure Captions

Figure 1. A. Summary of EPIC location and magnitude performance from October 2018 to May 2022. EPIC final event locations are plotted as green, yellow, orange, red, and dark red dots when the location error is below 25 km, between 25 km and 50 km, between 50 and 75 km, between 75 and 100 km, and greater than 100 km respectively. Dots are plotting at EPIC's location. For each event, a corresponding colored dashed line connects the EPIC location with the final USGS location. The blue dashed box outlines what we define as the Mendocino Triple Junction (MTJ) region and is used to classify the offshore events. **B.** Zoomed in view of the MTJ region, and location error between EPIC and USGS locations. BC, WA, OR, NV, and CA symbols mark British Columbia, Washington, Oregon, Nevada, and California, respectively.

Figure 2. Contemporary seismicity of M3.0 and larger events across the western United States, marked with black *x* icons. Black dashed lines indicate the inset panels B and C. Panel B shows the extent of the Mendocino Triple Junction region. Panel C shows a subsection of California. The entire region is shaded based on levels of seismicity where lighter colors have lower seismicity, and darker shades have higher seismicity.

Figure 3 A. Earthquake location accuracy across the ShakeAlert network when using the bEPIC modification for the same events as in Figure 1 across the western US. Blue dashed line outlines our defined MTJ zone. B. Zoomed in view of the MTJ region (blue dashed line region in subplot A), and location error between bEPIC and USGS locations.

Figure 4. A. Location error, in log scale, for EPIC (blue bars) and bEPIC (red bars) for all events in the replay catalog. **B**. Location error, with same scaling as subplot A, for only earthquakes located within the MTJ region. **C.** Magnitude error, linear scale, for all earthquakes in the replay catalog. **D.** Magnitude error for earthquakes located within the MTJ region.

Figure 5. Likelihood, prior seismicity, and posterior solutions for an offshore event example from the test catalog. Note the difference in color scales between the likelihood (a non-normalized value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network that did trigger. White triangles mark stations that did not trigger. The white star marks the USGS earthquake location for each event. Red star marks the EPIC event location, blue star marks the bEPIC event location.

Figure 6. Likelihood, prior seismicity, and posterior solutions for an in-network and onland event from the test catalog. Note the difference in color scales between the likelihood (a non-normalized value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network that did trigger. White triangles mark stations that did not trigger. The white star marks the USGS earthquake location for each event. Red star marks the EPIC event location, blue star marks the bEPIC event location.

Figure 7. Likelihood, prior seismicity, and posterior solutions for an offshore, non-MTJ event in an area of low past seismicity from the test catalog. Note the difference in color scales between the likelihood (a non-normalized value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network that did trigger. White triangles mark stations that did not

trigger. The white star marks the USGS earthquake location for each event. Red star marks the EPIC event location, blue star marks the bEPIC event location. *Note*: in this example, the red and blue stars are co-located.

References

Allen, R. M., Brown, H., Hellweg, M., Khainovski, O., Lombard, P., & Neuhauser, D. (2009). Real-time earthquake detection and hazard assessment by ElarmS across California. *Geophysical research letters*, *36*(5).

Allen, R. M., & Melgar, D. (2019). Earthquake early warning: Advances, scientific challenges, and societal needs. *Annual Review of Earth and Planetary Sciences*, *47*, 361-388.

Böse, M., Smith, D. E., Felizardo, C., Meier, M. A., Heaton, T. H., & Clinton, J. F. (2018). FinDer v. 2: Improved real-time ground-motion predictions for M2–M9 with seismic finite-source characterization. *Geophysical Journal International*, *212*(1), 725-742.

Böse, M., Felizardo, C., & Heaton, T. H. (2015). Finite-fault rupture detector (FinDer): Going real-time in Californian ShakeAlert warning system. *Seismological Research Letters*, *86*(6), 1692-1704.

Böse, M., Heaton, T. H., & Hauksson, E. (2012). Real-time finite fault rupture detector (FinDer) for large earthquakes. *Geophysical Journal International*, *191*(2), 803-812.

