# Supporting Information for "Benchmarking seismic phase associators: Insights from synthetic scenarios"

Jorge Puente[1,2], Christian Sippl[1], Jannes Münchmeyer[3], Ian W. McBrearty[4]

[1] Institute of Geophysics, Czech Academy of Sciences, Prague, Czech Republic

[2]Charles University, Faculty of Mathematics and Physics, Department of Geophysics, Prague, Czech Republic

[3]Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, Univ. Gustave Eiffel, ISTerre, Grenoble, France

[4]Department of Geophysics, Stanford University, Stanford, California, U.S.A.

**Contents of this file**



**Text S1: Evaluating the use of 0D vs. 1D velocity models**

In evaluating seismic phase associators, the choice between a simple homogeneous (0D) velocity model and a more detailed 1D velocity model can significantly impact performance. While 0D models offer simplicity and computational efficiency, they may overlook critical depth-dependent variations in seismic wave propagation that 1D models capture and introduce systematic errors when predicting traveltimes at larger distances. We thus compare the performance of REAL, GaMMA and PyOcto with 0D and 1D velocity models, for the crustal as well as the subduction zone scenario. Results from these runs are shown in Figures S2 and S3, and raw results are provided in Table S8. In the crustal scenario, where seismicity is shallow and thus occurs in a region of relatively uniform

velocities, the differences between 0D and 1D models are generally modest. GaMMA and PyOcto showed slight improvements in precision, recall, and runtime efficiency when using the 1D model, particularly at higher event counts and noise levels. For REAL, the 1D model reduced runtime but led to slightly lower performance scores overall. In the subduction zone scenario, characterized by a greater hypocentral depth range and longer raypaths, i.e. higher expected model errors, we encounter larger differences between 0D and 1D model versions. For GaMMA and especially PyOcto, the 1D version significantly outperforms the 0D one, whereas REAL once again shows better performance with the 0D model, suggesting that the simpler homogeneous model is more effective here. Based on these findings, we selected the 0D version of REAL, and 1D versions of GaMMA and PyOcto for comparison against the deep learning-based associators, as shown in Section 4.

**Text S2: Training the deep learning based algorithms**

The deep learning-based algorithms need to be trained prior to application, which is usually done with synthetic data. The algorithms' performance critically depends on how large and realistic the training datasets are, as well as on the adequate choice of a number of parameters that steer the training process. We here outline the training approaches for PhaseLink and GENIE. These associators require extensive synthetic data generation to expose the model to a wide range of event-station geometries. We used an approach highly similar to the one previously outlined (Section 3.2) to create synthetic arrival time data from 1D velocity models of the regions of interest.

PhaseLink is trained with a supervised learning approach, where the ground truth associations (labels) are known. We train PhaseLink for 100 epochs, saving model checkpoints at each epoch, and select the checkpoint with the lowest validation loss. During training, the selection of training parameters for PhaseLink is conducted through an iterative

process, similar to how we optimized parameters for the other associators (Section 3.4). For instance, a batch size of 64 is found to be optimal for the subduction zone scenario, whereas the higher station density of the crustal scenario necessitates a higher value of 300. Likewise, we vary the number of fake picks (n_fake) to simulate different noise levels in the training data. Higher values of fake picks were tested for the crustal scenario to reflect its higher noise environment, ultimately selecting 400 fake picks per batch. For the subduction scenario, we find that 25 fake picks provided a good model performance. Lastly, we generate 1,000,000 synthetic training samples for each scenario (for all parameter choices, refer to Table S6), ensuring that the model is exposed to a wide variety of event locations and noise conditions. The model's performance is evaluated by monitoring the validation loss and assessing the quality of the associations in preliminary runs. To illustrate the model's convergence during training, Figure S4 shows the evolution of validation loss through the 100 epochs, with the best model chosen at epoch 61.

The input of GENIE consists of any number of phase picks over an arbitrary station network, and the model is trained to predict source space-time likelihoods and source-arrival association assignments for the set of input picks. Internally, the model uses two graphs: one for the stations, and another for the source region. For each pair of source and station nodes, the misfit between observed arrivals and the theoretical arrivals is measured, and this information is then shared and transformed between both neighboring stations and source nodes with graph convolutions to detect when and where earthquakes have occurred, and the likely association assignments to these events. Through the training process the model can learn to detect subtle signatures of moveout patterns over seismic networks for both small and large events, and learn to account for the heterogeneous station distribution, noise level, and monitoring conditions. Similar to PhaseLink, GENIE is trained using supervised learning. To train the model, a diverse suite of synthetic

training data is generated, which includes sources with arbitrary positions and highly variable levels of noise and observational characteristics. Key training parameters include the maximum moveout distances of sources, the level of travel time noise, the amount of false and corrupted picks, and the maximum rate of events (Table S7). Additionally, users must set the target source region, velocity model, and choose kernel sizes for the space-time Gaussian labels. Hence, while the model can handle changing station distributions between training and future applications, for applying the model to entirely new regions it is helpful to retrain the model so that the chosen kernel sizes, velocity model, and spatial extent of the source graphs are all well calibrated to the study region of interest. The number of epochs, learning rate, and batch size can also be varied, however these are typically set to nominal values.

**Text S3: Testing different station densities**

To test the effect of different seismic network densities on associator performance, we conducted an additional test by modifying the California-based crustal scenario. While a higher station density can enable more accurate event detection, having more closely-spaced stations also increases the possibility of cross-associating phases to the wrong event in the case of dense seismicity such as aftershock sequences. We created two distinct station configurations derived from the crustal scenario within the same geographic area of $1.5° \times 1.5°$ (Figure S5), using real-world seismic networks from the Southern California Seismic Network (SCSN). The low-density configuration comprises a total of 21 stations, the high-density configuration has 91 stations.

When repeating the different runs from the crustal scenario (see Section 4) with the modified station sets, we find that the precision of most associators decreases significantly, which is mainly due to our choice of the same association threshold (10 picks) for all runs. What we find is an inherent trade-off between event detection sensitivity and precision. In

high-density networks, a low association threshold enhances sensitivity to smaller events but increases the risk of false associations due to random noise picks. Conversely, increasing the threshold improves precision by filtering out false associations, but will reduce sensitivity. Notably, GENIE is less affected by this issue. It consistently maintains high precision and recall across both scenarios without the need to adjust the association threshold or other parameters, and even in the high-density crustal case worked well with 10 required picks while maintaining a low rate of false positives. This independence of parameter optimization appears to be an important advantage of neural network based methods. While GENIE could be applied in all three cases (high density, low density, original) with the same training, we found that PhaseLink needs to be re-trained in order to perform well across different station densities (see Figure S6).

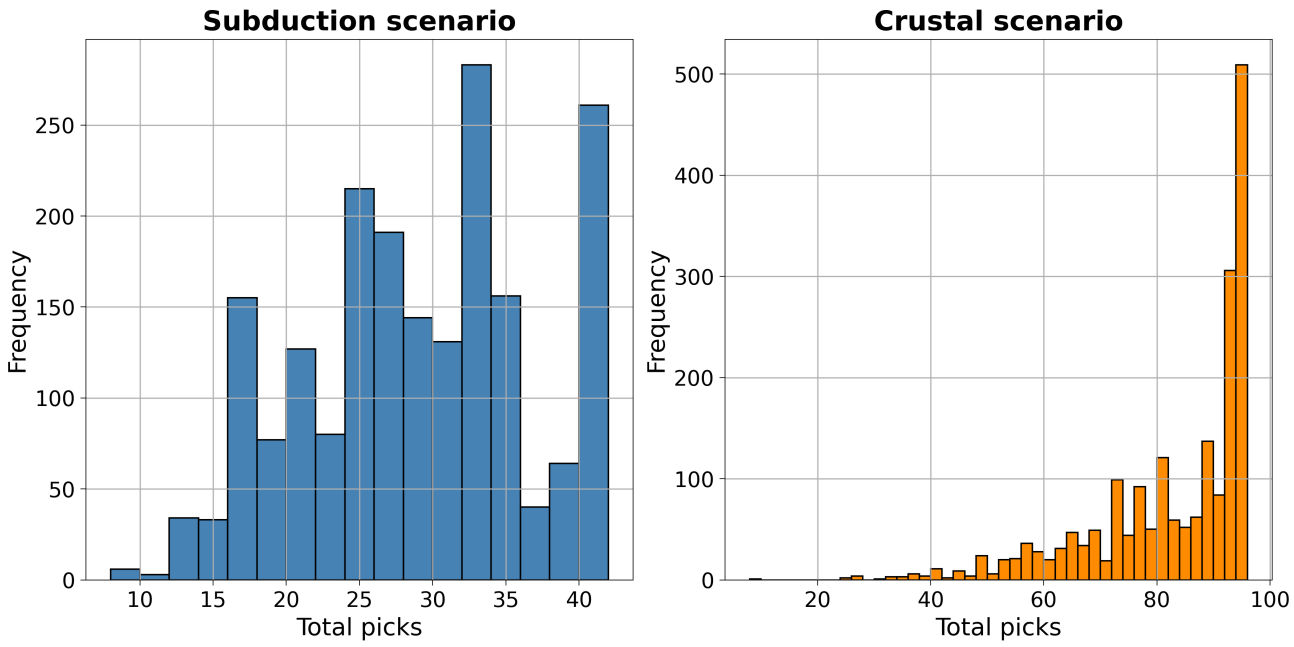


**Figure S1.** Distributions of event pick count for 2000 events in subduction zone (left) and crustal (right) scenario.

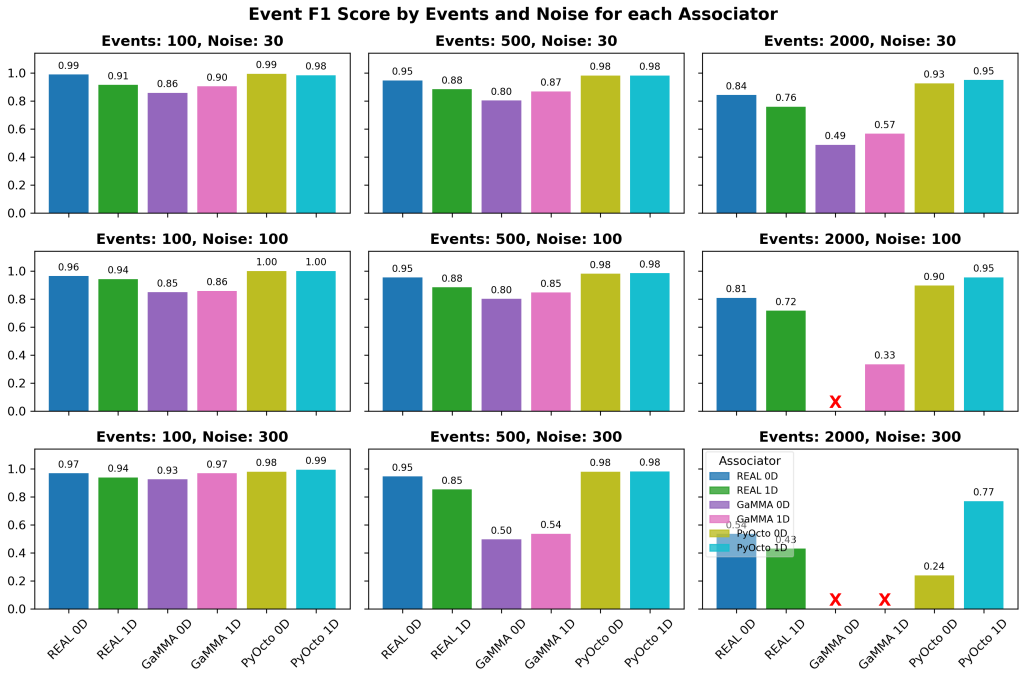


**Figure S2.** Comparison of Event F1 Score for GaMMA, REAL, and PyOcto using 0D (homogeneous) and 1D velocity models across different noise levels and event densities in subduction zone scenario. For the processing times of the different runs, please refer to Table S8.

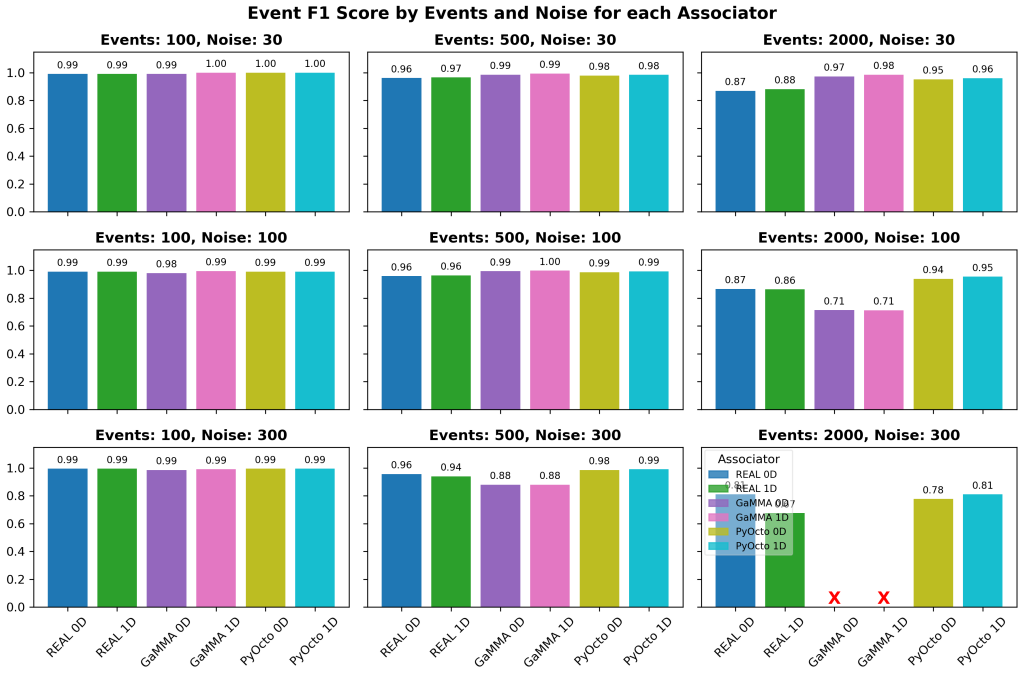


**Figure S3.** Comparison of Event F1 Score for GaMMA, REAL, and PyOcto using 0D (homogeneous) and 1D velocity models across different noise levels and event densities in the crustal scenario. For the processing times of the different runs, please refer to Table S8.

**Table S1.** Dataset statistics of subduction scenario.

| Events | Noise (%) | Event picks | False picks | Total picks | Picks per event |
|---|---|---|---|---|---|
| 100 | 30 | 2794 | 838 | 3632 | 27.940 |
| 100 | 100 | 2912 | 2912 | 5824 | 29.120 |
| 100 | 300 | 2946 | 8838 | 11784 | 29.460 |
| 500 | 30 | 14150 | 4244 | 18394 | 28.300 |
| 500 | 100 | 13864 | 13864 | 27728 | 27.728 |
| 500 | 300 | 14100 | 42300 | 56400 | 28.200 |
| 2000 | 30 | 55874 | 16762 | 72636 | 27.937 |
| 2000 | 100 | 55822 | 55822 | 111644 | 27.911 |
| 2000 | 300 | 55190 | 165570 | 220760 | 27.595 |

**Table S2.** Dataset statistics of shallow seismicity scenario.

| Events | Noise (%) | Event picks | False picks | Total picks | Picks per event |
|---|---|---|---|---|---|
| 100 | 30 | 8178 | 2452 | 10630 | 81.780 |
| 100 | 100 | 8298 | 8298 | 16596 | 82.980 |
| 100 | 300 | 8124 | 24372 | 32496 | 81.240 |
| 500 | 30 | 40490 | 12146 | 52636 | 80.980 |
| 500 | 100 | 40842 | 40842 | 81684 | 81.684 |
| 500 | 300 | 40788 | 122364 | 163152 | 81.576 |
| 2000 | 30 | 163270 | 48980 | 212250 | 81.635 |
| 2000 | 100 | 163240 | 163240 | 326480 | 81.620 |
| 2000 | 300 | 161874 | 485622 | 647496 | 80.937 |

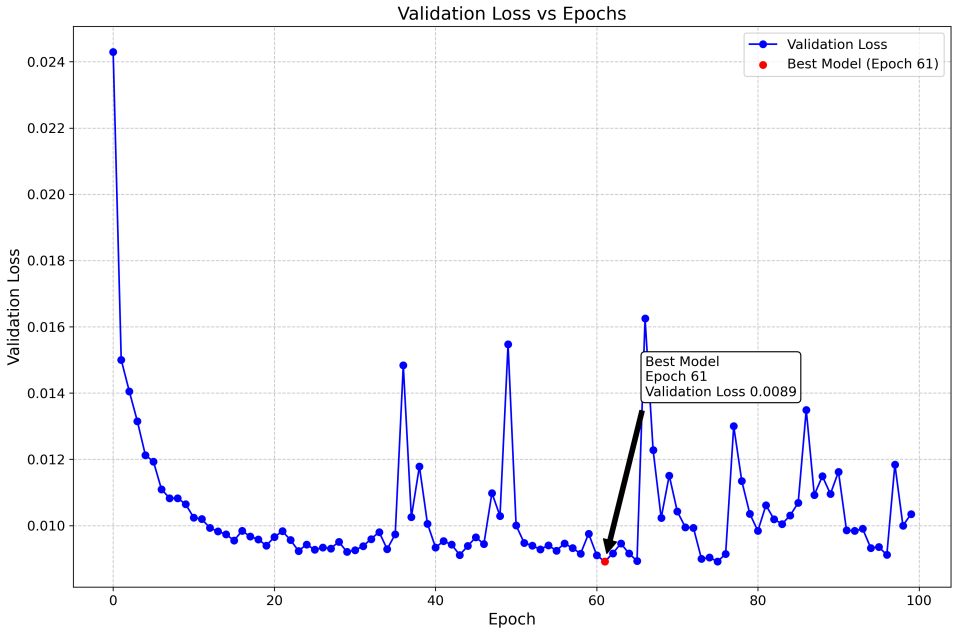


**Figure S4.** Validation loss vs. epochs for an example training of PhaseLink. The plot shows the validation loss at each epoch during the training process. The best model, indicated by the red marker, was achieved here at epoch 61 with a validation loss of 0.0089. The plot shows a general trend of decreasing validation loss as the training progresses, demonstrating the model's improvement over time.

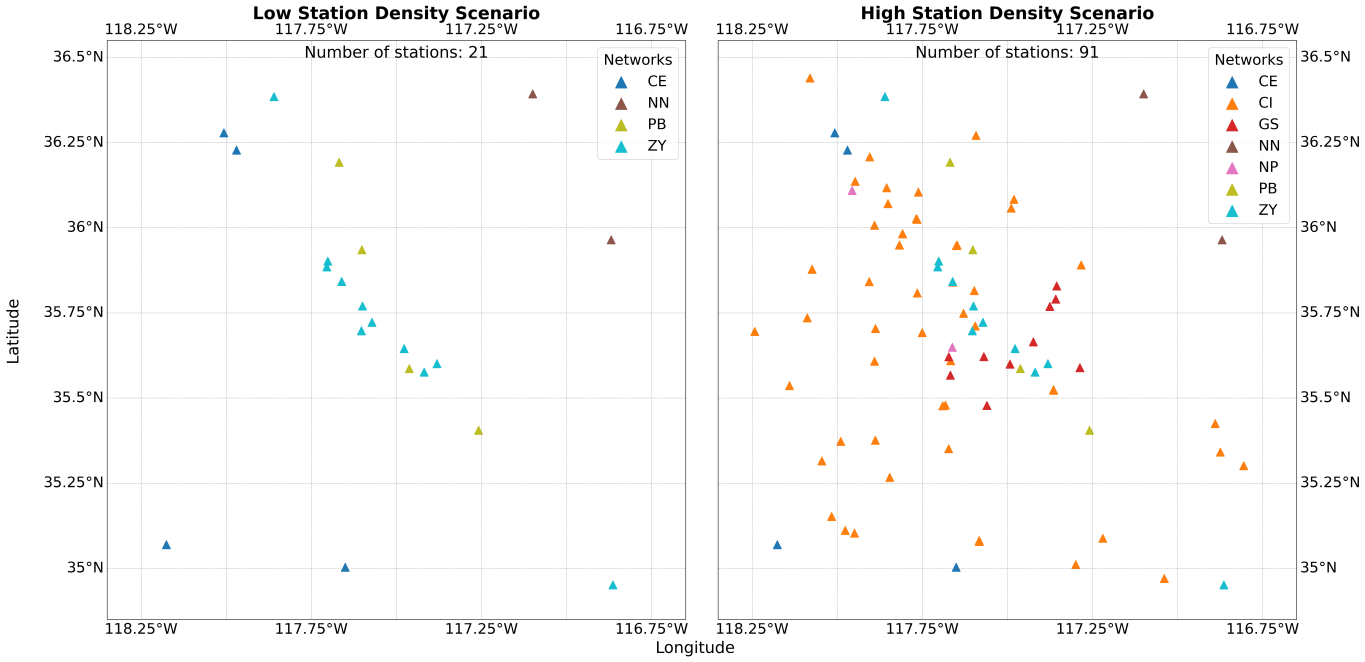


**Figure S5.** Station density configurations within the 1.5° x 1.5° area of the crustal scenario. **Left:** Low station density configuration (21 stations). **Right:** High station density configuration (91 stations). The seismic networks (CE, PB, ZY, NN, CI, GS, NP) are indicated in the legend by colors.

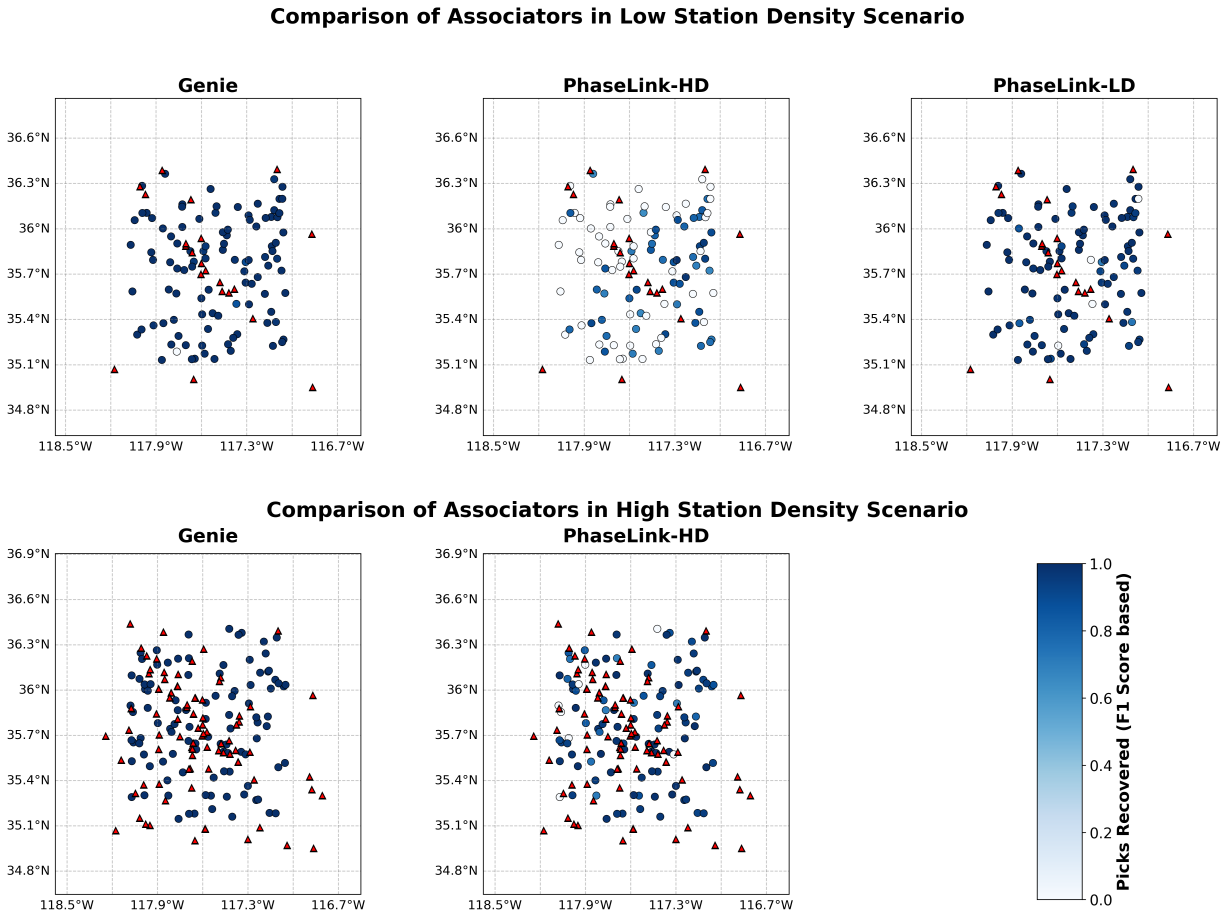


**Figure S6.** Comparison of associators for low (top) and high station density (bottom) scenarios. Triangles indicate station locations and circles represent events, colored by the achieved F1 score on pick level by each associator. The increase in station density (bottom row) generally improves event association and pick recovery, as shown by the more densely populated and darker-colored events in those subplots. PhaseLink-HD and PhaseLink-LD refer to the PhaseLink associator that was trained with the high-density and low-density scenario, respectively. GENIE was only trained on the high-density scenario here.