572

573 Chung, A. I., Henson, I., & Allen, R. M. (2019). Optimizing earthquake early warning
574 performance: ElarmS-3. *Seismological Research Letters*, *90*(2A), 727-743.
575

576 Chung, A. I., Meier, M. A., Andrews, J., Böse, M., Crowell, B. W., McGuire, J. J., & Smith, D.
577 E. (2020). ShakeAlert earthquake early warning system performance during the 2019 Ridgecrest
578 earthquake sequence. *Bulletin of the Seismological Society of America*, *110*(4), 1904-1923.
579

580 Espinosa-Aranda, J. M., Cuellar, A., Garcia, A., Ibarrola, G., Islas, R., Maldonado, S., &
581 Rodriguez, F. H. (2009). Evolution of the Mexican seismic alert system
582 (SASMEX). *Seismological Research Letters*, *80*(5), 694-706.
583

584 Farghal, N. S., Saunders, J. K., & Parker, G. A. (2022). The Potential of Using Fiber Optic
585 Distributed Acoustic Sensing (DAS) in Earthquake Early Warning Applications. *Bulletin of the*
586 *Seismological Society of America*, *112*(3), 1416-1435.
587

588 González, F. I., Satake, K., Boss, E. F., & Mofjeld, H. O. (1995). Edge wave and non-trapped
589 modes of the 25 April 1992 Cape Mendocino tsunami. In *Tsunamis: 1992–1994* (pp. 409-426).
590 Birkhäuser Basel.
591

592 Hauksson, E., Stock, J. M., & Husker, A. L. (2022). Seismicity in a Weak Crust: The
593 Transtensional Tectonics of the Brawley Seismic Zone Section of the Pacific-North America
594 Plate Boundary in Southern California, USA. *Geophysical Journal International*.

595

596 Hoshiba, M., Iwakiri, K., Hayashimoto, N., Shimoyama, T., Hirano, K., Yamada, Y., ... &
597 Kikuta, H. (2011). Outline of the 2011 off the Pacific coast of Tohoku earthquake (Mw 9.0)—
598 Earthquake early warning and observed seismic intensity—. *Earth, planets and space*, *63*(7),
599 547-551.

600

601 Kohler, M. D., Smith, D. E., Andrews, J., Chung, A. I., Hartog, R., Henson, I., ... & Guiwits, S.
602 (2020). Earthquake early warning ShakeAlert 2.0: Public rollout. *Seismological Research*
603 *Letters*, *91*(3), 1763-1775.

604

605 Kuyuk, H. S., & Allen, R. M. (2013). A global approach to provide magnitude estimates for
606 earthquake early warning alerts. *Geophysical Research Letters*, *40*(24), 6329-6333.

607

608 Lomax, A., Michelini, A., Curtis, A., & Meyers, R. A. (2009). Earthquake location, direct,
609 global-search methods. *Encyclopedia of complexity and systems science*, *5*, 2449-2473.

610

611 McBride, S. K., Smith, H., Morgoch, M., Sumy, D., Jenkins, M., Peek, L., ... & Wood, M.
612 (2022). Evidence-based guidelines for protective actions and earthquake early warning
613 systems. *Geophysics*, *87*(1), WA77-WA102.

614

615 Oppenheimer, D., Eaton, J., Jayko, A., Lisowski, M., Marshall, G., Murray, M., ... & Valentine,
616 D. (1993). The Cape Mendocino, California, earthquakes of April 1992: Subduction at the triple
617 junction. *Science*, *261*(5120), 433-438.

Rollins, J. C., & Stein, R. S. (2010). Coulomb stress interactions among M≥ 5.9 earthquakes in the Gorda deformation zone and on the Mendocino Fault Zone, Cascadia subduction zone, and northern San Andreas Fault. *Journal of Geophysical Research: Solid Earth*, *115*(B12).

Salaree, A., Howe, B. M., Huang, Y., Weinstein, S. A., & Sakya, A. E. (2022). A numerical study of SMART Cables potential in marine hazard early warning for the Sumatra and Java regions. *Pure and Applied Geophysics*, 1-33.

Saunders, J. K., Minson, S. E., & Baltay, A. S. (2022). How low should we alert? Quantifying intensity threshold alerting strategies for earthquake early warning in the United States. *Earth's Future*, *10*(3), e2021EF002515.

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Serdar Kuyuk, H., Allen, R. M., Brown, H., Hellweg, M., Henson, I., & Neuhauser, D. (2014). Designing a network-based earthquake early warning algorithm for California: ElarmS-2. *Bulletin of the Seismological Society of America*, *104*(1), 162-173.