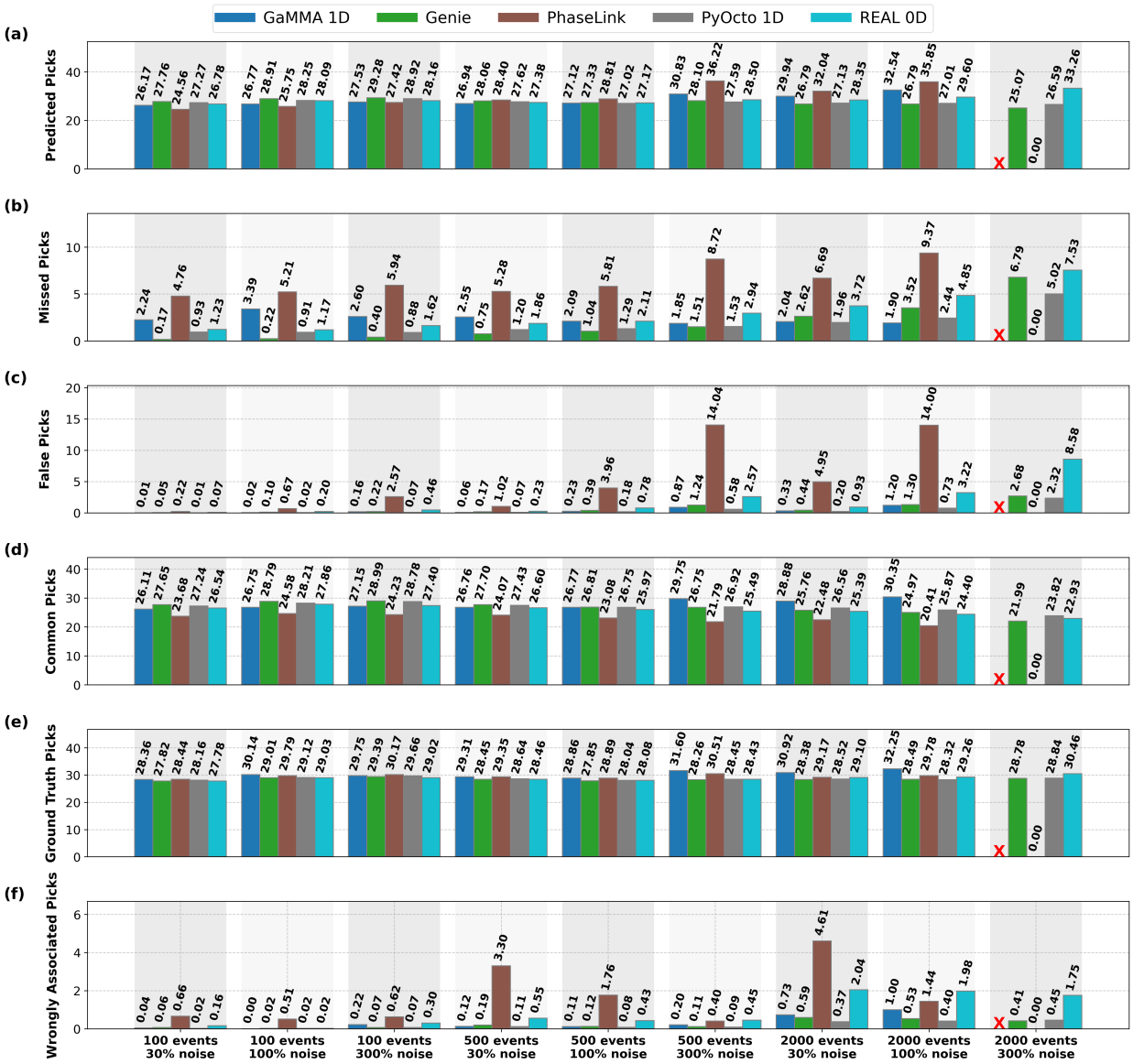


**Figure S7.** Columns step from the easiest subduction scenario catalog (100 events / 30% noise) to the hardest (2000 events / 300% noise). Metrics reflect only those seismic events that met or exceeded a 50% matching threshold with the ground truth synthetic dataset. Bars are color-coded by associator (legend, top); numbers above each bar give the mean value per event, while blank slots with a red "×" denote runs that did not finish. Sub-plots: (a) Predicted picks returned by the associator; (b) Missed picks that should have been returned but were not; (c) False picks newly attached to an event; (d) Common picks correctly shared between prediction and ground truth; (e) total ground-truth picks available (baseline); (f) picks wrongly associated with a different event. The slight algorithm-to-algorithm variation in panel (e) arises because averages are taken only over the events each method recovered.

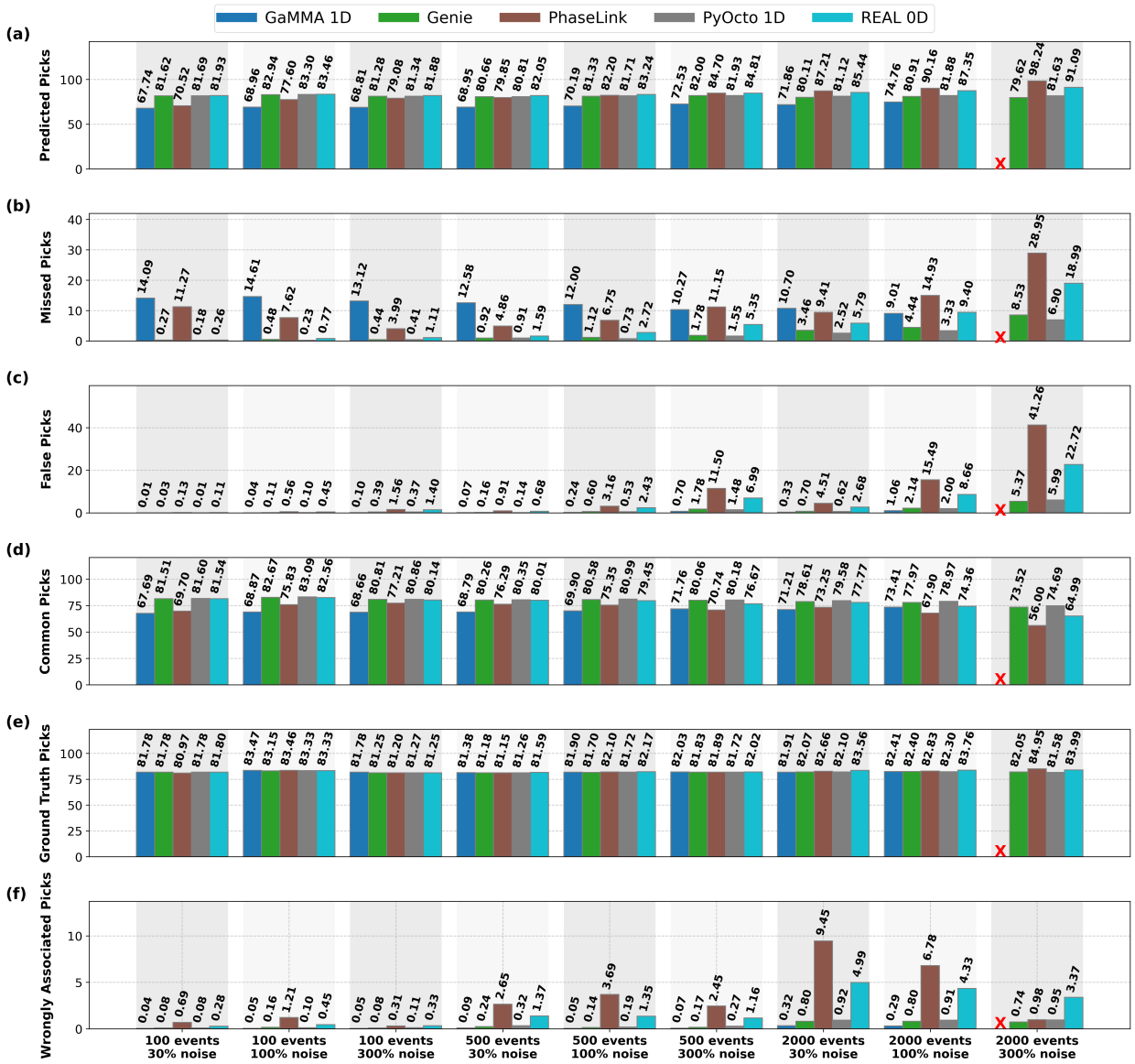


**Figure S8.** Columns step from the easiest crustal scenario catalog (100 events / 30% noise) to the hardest (2000 events / 300% noise). Metrics reflect only those seismic events that met or exceeded a 50% matching threshold with the ground truth synthetic dataset. Bars are color-coded by associator (legend, top); numbers above each bar give the mean value per event, while blank slots with a red "×" denote runs that did not finish. Sub-plots: (a) Predicted picks returned by the associator; (b) Missed picks that should have been returned but were not; (c) False picks newly attached to an event; (d) Common picks correctly shared between prediction and ground truth; (e) total ground-truth picks available (baseline); (f) picks wrongly associated with a different event. The slight algorithm-to-algorithm variation in panel (e) arises because averages are taken only over the events each method recovered.