Strauss, J. A., & Allen, R. M. (2016). Benefits and costs of earthquake early warning. *Seismological Research Letters*, *87*(3), 765-772.

Strauss, J. A., Kong, Q., Pothan, S., Thompson, S., Mejia, R. F., Allen, S., ... & Allen, R. M. (2020). MyShake Citizen Seismologists help launch dual-use seismic network in California. *Frontiers in Communication*, 32.

Thompson, M., Hartog, J. R., & Wirth, E. A. (2022). Effect of Fixing Earthquake Depth in ShakeAlert Algorithms on Performance for Intraslab Earthquakes. *Seismological Society of America*, *93*(1), 277-287.

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, *20*(11), 5556-5564.

Wurman, G., Allen, R. M., & Lombard, P. (2007). Toward earthquake early warning in northern California. *Journal of Geophysical Research: Solid Earth*, *112*(B8).

Yin, L., Andrews, J., & Heaton, T. (2018). Rapid earthquake discrimination for earthquake early warning: A Bayesian probabilistic approach using three-component single-station waveforms and seismicity forecast. *Bulletin of the Seismological Society of America*, *108*(4), 2054-2067.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C. (2022). Earthquake phase association using a Bayesian Gaussian mixture model. *Journal of Geophysical Research: Solid Earth*, e2021JB023249.

Amy Williamson
Berkeley Seismological Laboratory
University of California, Berkeley
Berkeley, California 94703

Angela Lux
Berkeley Seismological Laboratory
University of California, Berkeley
Berkeley, California 94703

Richard Allen
Berkeley Seismological Laboratory
University of California, Berkeley
Berkeley, California 94703