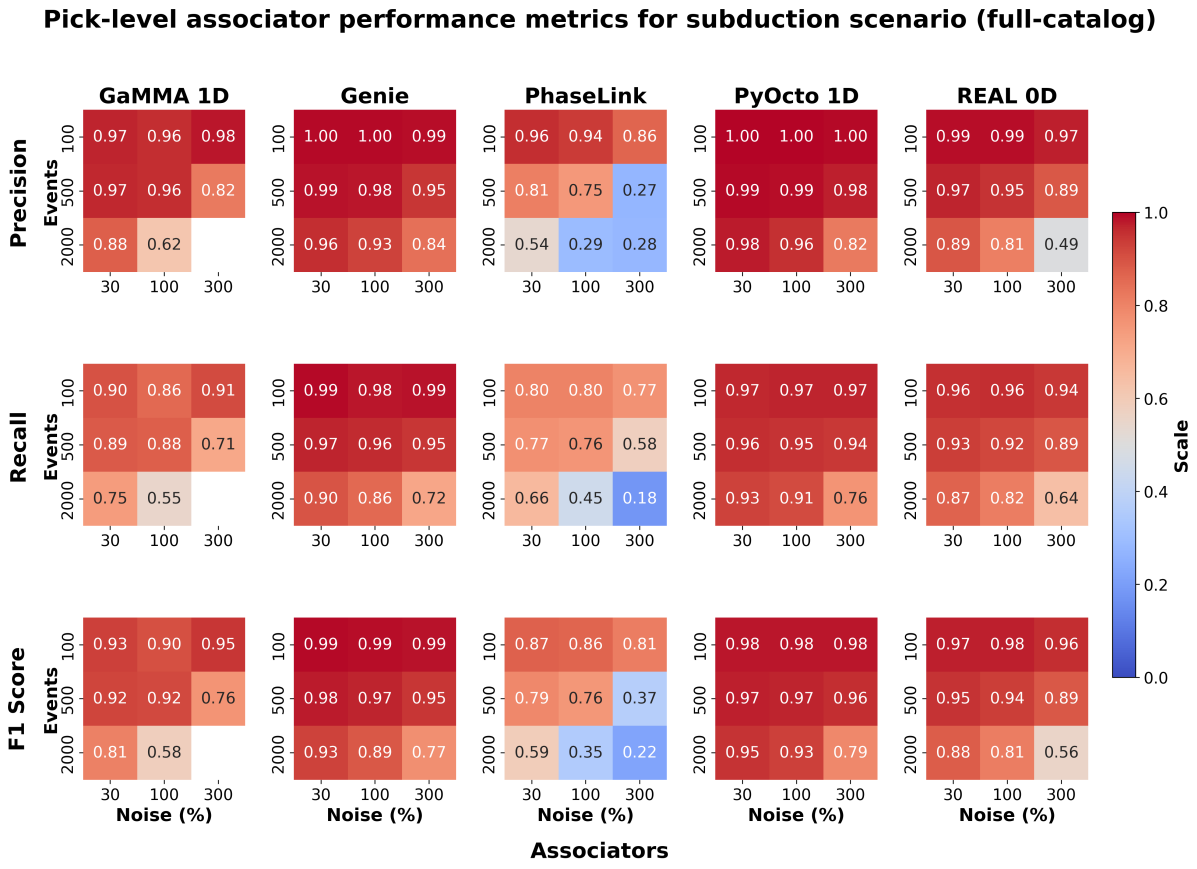


**Figure S9.** Pick-level performance of five associators in the subduction scenario after accounting for all predicted events, including false-positive events (no 50 %-match filter). Columns group the associators (GaMMA 1D, Genie, PhaseLink, PyOcto 1D, REAL 0D). Rows give the three metrics: Precision, Recall, and F1 score, shown once as row labels on the left. Within each heat-map the x-axis steps through increasing catalog noise (30%, 100%, 300% false picks) and the y-axis through higher pick rates (100, 500, 2000 events). The three rows display, from top to bottom, precision, recall, and F1-score; warmer colors indicate better performance according to the shared scale bar, which now spans the full range 0 - 1. Blank cells mark runs that were not completed. Each panel shows the mean performance derived from events that matched the synthetic ground truth.

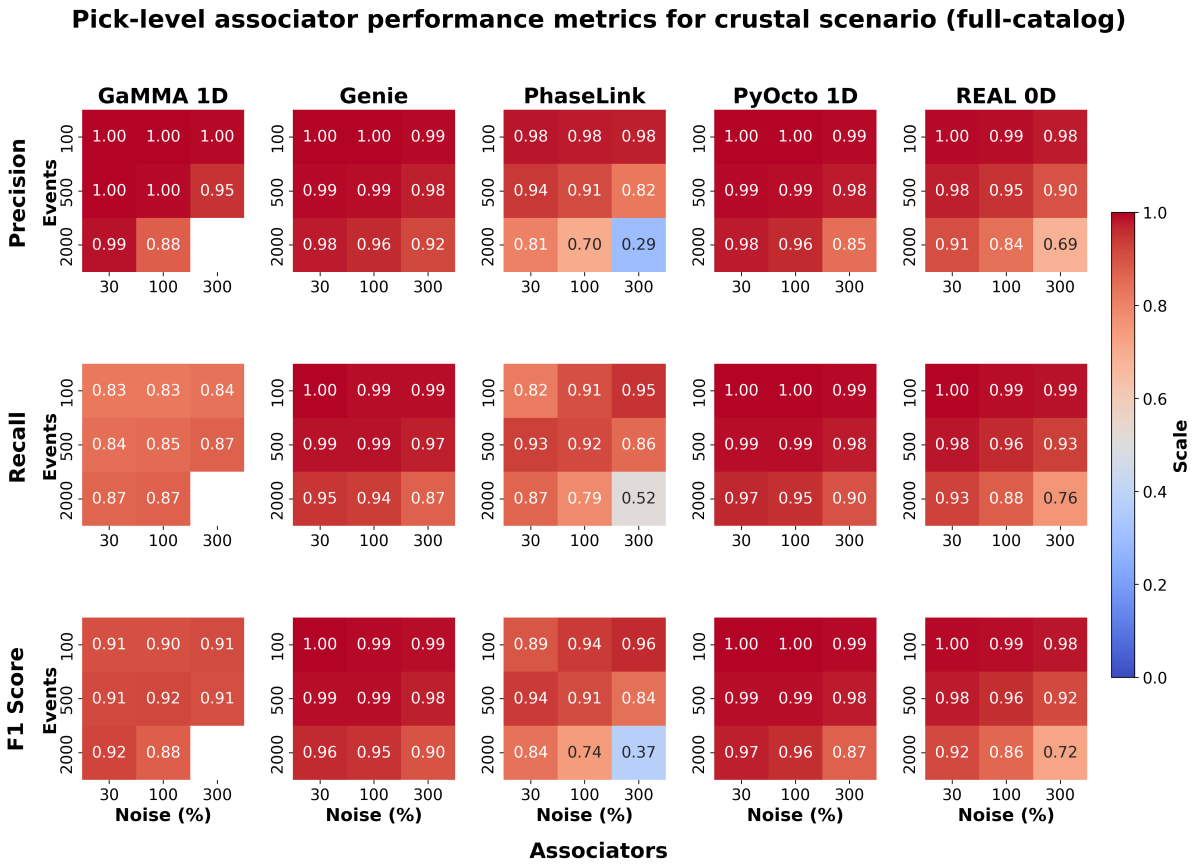


**Figure S10.** Pick-level performance of five associators in the crustal scenario after accounting for all predicted events, including false-positive events (no 50 %-match filter). Columns group the associators (GaMMA 1D, Genie, PhaseLink, PyOcto 1D, REAL 0D). Rows give the three metrics: Precision, Recall, and F1 score, shown once as row labels on the left. Within each heat-map the x-axis steps through increasing catalog noise (30%, 100%, 300% false picks) and the y-axis through higher pick rates (100, 500, 2000 events). The three rows display, from top to bottom, precision, recall, and F1-score; warmer colors indicate better performance according to the shared scale bar, which now spans the full range 0 - 1. Blank cells mark runs that were not completed. Each panel shows the mean performance derived from events that matched the synthetic ground truth.

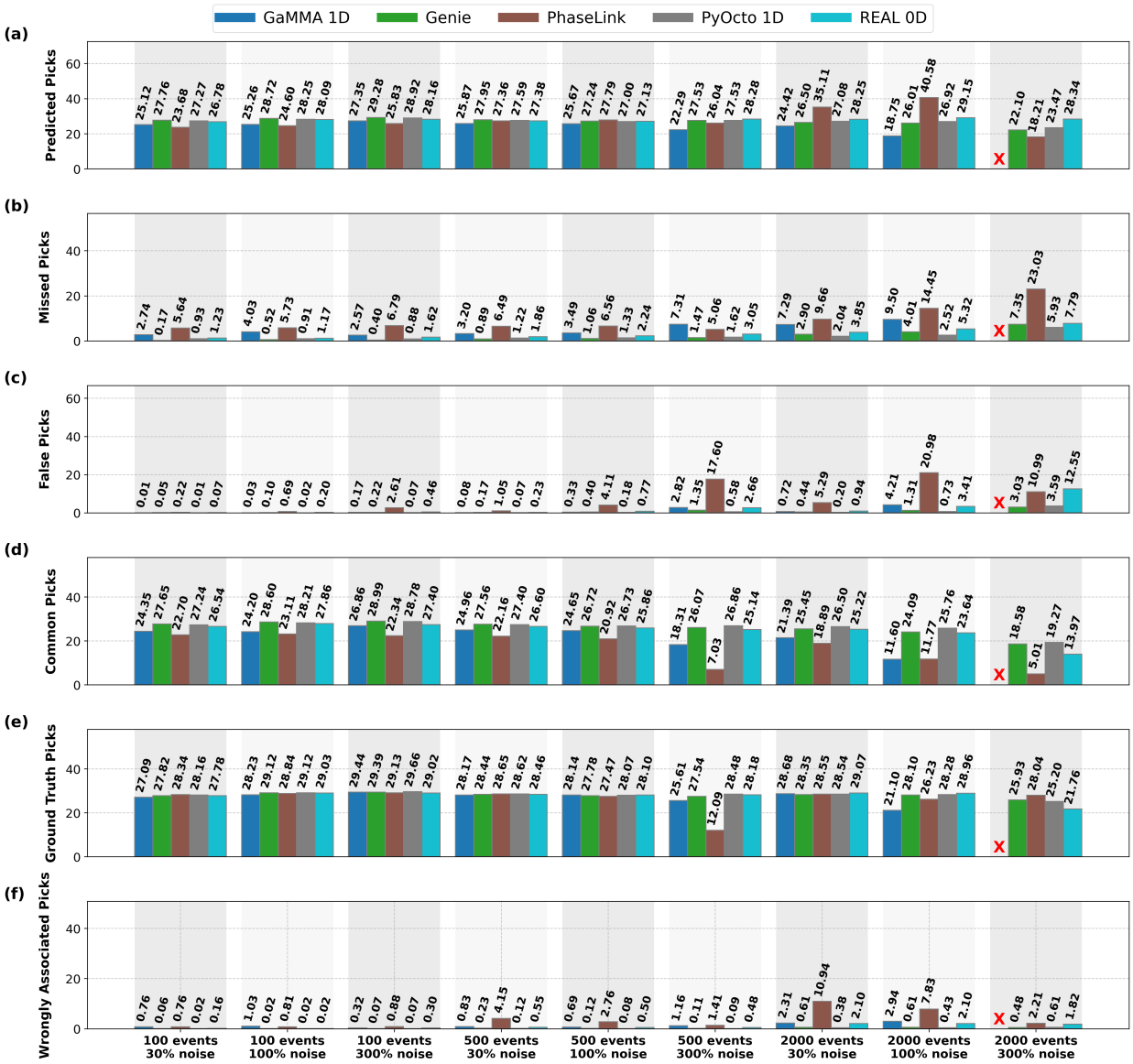


**Figure S11.** Columns step from the easiest subduction scenario catalog (100 events / 30% noise) to the hardest (2000 events / 300% noise). No 50 % match threshold is applied here; every event returned by an association, including false positives, is evaluated. Bars are color-coded by associator (legend, top); numbers above each bar give the mean value per event, while blank slots with a red "×" denote runs that did not finish. Sub-plots: (a) Predicted picks returned by the associator; (b) Missed picks that should have been returned but were not; (c) False picks newly attached to an event; (d) Common picks correctly shared between prediction and ground truth; (e) total ground-truth picks available (baseline); (f) picks wrongly associated with a different event. The slight algorithm-to-algorithm variation in panel (e) arises because averages are taken only over the events each method recovered.