Figures

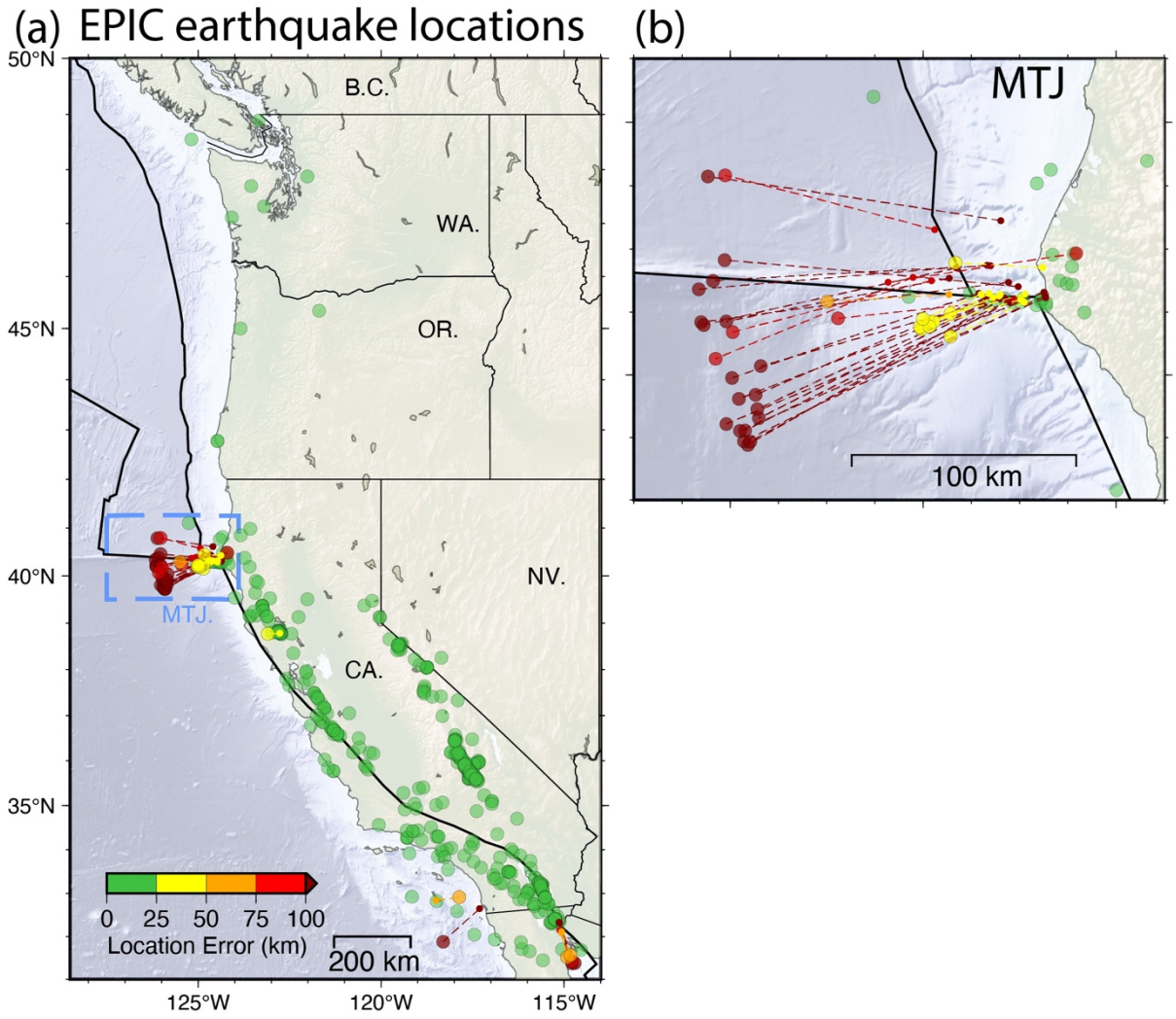


Figure 1. A. Summary of EPIC location and magnitude performance from October 2018 to May 2022. EPIC final event locations are plotted as green, yellow, orange, red, and dark red dots when the location error is below 25 km, between 25 km and 50 km, between 50 and 75 km, between 75 and 100 km, and greater than 100 km respectively. Dots are plotting at EPIC's location. For each event, a corresponding colored dashed line connects the EPIC location with the final USGS

location. The blue dashed box outlines what we define as the Mendocino Triple Junction (MTJ) region and is used to classify the offshore events. **B.** Zoomed in view of the MTJ region, and location error between EPIC and USGS locations. BC, WA, OR, NV, and CA symbols mark British Columbia, Washington, Oregon, Nevada, and California, respectively.