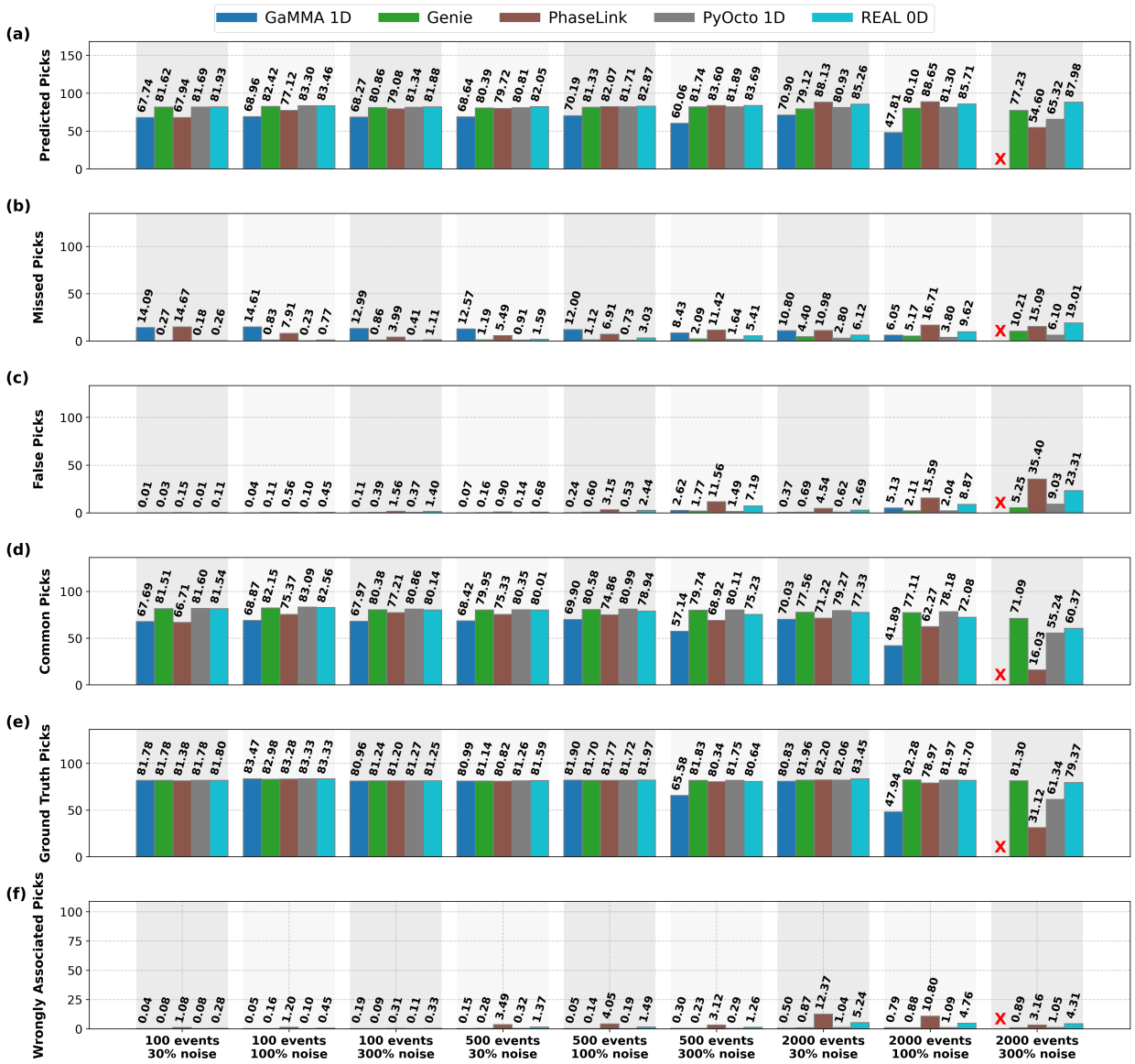


**Figure S12.** Columns step from the easiest crustal scenario catalog (100 events / 30% noise) to the hardest (2000 events / 300% noise). No 50 % match threshold is applied here; every event returned by an association, including false positives, is evaluated. Bars are color-coded by associator (legend, top); numbers above each bar give the mean value per event, while blank slots with a red "×" denote runs that did not finish. Sub-plots: (a) Predicted picks returned by the associator; (b) Missed picks that should have been returned but were not; (c) False picks newly attached to an event; (d) Common picks correctly shared between prediction and ground truth; (e) total ground-truth picks available (baseline); (f) picks wrongly associated with a different event. The slight algorithm-to-algorithm variation in panel (e) arises because averages are taken only over the events each method recovered.

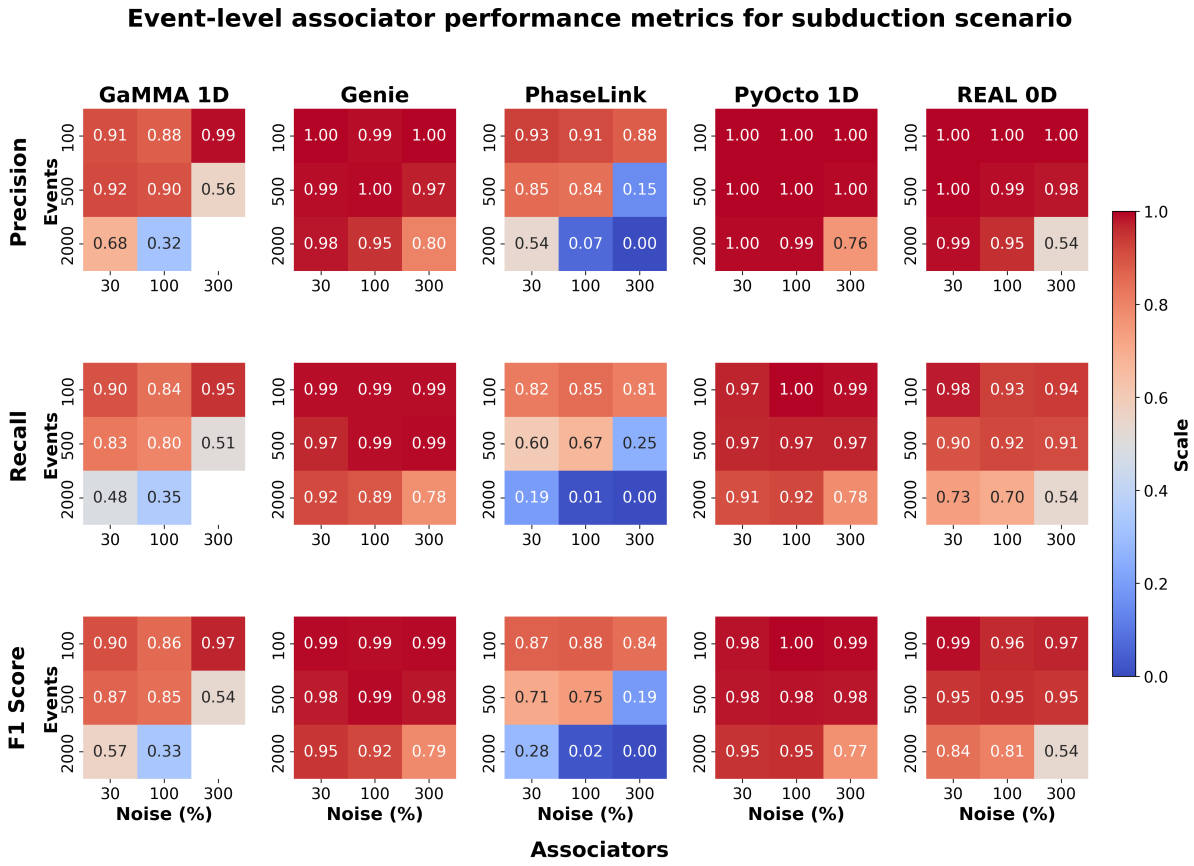


**Figure S13.** Event-level performance of five associators in the subduction scenario. Each column corresponds to an associator (GaMMA 1D, GENIE, PhaseLink, PyOcto 1D, REAL 0D). Within every heatmap, the x-axis steps through higher proportions of false picks (30%, 100%, 300%), while the y-axis steps through denser catalogues (100, 500, 2000 events). The three rows show precision, recall, and F1-score (top-to-bottom); warmer colors indicate better performance according to the shared scale bar. Blank cells mark runs that did not complete.

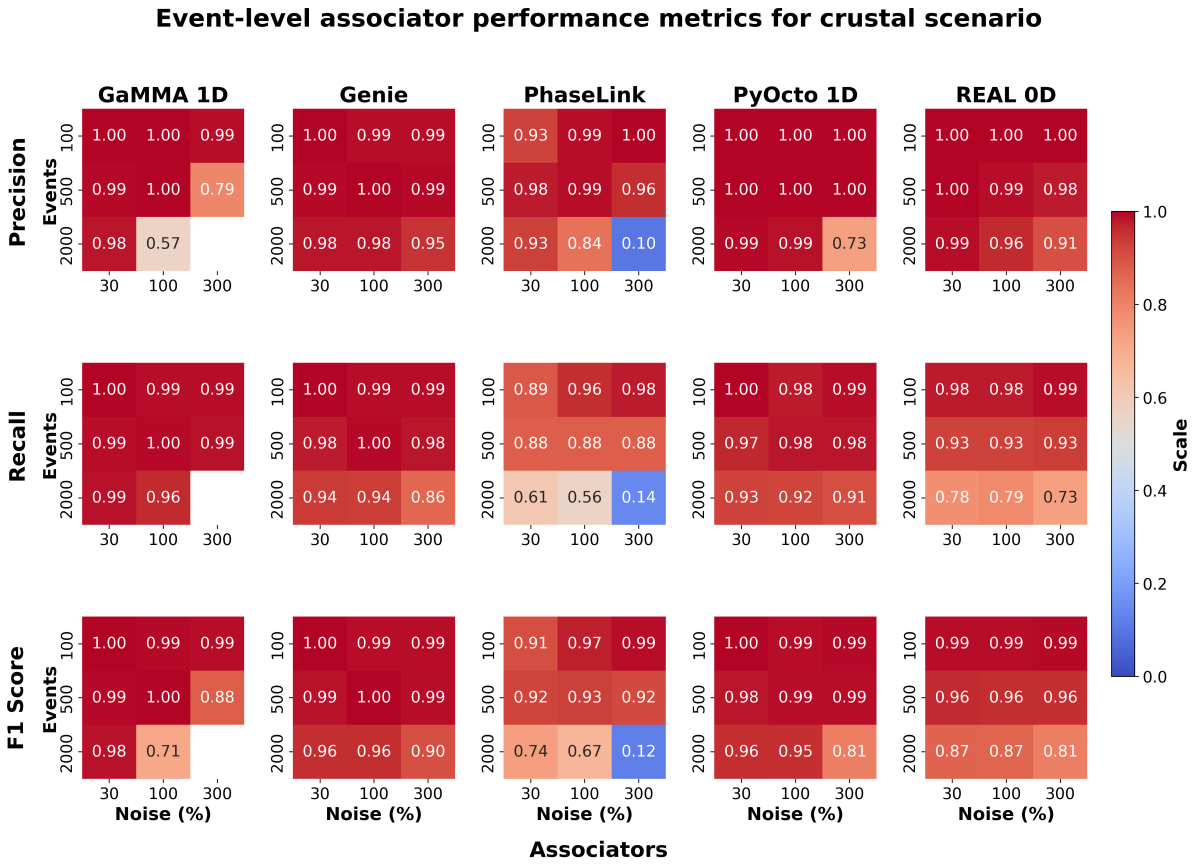


**Figure S14.** Event-level performance of five associators in the crustal scenario. Each column corresponds to an associator (GaMMA 1D, GENIE, PhaseLink, PyOcto 1D, REAL 0D). Within every heatmap, the x-axis steps through higher proportions of false picks (30%, 100%, 300%), while the y-axis steps through denser catalogues (100, 500, 2000 events). The three rows show precision, recall, and F1-score (top-to-bottom); warmer colors indicate better performance according to the shared scale bar. Blank cells mark runs that did not complete.