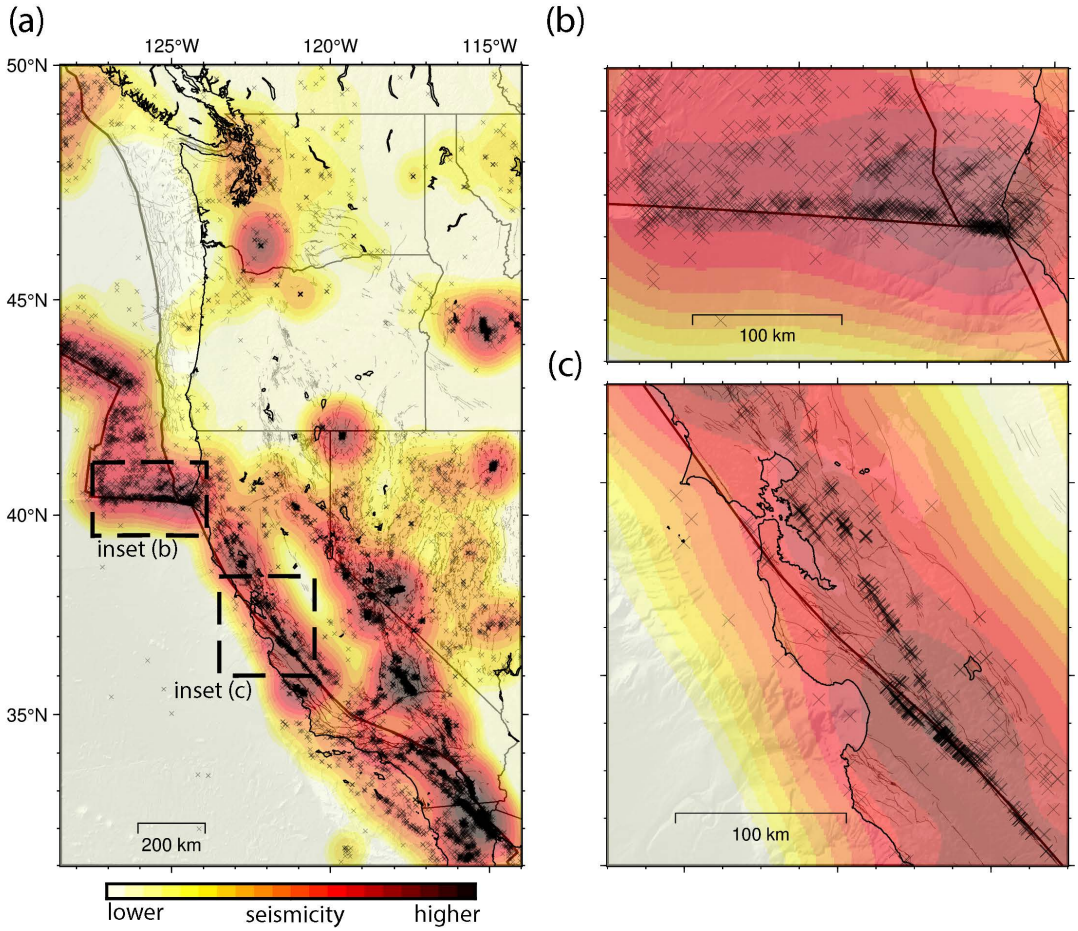


Figure 2. Contemporary seismicity of M3.0 and larger events across the western United States, marked with black *x* icons. Black dashed lines indicate the inset panels B and C. Panel B shows the extent of the Mendocino Triple Junction region. Panel C shows a subsection of California. The entire region is shaded based on levels of seismicity where lighter colors have lower seismicity, and darker shades have higher seismicity.

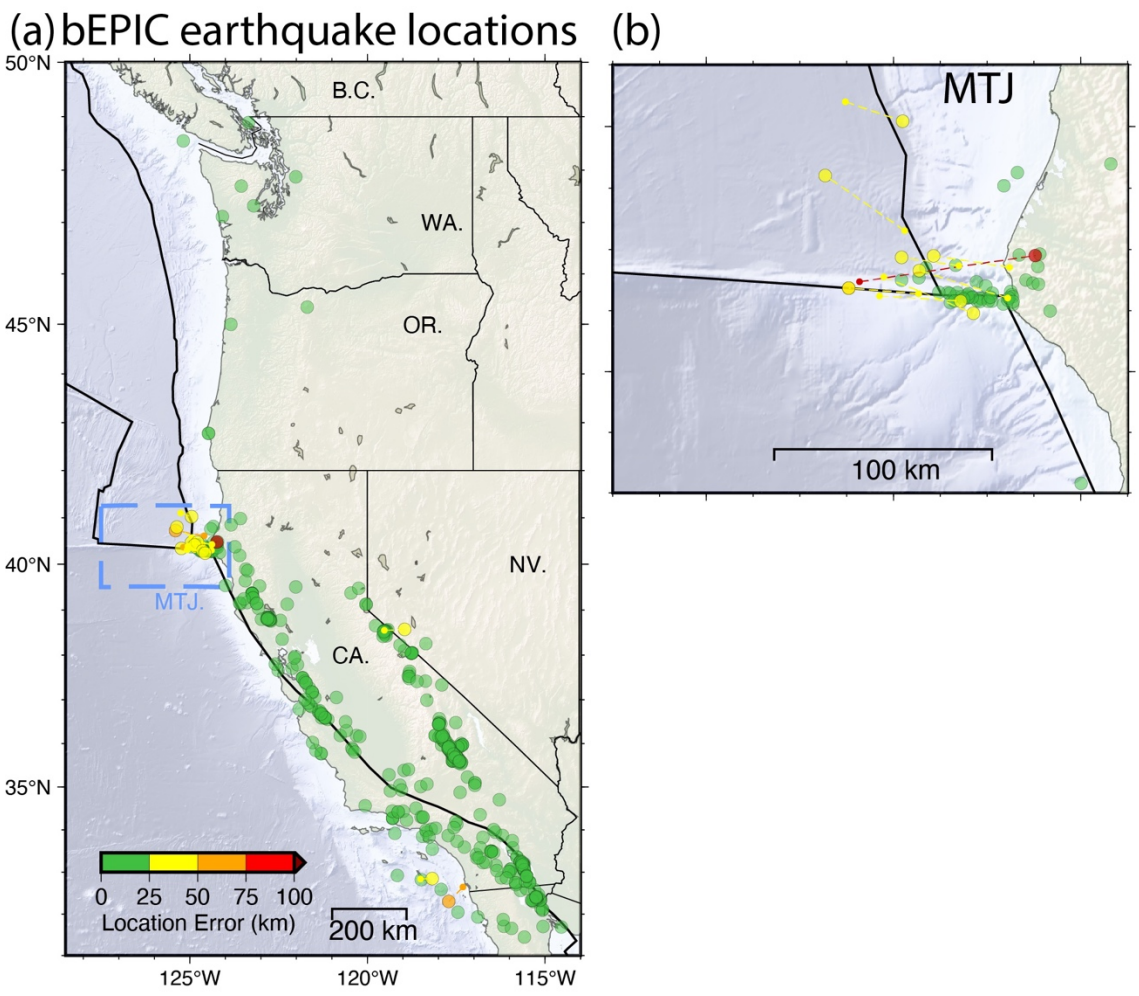
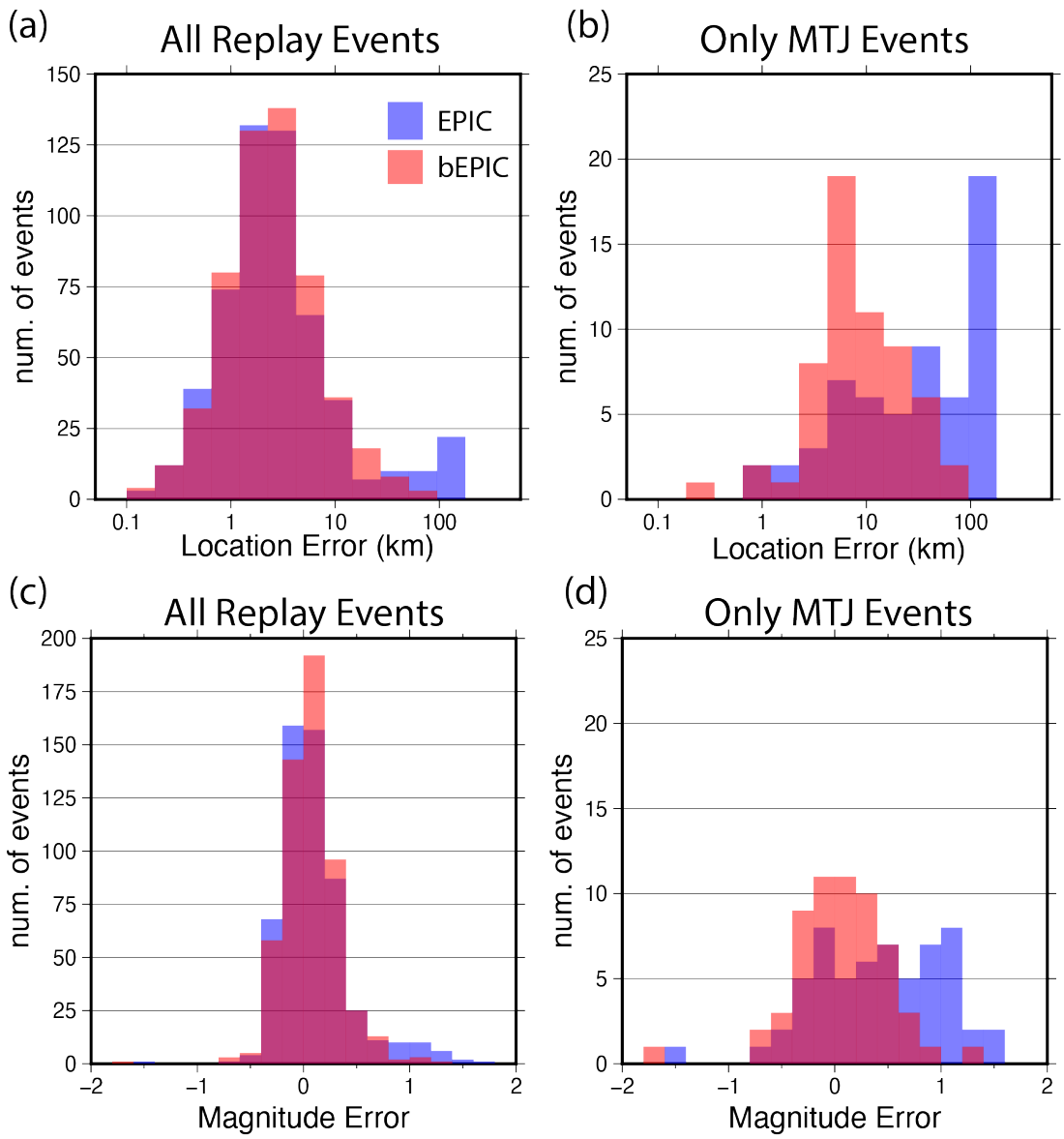