**Table S3.** Parameters for GaMMA 1D in the Crustal and Subduction Scenarios. Italicized values represent parameters that were adjusted during the tuning process, while non-italicized values indicate parameters that were kept fixed.

| Parameter | Crustal Scenario | Subduction Scenario |
|---|---|---|
| `use amplitude` | False | False |
| `vel p` | *6.2* | *7.0* |
| `vel s` | *3.4* | *4.0* |
| `method` | BGMM | BGMM |
| `use dbscan` | True | True |
| `oversample factor` | *3* | *2* |
| `dbscan eps` | *7* | *20* |
| `dbscan min samples` | *20* | *5* |
| `min picks per eq` | 10 | 10 |
| `max sigma11` | *2.0* | *2.0* |
| `max sigma22` | *1.0* | *1.0* |
| `max sigma12` | *1.0* | *1.0* |
| `ncpu` | 25 | 25 |
| `1D velocity model` | True | True |
| `x(km)` | [385, 520] | [250, 600] |
| `y(km)` | [3860, 4040] | [7200, 8000] |
| `z(km)` | [0, 30.0] | [0, 250] |
| `local crs` | 32611 | 9155 |

**Table S4.** Parameters for PyOcto 1D in the Crustal and Subduction Scenarios. Italicized values represent parameters that were adjusted during the tuning process, while non-italicized values indicate parameters that were kept fixed.

| Parameter | Crustal Scenario | Subduction Scenario |
|---|---|---|
| `spatial limits xlim` | [385, 520] | [250.0, 600.0] |
| `spatial limits ylim` | [3860, 4040] | [7200.0, 8000.0] |
| `spatial limits zlim` | [0, 30] | [0, 250.0] |
| `association cutoff distance` | *200* | *350* |
| `time before` | *100.0* | *300.0* |
| `min node size` | *10* | *10* |
| `min node size location` | *2.5* | *1.5* |
| `pick match tolerance` | *0.8* | *0.8* |
| `min interevent time` | *3.0* | *3.0* |
| `max pick overlap` | *4* | *4* |
| `n picks` | 10 | 10 |
| `n p picks` | 5 | 5 |
| `n s picks` | 5 | 5 |
| `n p and s picks` | 4 | 4 |
| `refinement iterations` | *3* | *3* |
| `time slicing` | *1200.0* | *1200.0* |
| `location split depth` | *6* | *6* |
| `location split return` | *4* | *4* |
| `min pick fraction` | *0.0* | *0.25* |
| `n threads` | 25 | 25 |
| `VelMod1D` | True | True |
| `velocity model tolerance` | *1.0* | *1.0* |
| `local crs` | 32611 | 9155 |
| `tt table grid spacing` | *1.0* | *0.5* |
| `tt table x extent` | *300* | *500* |
| `tt table y extent` | *300* | *800* |

**Table S5.** Parameters for REAL 0D in the Crustal and Subduction Scenarios. Italicized values represent parameters that were adjusted during the tuning process, while non-italicized values indicate parameters that were kept fixed.

| Parameter | Crustal Scenario | Subduction Scenario |
|---|---|---|
| `tt config dist` | *4* | *9* |
| `tt config dep` | *30* | *250* |
| `tt config ddist` | *0.6* | *1.0* |
| `tt config ddep` | *1* | *8* |
| `1D velocity model` | False | False |
| `latitude` | 35.0 | -21.18148 |
| `R rx` | *1* | *1* |
| `R rh` | *30* | *250* |
| `R tdx` | *0.1* | *0.1* |
| `R tdh` | *8* | *10* |
| `R tint` | *0.1* | *0.1* |
| `V vp0` | *6.2* | *6.8* |
| `V vs0` | *3.3* | *4.0* |
| `V s_vp0` | *5.4* | *5.3* |
| `V s_vs0` | *3.3* | *3.1* |
| `V ielev` | 1 | 1 |
| `S np0` | 4 | 4 |
| `S ns0` | *4* | *4* |
| `S nps0` | 10 | 10 |
| `S npsboth0` | 4 | 4 |
| `S std0` | *2.0* | *2.0* |
| `S dtps` | *0* | *0* |
| `S nrt` | *2* | *2* |
| `S drt` | *2* | *2* |
| `S nxd` | *0.4* | *0.4* |
| `S rsel` | *6* | *6* |

**Table S6.** Parameters for PhaseLink in the Crustal and Subduction Scenarios. Italicized values represent parameters that were adjusted during the tuning process, while non-italicized values indicate parameters that were kept fixed.

| Parameter | Crustal Scenario | Subduction Scenario |
|---|---|---|
| `t win` | *250* | *120* |
| `n epochs` | *100* | *100* |
| `n max picks` | *300* | *120* |
| `batch size` | *64* | *64* |
| `n min nucl` | *12* | *6* |
| `n min merge` | *2* | *2* |
| `n min det` | 10 | 10 |
| `avg eve sep` | *20* | *12* |
| `pr min` | *0.5* | *0.5* |
| `n train samp` | 1000000 | 1000000 |
| `n min radius` | *8* | *8* |
| `n fake` | *400* | *25* |
| `max event depth` | 30 | 250 |
| `min hypo dist` | *50.0* | *80.0* |
| `max hypo dist` | *80.0* | *450.0* |
| `max pick error` | 1.0 | 1.0 |
| `min pick dist` | 0.5 | 0.5 |
| `min sep` | *0.6* | *0.6* |
| `lat min` | 34.87 | -25.0 |
| `lat max` | 36.5 | -17.0 |
| `lon min` | 118.28 | -71.0 |
| `lon max` | 116.7 | -66.0 |

**Table S7.** Chosen parameters for GENIE 1D in the Crustal and Subduction Scenarios. Italicized entries were partially tuned with 2−3 rounds of re-training, all other values were chosen to reflect the characteristic spatial scale and expected event rates of either scenario.

| Parameter | Crustal Scenario | Subduction Scenario |
|---|---|---|
| `k_sta_edges` | 8 | 8 |
| `k_spc_edges` | 15 | 15 |
| `n_of_spatial_nodes` | 1000 | 1500 |
| `kernel_sig_t` | 3.0 | 8.0 |
| `src_t_kernel` | 3.0 | 8.0 |
| `src_x_kernel` | 15000 | 45000 |
| `spc_random` | 15000 | 10000 |
| `spc_thresh_ran` | 15000 | 135000 |
| `sig_t` | 0.01 | *0.0075* |
| `min_sta_arrival` | 12 | 8 |
| `thresh_noise_max` | 2.25 | *0.75* |
| `total_bias` | 0.01 | *0.0075* |
| `dist_range` | [5000, 250000] | [100000, 1490000] |
| `max_rate_events` | *225* | *280* |
| `max_false_events` | *650* | *650* |
| `miss_pick_fraction` | [0.05, 0.35] | [0.05, 0.35] |
| `thresh` | 0.6 | 0.6 |
| `thresh_assoc` | 0.6 | 0.6 |
| `tc_win` | 2.5 | 8.0 |
| `sp_win` | 12500 | 45000 |
| `d_win` | 0.2 | 0.45 |
| `d_win_depth` | 20000 | 50000 |
| Latitude | [34.82°, 36.55°] N | [-25.0°, -17.0°] N |
| Longitude | [-118.33°, -116.65°] E | [-71.0°, -66.0°] E |
| Depths | [-35, 5] km | [-250, 5] km |

**Table S8.** Performance comparison using homogeneous (0D) and 1D velocity models

| Events | Noise | Associator | Event_Precision | Event_Recall | Event_F1_Score | Runtime (s) |
|---|---|---|---|---|---|---|
| 100 | 30 | GaMMA 0D | 0.84 | 0.88 | 0.86 | 2.91 |
| 100 | 30 | GaMMA 1D | 0.91 | 0.90 | 0.90 | 5.03 |
| 100 | 30 | PyOcto 0D | 1.00 | 0.99 | 0.99 | 1.29 |
| 100 | 30 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 3.16 |
| 100 | 30 | REAL 0D | 1.00 | 0.98 | 0.99 | 60.98 |
| 100 | 30 | REAL 1D | 0.98 | 0.86 | 0.91 | 72.06 |
| 100 | 100 | GaMMA 0D | 0.86 | 0.84 | 0.85 | 2.93 |
| 100 | 100 | GaMMA 1D | 0.88 | 0.84 | 0.86 | 5.28 |
| 100 | 100 | PyOcto 0D | 1.00 | 1.00 | 1.00 | 0.99 |
| 100 | 100 | PyOcto 1D | 1.00 | 1.00 | 1.00 | 1.42 |
| 100 | 100 | REAL 0D | 1.00 | 0.93 | 0.96 | 169.51 |
| 100 | 100 | REAL 1D | 0.99 | 0.90 | 0.94 | 209.14 |
| 100 | 300 | GaMMA 0D | 0.92 | 0.93 | 0.93 | 6.10 |
| 100 | 300 | GaMMA 1D | 0.99 | 0.95 | 0.97 | 6.71 |
| 100 | 300 | PyOcto 0D | 0.98 | 0.98 | 0.98 | 42.68 |
| 100 | 300 | PyOcto 1D | 1.00 | 0.99 | 0.99 | 1.19 |
| 100 | 300 | REAL 0D | 1.00 | 0.94 | 0.97 | 453.73 |
| 100 | 300 | REAL 1D | 0.96 | 0.92 | 0.94 | 626.48 |
| 500 | 30 | GaMMA 0D | 0.83 | 0.78 | 0.80 | 41.84 |
| 500 | 30 | GaMMA 1D | 0.92 | 0.83 | 0.87 | 17.73 |
| 500 | 30 | PyOcto 0D | 0.99 | 0.97 | 0.98 | 91.96 |
| 500 | 30 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 6.74 |
| 500 | 30 | REAL 0D | 1.00 | 0.90 | 0.95 | 306.16 |
| 500 | 30 | REAL 1D | 0.96 | 0.82 | 0.88 | 424.37 |
| 500 | 100 | GaMMA 0D | 0.86 | 0.75 | 0.80 | 57.15 |
| 500 | 100 | GaMMA 1D | 0.90 | 0.80 | 0.85 | 25.74 |
| 500 | 100 | PyOcto 0D | 0.99 | 0.97 | 0.98 | 45.06 |
| 500 | 100 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 12.29 |
| 500 | 100 | REAL 0D | 0.99 | 0.92 | 0.95 | 754.36 |
| 500 | 100 | REAL 1D | 0.94 | 0.83 | 0.88 | 931.85 |
| 500 | 300 | GaMMA 0D | 0.53 | 0.47 | 0.50 | 584.25 |
| 500 | 300 | GaMMA 1D | 0.56 | 0.51 | 0.54 | 145.93 |
| 500 | 300 | PyOcto 0D | 0.99 | 0.97 | 0.98 | 85.44 |
| 500 | 300 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 33.46 |
| 500 | 300 | REAL 0D | 0.98 | 0.91 | 0.95 | 2032.63 |
| 500 | 300 | REAL 1D | 0.90 | 0.81 | 0.85 | 2226.60 |
| 2000 | 30 | GaMMA 0D | 0.58 | 0.42 | 0.49 | 1789.12 |
| 2000 | 30 | GaMMA 1D | 0.68 | 0.48 | 0.57 | 210.15 |
| 2000 | 30 | PyOcto 0D | 0.96 | 0.89 | 0.93 | 201.67 |
| 2000 | 30 | PyOcto 1D | 1.00 | 0.91 | 0.95 | 96.97 |
| 2000 | 30 | REAL 0D | 0.99 | 0.73 | 0.84 | 1300.54 |
| 2000 | 30 | REAL 1D | 0.93 | 0.64 | 0.76 | 1427.40 |
| 2000 | 100 | GaMMA 1D | 0.32 | 0.35 | 0.33 | 3598.16 |
| 2000 | 100 | PyOcto 0D | 0.91 | 0.88 | 0.90 | 417.81 |
| 2000 | 100 | PyOcto 1D | 0.99 | 0.92 | 0.95 | 131.74 |
| 2000 | 100 | REAL 0D | 0.95 | 0.70 | 0.81 | 2848.73 |
| 2000 | 100 | REAL 1D | 0.87 | 0.61 | 0.72 | 2830.46 |
| 2000 | 300 | PyOcto 0D | 0.16 | 0.44 | 0.24 | 1023.65 |
| 2000 | 300 | PyOcto 1D | 0.76 | 0.78 | 0.77 | 1790.04 |
| 2000 | 300 | REAL 0D | 0.54 | 0.54 | 0.54 | 6706.46 |
| 2000 | 300 | REAL 1D | 0.43 | 0.43 | 0.43 | 5274.96 |