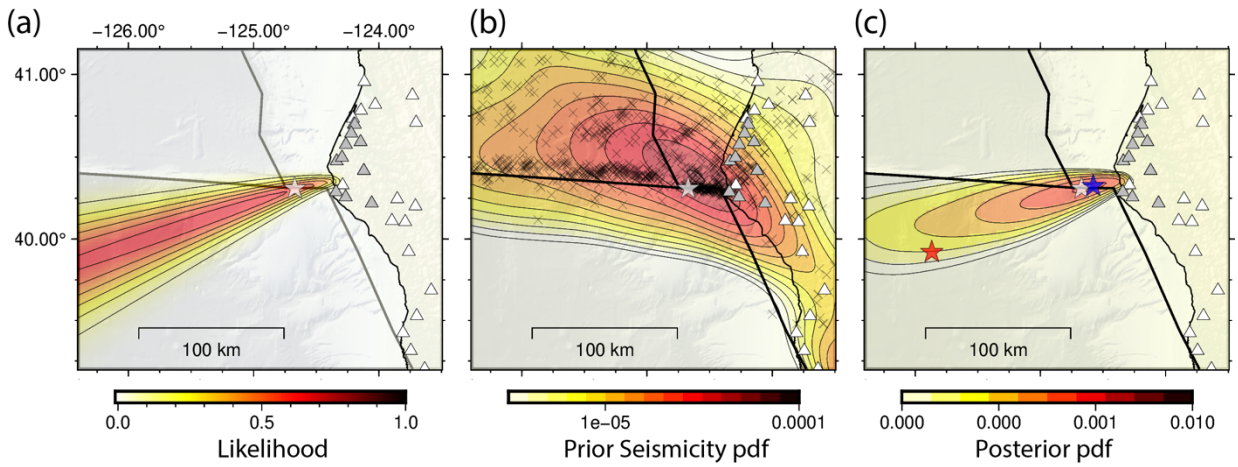
Figure 3 A. Earthquake location accuracy across the ShakeAlert network when using the bEPIC modification for the same events as in Figure 1 across the western US. Blue dashed line outlines our defined MTJ zone. B. Zoomed in view of the MTJ region (blue dashed line region in subplot A), and location error between bEPIC and USGS locations.



719

720 **Figure 4. A.** Location error, in log scale, for EPIC (blue bars) and bEPIC (red bars) for all events
721 in the replay catalog. **B**. Location error, with same scaling as subplot A, for only earthquakes
722 located within the MTJ region. **C.** Magnitude error, linear scale, for all earthquakes in the replay
723 catalog. **D.** Magnitude error for earthquakes located within the MTJ region.

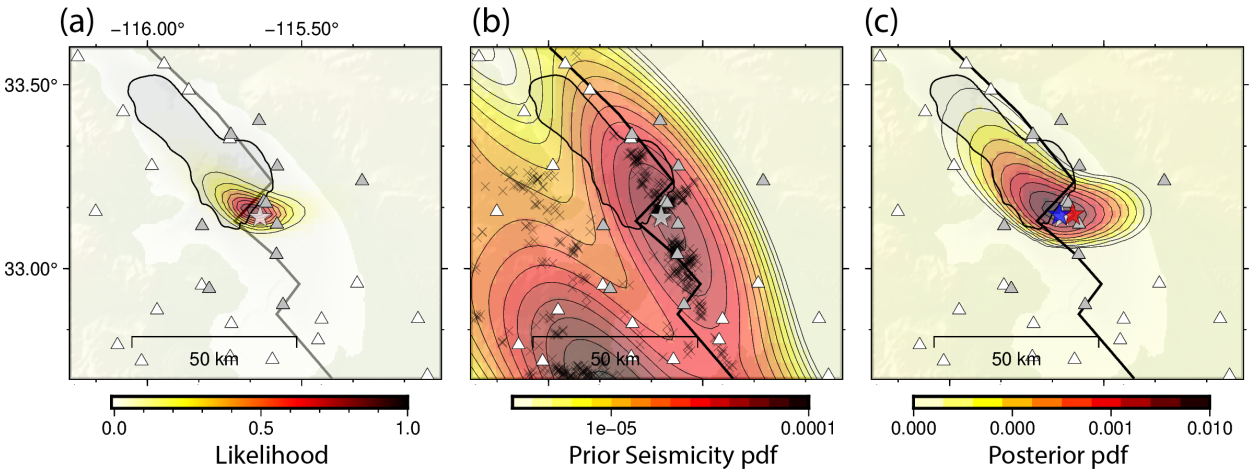
724



725

726 **Figure 5.** Likelihood, prior seismicity, and posterior solutions for an offshore event example from
727 the test catalog. Note the difference in color scales between the likelihood (a non-normalized
728 value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network
729 that did trigger. White triangles mark stations that did not trigger. The white star marks the USGS
730 earthquake location for each event. Red star marks the EPIC event location, blue star marks the
731 bEPIC event location.

732



733

734 **Figure 6.** Likelihood, prior seismicity, and posterior solutions for an in-network and onland event
735 from the test catalog. Note the difference in color scales between the likelihood (a non-normalized
736 value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network

that did trigger. White triangles mark stations that did not trigger. The white star marks the USGS earthquake location for each event. Red star marks the EPIC event location, blue star marks the bEPIC event location.

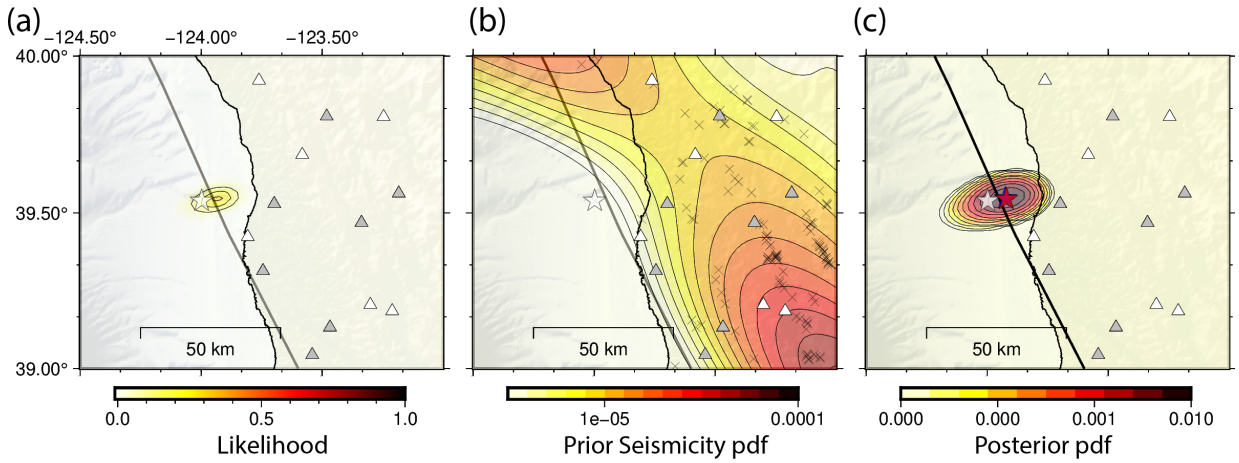


Figure 7. Likelihood, prior seismicity, and posterior solutions for an offshore, non-MTJ event in an area of low past seismicity from the test catalog. Note the difference in color scales between the likelihood (a non-normalized value) and the prior and posterior pdfs. Gray triangles mark stations in the ShakeAlert network that did trigger. White triangles mark stations that did not trigger. The white star marks the USGS earthquake location for each event. Red star marks the EPIC event location, blue star marks the bEPIC event location. *Note*: in this example, the red and blue stars are co-located.