**Table S9.** Subduction Zone Scenario: event-level evaluation of seismic phase associators

| Events | Noise | Associator | Event_Precision | Event_Recall | Event_F1_Score | Runtime (s) |
|---|---|---|---|---|---|---|
| 100 | 30 | GaMMA 1D | 0.91 | 0.90 | 0.90 | 5.03 |
| 100 | 30 | Genie | 1.00 | 0.99 | 0.99 | 294.52 |
| 100 | 30 | PhaseLink | 0.93 | 0.82 | 0.87 | 4.99 |
| 100 | 30 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 3.16 |
| 100 | 30 | REAL 0D | 1.00 | 0.98 | 0.99 | 60.98 |
| 100 | 100 | GaMMA 1D | 0.88 | 0.84 | 0.86 | 5.28 |
| 100 | 100 | Genie | 0.99 | 0.99 | 0.99 | 313.33 |
| 100 | 100 | PhaseLink | 0.91 | 0.85 | 0.88 | 3.79 |
| 100 | 100 | PyOcto 1D | 1.00 | 1.00 | 1.00 | 1.42 |
| 100 | 100 | REAL 0D | 1.00 | 0.93 | 0.96 | 169.51 |
| 100 | 300 | GaMMA 1D | 0.99 | 0.95 | 0.97 | 6.71 |
| 100 | 300 | Genie | 1.00 | 0.99 | 0.99 | 325.05 |
| 100 | 300 | PhaseLink | 0.88 | 0.81 | 0.84 | 4.15 |
| 100 | 300 | PyOcto 1D | 1.00 | 0.99 | 0.99 | 1.19 |
| 100 | 300 | REAL 0D | 1.00 | 0.94 | 0.97 | 453.73 |
| 500 | 30 | GaMMA 1D | 0.92 | 0.83 | 0.87 | 17.73 |
| 500 | 30 | Genie | 0.99 | 0.97 | 0.98 | 530.98 |
| 500 | 30 | PhaseLink | 0.85 | 0.60 | 0.71 | 5.11 |
| 500 | 30 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 6.74 |
| 500 | 30 | REAL 0D | 1.00 | 0.90 | 0.95 | 306.16 |
| 500 | 100 | GaMMA 1D | 0.90 | 0.80 | 0.85 | 25.74 |
| 500 | 100 | Genie | 1.00 | 0.99 | 0.99 | 548.01 |
| 500 | 100 | PhaseLink | 0.84 | 0.67 | 0.75 | 5.99 |
| 500 | 100 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 12.29 |
| 500 | 100 | REAL 0D | 0.99 | 0.92 | 0.95 | 754.36 |
| 500 | 300 | GaMMA 1D | 0.56 | 0.51 | 0.54 | 145.93 |
| 500 | 300 | Genie | 0.97 | 0.99 | 0.98 | 593.40 |
| 500 | 300 | PhaseLink | 0.15 | 0.25 | 0.19 | 11.36 |
| 500 | 300 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 33.46 |
| 500 | 300 | REAL 0D | 0.98 | 0.91 | 0.95 | 2032.63 |
| 2000 | 30 | GaMMA 1D | 0.68 | 0.48 | 0.57 | 210.15 |
| 2000 | 30 | Genie | 0.98 | 0.92 | 0.95 | 1211.63 |
| 2000 | 30 | PhaseLink | 0.54 | 0.19 | 0.28 | 14.08 |
| 2000 | 30 | PyOcto 1D | 1.00 | 0.91 | 0.95 | 96.97 |
| 2000 | 30 | REAL 0D | 0.99 | 0.73 | 0.84 | 1300.54 |
| 2000 | 100 | GaMMA 1D | 0.32 | 0.35 | 0.33 | 3598.16 |
| 2000 | 100 | Genie | 0.95 | 0.89 | 0.92 | 1256.66 |
| 2000 | 100 | PhaseLink | 0.07 | 0.01 | 0.02 | 16.87 |
| 2000 | 100 | PyOcto 1D | 0.99 | 0.92 | 0.95 | 131.74 |
| 2000 | 100 | REAL 0D | 0.95 | 0.70 | 0.81 | 2848.73 |
| 2000 | 300 | Genie | 0.80 | 0.78 | 0.79 | 1474.35 |
| 2000 | 300 | PhaseLink | 0.00 | 0.00 | 0.00 | 25.66 |
| 2000 | 300 | PyOcto 1D | 0.76 | 0.78 | 0.77 | 1790.04 |
| 2000 | 300 | REAL 0D | 0.54 | 0.54 | 0.54 | 6706.46 |

**Table S10.** Crustal Scenario: event-level evaluation of seismic phase associators

| Events | Noise | Associator | Event_Precision | Event_Recall | Event_F1_Score | Runtime (s) |
|---|---|---|---|---|---|---|
| 100 | 30 | GaMMA 1D | 1.00 | 1.00 | 1.00 | 9.89 |
| 100 | 30 | Genie | 1.00 | 1.00 | 1.00 | 319.66 |
| 100 | 30 | PhaseLink | 0.93 | 0.89 | 0.91 | 5.82 |
| 100 | 30 | PyOcto 1D | 1.00 | 1.00 | 1.00 | 2.71 |
| 100 | 30 | REAL 0D | 1.00 | 0.98 | 0.99 | 9.46 |
| 100 | 100 | GaMMA 1D | 1.00 | 0.99 | 0.99 | 13.22 |
| 100 | 100 | Genie | 0.99 | 0.99 | 0.99 | 333.67 |
| 100 | 100 | PhaseLink | 0.99 | 0.96 | 0.97 | 5.10 |
| 100 | 100 | PyOcto 1D | 1.00 | 0.98 | 0.99 | 3.42 |
| 100 | 100 | REAL 0D | 1.00 | 0.98 | 0.99 | 31.51 |
| 100 | 300 | GaMMA 1D | 0.99 | 0.99 | 0.99 | 18.56 |
| 100 | 300 | Genie | 0.99 | 0.99 | 0.99 | 345.50 |
| 100 | 300 | PhaseLink | 1.00 | 0.98 | 0.99 | 6.45 |
| 100 | 300 | PyOcto 1D | 1.00 | 0.99 | 0.99 | 5.84 |
| 100 | 300 | REAL 0D | 1.00 | 0.99 | 0.99 | 75.32 |
| 500 | 30 | GaMMA 1D | 0.99 | 0.99 | 0.99 | 46.43 |
| 500 | 30 | Genie | 0.99 | 0.98 | 0.99 | 582.32 |
| 500 | 30 | PhaseLink | 0.98 | 0.88 | 0.92 | 10.78 |
| 500 | 30 | PyOcto 1D | 1.00 | 0.97 | 0.98 | 19.02 |
| 500 | 30 | REAL 0D | 1.00 | 0.93 | 0.96 | 48.30 |
| 500 | 100 | GaMMA 1D | 1.00 | 1.00 | 1.00 | 67.81 |
| 500 | 100 | Genie | 1.00 | 1.00 | 1.00 | 663.37 |
| 500 | 100 | PhaseLink | 0.99 | 0.88 | 0.93 | 13.31 |
| 500 | 100 | PyOcto 1D | 1.00 | 0.98 | 0.99 | 10.95 |
| 500 | 100 | REAL 0D | 0.99 | 0.93 | 0.96 | 145.26 |
| 500 | 300 | GaMMA 1D | 0.79 | 0.99 | 0.88 | 150.07 |
| 500 | 300 | Genie | 0.99 | 0.98 | 0.99 | 635.51 |
| 500 | 300 | PhaseLink | 0.96 | 0.88 | 0.92 | 20.48 |
| 500 | 300 | PyOcto 1D | 1.00 | 0.98 | 0.99 | 21.26 |
| 500 | 300 | REAL 0D | 0.98 | 0.93 | 0.96 | 419.16 |
| 2000 | 30 | GaMMA 1D | 0.98 | 0.99 | 0.98 | 312.27 |
| 2000 | 30 | Genie | 0.98 | 0.94 | 0.96 | 1445.19 |
| 2000 | 30 | PhaseLink | 0.93 | 0.61 | 0.74 | 48.07 |
| 2000 | 30 | PyOcto 1D | 0.99 | 0.93 | 0.96 | 51.07 |
| 2000 | 30 | REAL 0D | 0.99 | 0.78 | 0.87 | 239.90 |
| 2000 | 100 | GaMMA 1D | 0.57 | 0.96 | 0.71 | 900.20 |
| 2000 | 100 | Genie | 0.98 | 0.94 | 0.96 | 1463.43 |
| 2000 | 100 | PhaseLink | 0.84 | 0.56 | 0.67 | 61.19 |
| 2000 | 100 | PyOcto 1D | 0.99 | 0.92 | 0.95 | 136.43 |
| 2000 | 100 | REAL 0D | 0.96 | 0.79 | 0.87 | 551.88 |
| 2000 | 300 | Genie | 0.95 | 0.86 | 0.90 | 1769.01 |
| 2000 | 300 | PhaseLink | 0.10 | 0.14 | 0.12 | 143.76 |
| 2000 | 300 | PyOcto 1D | 0.73 | 0.91 | 0.81 | 1170.82 |
| 2000 | 300 | REAL 0D | 0.91 | 0.73 | 0.81 | 1985.60 |

**Table S11.** Subduction zone scenario: evaluation of seismic phase associators at pick level across different event and noise levels. GT: Ground Truth Picks, Pred: Predicted Picks, CA: Commonly Associated Picks, Missed: Missed Picks, FP: False Picks, WAP: Wrongly Associated Picks.

| Associator | Ev. | Noise | GT | Pred | CA | Missed | FP | WAP | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GaMMA 1D | 100 | 30 | 28.36 | 26.17 | 26.11 | 2.24 | 0.01 | 0.04 | 1.00 | 0.92 | 0.96 |
| Genie | 100 | 30 | 27.82 | 27.76 | 27.65 | 0.17 | 0.05 | 0.06 | 1.00 | 0.99 | 1.00 |
| PhaseLink | 100 | 30 | 28.44 | 24.56 | 23.68 | 4.76 | 0.22 | 0.66 | 0.97 | 0.84 | 0.89 |
| PyOcto 1D | 100 | 30 | 28.16 | 27.27 | 27.24 | 0.93 | 0.01 | 0.02 | 1.00 | 0.97 | 0.98 |
| REAL 0D | 100 | 30 | 27.78 | 26.78 | 26.54 | 1.23 | 0.07 | 0.16 | 0.99 | 0.96 | 0.98 |
| GaMMA 1D | 100 | 100 | 30.14 | 26.77 | 26.75 | 3.39 | 0.02 | 0.00 | 1.00 | 0.90 | 0.94 |
| Genie | 100 | 100 | 29.01 | 28.91 | 28.79 | 0.22 | 0.10 | 0.02 | 1.00 | 0.99 | 0.99 |
| PhaseLink | 100 | 100 | 29.79 | 25.75 | 24.58 | 5.21 | 0.67 | 0.51 | 0.96 | 0.83 | 0.89 |
| PyOcto 1D | 100 | 100 | 29.12 | 28.25 | 28.21 | 0.91 | 0.02 | 0.02 | 1.00 | 0.97 | 0.98 |
| REAL 0D | 100 | 100 | 29.03 | 28.09 | 27.86 | 1.17 | 0.20 | 0.02 | 0.99 | 0.97 | 0.98 |
| GaMMA 1D | 100 | 300 | 29.75 | 27.53 | 27.15 | 2.60 | 0.16 | 0.22 | 0.99 | 0.92 | 0.95 |
| Genie | 100 | 300 | 29.39 | 29.28 | 28.99 | 0.40 | 0.22 | 0.07 | 0.99 | 0.99 | 0.99 |
| PhaseLink | 100 | 300 | 30.17 | 27.42 | 24.23 | 5.94 | 2.57 | 0.62 | 0.89 | 0.81 | 0.84 |
| PyOcto 1D | 100 | 300 | 29.66 | 28.92 | 28.78 | 0.88 | 0.07 | 0.07 | 1.00 | 0.97 | 0.98 |
| REAL 0D | 100 | 300 | 29.02 | 28.16 | 27.40 | 1.62 | 0.46 | 0.30 | 0.98 | 0.95 | 0.96 |
| GaMMA 1D | 500 | 30 | 29.31 | 26.94 | 26.76 | 2.55 | 0.06 | 0.12 | 0.99 | 0.92 | 0.95 |
| Genie | 500 | 30 | 28.45 | 28.06 | 27.70 | 0.75 | 0.17 | 0.19 | 0.99 | 0.97 | 0.98 |
| PhaseLink | 500 | 30 | 29.35 | 28.40 | 24.07 | 5.28 | 1.02 | 3.30 | 0.88 | 0.83 | 0.84 |
| PyOcto 1D | 500 | 30 | 28.64 | 27.62 | 27.43 | 1.20 | 0.07 | 0.11 | 0.99 | 0.96 | 0.97 |
| REAL 0D | 500 | 30 | 28.46 | 27.38 | 26.60 | 1.86 | 0.23 | 0.55 | 0.97 | 0.94 | 0.96 |
| GaMMA 1D | 500 | 100 | 28.86 | 27.12 | 26.77 | 2.09 | 0.23 | 0.11 | 0.99 | 0.93 | 0.95 |
| Genie | 500 | 100 | 27.85 | 27.33 | 26.81 | 1.04 | 0.39 | 0.12 | 0.98 | 0.96 | 0.97 |
| PhaseLink | 500 | 100 | 28.89 | 28.81 | 23.08 | 5.81 | 3.96 | 1.76 | 0.81 | 0.80 | 0.80 |
| PyOcto 1D | 500 | 100 | 28.04 | 27.02 | 26.75 | 1.29 | 0.18 | 0.08 | 0.99 | 0.96 | 0.97 |
| REAL 0D | 500 | 100 | 28.08 | 27.17 | 25.97 | 2.11 | 0.78 | 0.43 | 0.96 | 0.93 | 0.94 |
| GaMMA 1D | 500 | 300 | 31.60 | 30.83 | 29.75 | 1.85 | 0.87 | 0.20 | 0.96 | 0.94 | 0.95 |
| Genie | 500 | 300 | 28.26 | 28.10 | 26.75 | 1.51 | 1.24 | 0.11 | 0.95 | 0.95 | 0.95 |
| PhaseLink | 500 | 300 | 30.51 | 36.22 | 21.79 | 8.72 | 14.04 | 0.40 | 0.61 | 0.72 | 0.66 |
| PyOcto 1D | 500 | 300 | 28.45 | 27.59 | 26.92 | 1.53 | 0.58 | 0.09 | 0.97 | 0.95 | 0.96 |
| REAL 0D | 500 | 300 | 28.43 | 28.50 | 25.49 | 2.94 | 2.57 | 0.45 | 0.89 | 0.90 | 0.89 |
| GaMMA 1D | 2000 | 30 | 30.92 | 29.94 | 28.88 | 2.04 | 0.33 | 0.73 | 0.96 | 0.94 | 0.95 |
| Genie | 2000 | 30 | 28.38 | 26.79 | 25.76 | 2.62 | 0.44 | 0.59 | 0.96 | 0.91 | 0.93 |
| PhaseLink | 2000 | 30 | 29.17 | 32.04 | 22.48 | 6.69 | 4.95 | 4.61 | 0.72 | 0.78 | 0.74 |
| PyOcto 1D | 2000 | 30 | 28.52 | 27.13 | 26.56 | 1.96 | 0.20 | 0.37 | 0.98 | 0.93 | 0.95 |
| REAL 0D | 2000 | 30 | 29.10 | 28.35 | 25.39 | 3.72 | 0.93 | 2.04 | 0.90 | 0.88 | 0.88 |
| GaMMA 1D | 2000 | 100 | 32.25 | 32.54 | 30.35 | 1.90 | 1.20 | 1.00 | 0.93 | 0.94 | 0.93 |
| Genie | 2000 | 100 | 28.49 | 26.79 | 24.97 | 3.52 | 1.30 | 0.53 | 0.93 | 0.88 | 0.90 |
| PhaseLink | 2000 | 100 | 29.78 | 35.85 | 20.41 | 9.37 | 14.00 | 1.44 | 0.58 | 0.69 | 0.63 |
| PyOcto 1D | 2000 | 100 | 28.32 | 27.01 | 25.87 | 2.44 | 0.73 | 0.40 | 0.96 | 0.91 | 0.93 |
| REAL 0D | 2000 | 100 | 29.26 | 29.60 | 24.40 | 4.85 | 3.22 | 1.98 | 0.82 | 0.84 | 0.83 |
| Genie | 2000 | 300 | 28.78 | 25.07 | 21.99 | 6.79 | 2.68 | 0.41 | 0.88 | 0.76 | 0.81 |
| PhaseLink | 2000 | 300 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | nan | nan | nan |
| PyOcto 1D | 2000 | 300 | 28.84 | 26.59 | 23.82 | 5.02 | 2.32 | 0.45 | 0.89 | 0.82 | 0.84 |
| REAL 0D | 2000 | 300 | 30.46 | 33.26 | 22.93 | 7.53 | 8.58 | 1.75 | 0.69 | 0.76 | 0.72 |

**Table S12.** Crustal scenario: evaluation of seismic phase associators at the pick level across different event densities and noise conditions.

| Associator | Ev. | Noise | GT | Pred | CA | Missed | FP | WAP | Precision | Recall | F1 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| GaMMA 1D | 100 | 30 | 81.78 | 67.74 | 67.69 | 14.09 | 0.01 | 0.04 | 1.00 | 0.83 | 0.91 |
| Genie | 100 | 30 | 81.78 | 81.62 | 81.51 | 0.27 | 0.03 | 0.08 | 1.00 | 1.00 | 1.00 |
| PhaseLink | 100 | 30 | 80.97 | 70.52 | 69.70 | 11.27 | 0.13 | 0.69 | 0.99 | 0.86 | 0.92 |
| PyOcto 1D | 100 | 30 | 81.78 | 81.69 | 81.60 | 0.18 | 0.01 | 0.08 | 1.00 | 1.00 | 1.00 |
| REAL 0D | 100 | 30 | 81.80 | 81.93 | 81.54 | 0.26 | 0.11 | 0.28 | 1.00 | 1.00 | 1.00 |
| GaMMA 1D | 100 | 100 | 83.47 | 68.96 | 68.87 | 14.61 | 0.04 | 0.05 | 1.00 | 0.83 | 0.91 |
| Genie | 100 | 100 | 83.15 | 82.94 | 82.67 | 0.48 | 0.11 | 0.16 | 1.00 | 0.99 | 0.99 |
| PhaseLink | 100 | 100 | 83.46 | 77.60 | 75.83 | 7.62 | 0.56 | 1.21 | 0.98 | 0.91 | 0.94 |
| PyOcto 1D | 100 | 100 | 83.33 | 83.30 | 83.09 | 0.23 | 0.10 | 0.10 | 1.00 | 1.00 | 1.00 |
| REAL 0D | 100 | 100 | 83.33 | 83.46 | 82.56 | 0.77 | 0.45 | 0.45 | 0.99 | 0.99 | 0.99 |
| GaMMA 1D | 100 | 300 | 81.78 | 68.81 | 68.66 | 13.12 | 0.10 | 0.05 | 1.00 | 0.84 | 0.91 |
| Genie | 100 | 300 | 81.25 | 81.28 | 80.81 | 0.44 | 0.39 | 0.08 | 0.99 | 0.99 | 0.99 |
| PhaseLink | 100 | 300 | 81.20 | 79.08 | 77.21 | 3.99 | 1.56 | 0.31 | 0.98 | 0.95 | 0.96 |
| PyOcto 1D | 100 | 300 | 81.27 | 81.34 | 80.86 | 0.41 | 0.37 | 0.11 | 0.99 | 0.99 | 0.99 |
| REAL 0D | 100 | 300 | 81.25 | 81.88 | 80.14 | 1.11 | 1.40 | 0.33 | 0.98 | 0.99 | 0.98 |
| GaMMA 1D | 500 | 30 | 81.38 | 68.95 | 68.79 | 12.58 | 0.07 | 0.09 | 1.00 | 0.85 | 0.92 |
| Genie | 500 | 30 | 81.18 | 80.66 | 80.26 | 0.92 | 0.16 | 0.24 | 0.99 | 0.99 | 0.99 |
| PhaseLink | 500 | 30 | 81.15 | 79.85 | 76.29 | 4.86 | 0.91 | 2.65 | 0.96 | 0.94 | 0.95 |
| PyOcto 1D | 500 | 30 | 81.26 | 80.81 | 80.35 | 0.91 | 0.14 | 0.32 | 0.99 | 0.99 | 0.99 |
| REAL 0D | 500 | 30 | 81.59 | 82.05 | 80.01 | 1.59 | 0.68 | 1.37 | 0.98 | 0.98 | 0.98 |
| GaMMA 1D | 500 | 100 | 81.90 | 70.19 | 69.90 | 12.00 | 0.24 | 0.05 | 1.00 | 0.86 | 0.92 |
| Genie | 500 | 100 | 81.70 | 81.33 | 80.58 | 1.12 | 0.60 | 0.14 | 0.99 | 0.99 | 0.99 |
| PhaseLink | 500 | 100 | 82.10 | 82.20 | 75.35 | 6.75 | 3.16 | 3.69 | 0.93 | 0.92 | 0.92 |
| PyOcto 1D | 500 | 100 | 81.72 | 81.71 | 80.99 | 0.73 | 0.53 | 0.19 | 0.99 | 0.99 | 0.99 |
| REAL 0D | 500 | 100 | 82.17 | 83.24 | 79.45 | 2.72 | 2.43 | 1.35 | 0.95 | 0.97 | 0.96 |
| GaMMA 1D | 500 | 300 | 82.03 | 72.53 | 71.76 | 10.27 | 0.70 | 0.07 | 0.99 | 0.88 | 0.93 |
| Genie | 500 | 300 | 81.83 | 82.00 | 80.06 | 1.78 | 1.78 | 0.17 | 0.98 | 0.98 | 0.98 |
| PhaseLink | 500 | 300 | 81.89 | 84.70 | 70.74 | 11.15 | 11.50 | 2.45 | 0.84 | 0.87 | 0.85 |
| PyOcto 1D | 500 | 300 | 81.72 | 81.93 | 80.18 | 1.55 | 1.48 | 0.27 | 0.98 | 0.98 | 0.98 |
| REAL 0D | 500 | 300 | 82.02 | 84.81 | 76.67 | 5.35 | 6.99 | 1.16 | 0.90 | 0.94 | 0.92 |
| GaMMA 1D | 2000 | 30 | 81.91 | 71.86 | 71.21 | 10.70 | 0.33 | 0.32 | 0.99 | 0.87 | 0.93 |
| Genie | 2000 | 30 | 82.07 | 80.11 | 78.61 | 3.46 | 0.70 | 0.80 | 0.98 | 0.96 | 0.97 |
| PhaseLink | 2000 | 30 | 82.66 | 87.21 | 73.25 | 9.41 | 4.51 | 9.45 | 0.86 | 0.89 | 0.87 |
| PyOcto 1D | 2000 | 30 | 82.10 | 81.12 | 79.58 | 2.52 | 0.62 | 0.92 | 0.98 | 0.97 | 0.97 |
| REAL 0D | 2000 | 30 | 83.56 | 85.44 | 77.77 | 5.79 | 2.68 | 4.99 | 0.91 | 0.93 | 0.92 |
| GaMMA 1D | 2000 | 100 | 82.41 | 74.76 | 73.41 | 9.01 | 1.06 | 0.29 | 0.98 | 0.89 | 0.93 |
| Genie | 2000 | 100 | 82.40 | 80.91 | 77.97 | 4.44 | 2.14 | 0.80 | 0.96 | 0.95 | 0.95 |
| PhaseLink | 2000 | 100 | 82.83 | 90.16 | 67.90 | 14.93 | 15.49 | 6.78 | 0.77 | 0.82 | 0.79 |
| PyOcto 1D | 2000 | 100 | 82.30 | 81.88 | 78.97 | 3.33 | 2.00 | 0.91 | 0.96 | 0.96 | 0.96 |
| REAL 0D | 2000 | 100 | 83.76 | 87.35 | 74.36 | 9.40 | 8.66 | 4.33 | 0.85 | 0.89 | 0.87 |
| Genie | 2000 | 300 | 82.05 | 79.62 | 73.52 | 8.53 | 5.37 | 0.74 | 0.92 | 0.89 | 0.91 |
| PhaseLink | 2000 | 300 | 84.95 | 98.24 | 56.00 | 28.95 | 41.26 | 0.98 | 0.57 | 0.66 | 0.61 |
| PyOcto 1D | 2000 | 300 | 81.58 | 81.63 | 74.69 | 6.90 | 5.99 | 0.95 | 0.91 | 0.91 | 0.91 |
| REAL 0D | 2000 | 300 | 83.99 | 91.09 | 64.99 | 18.99 | 22.72 | 3.37 | 0.71 | 0.77 | 0.74 |