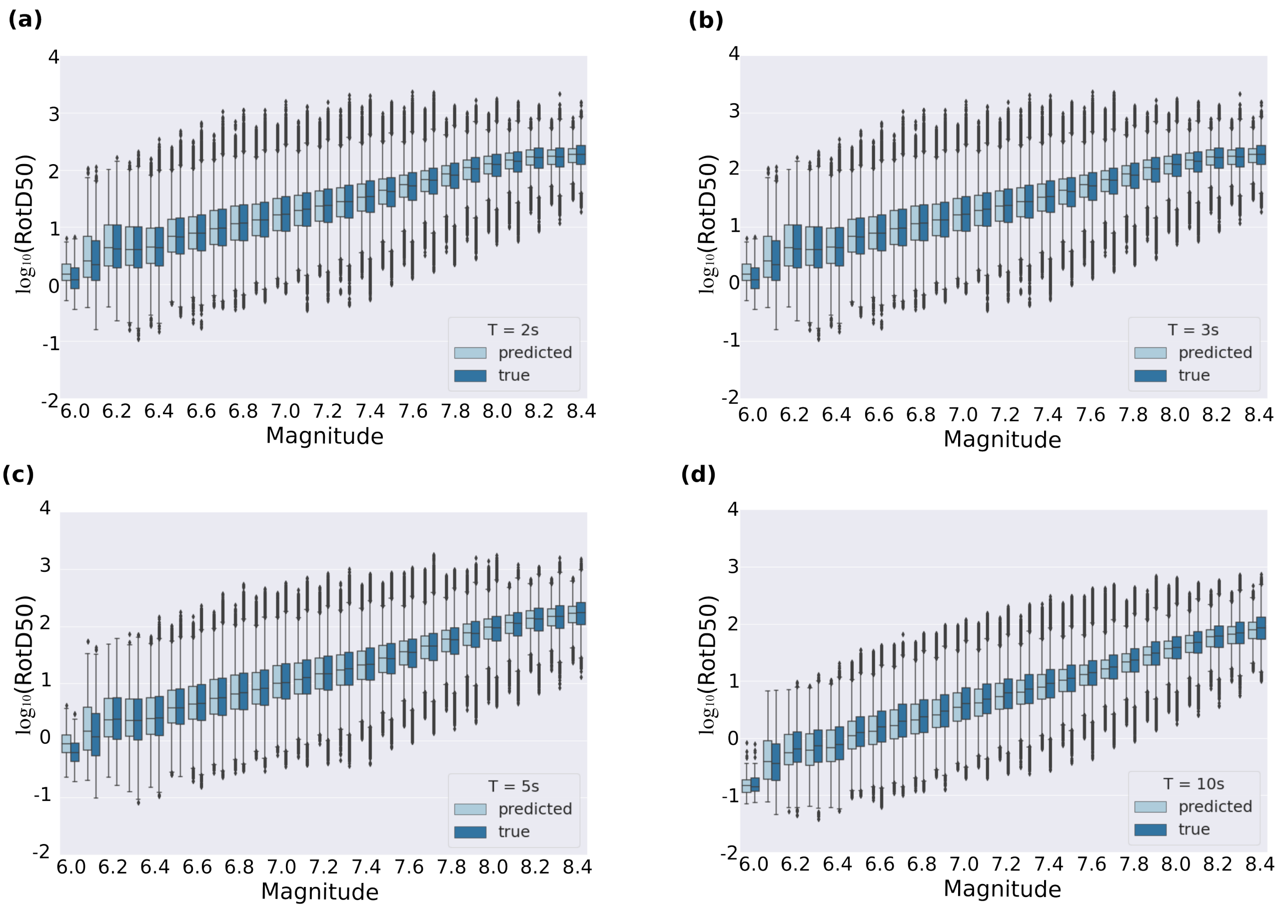
A machine learning estimator trained on synthetic data for real-time earthquake ground-shaking predictions in Southern California

[1,*]Marisol Monterrubio-Velasco [2]Scott Callaghan [1]David Modesto [1]Jose Carlos Carrasco [1]Rosa M. Badia [3]Pablo Pallares [1]Fernando Vazquez [3]Enrique Quintana [4]Marta Pienkowska [1]Josep de la Puente

Supplementary Material



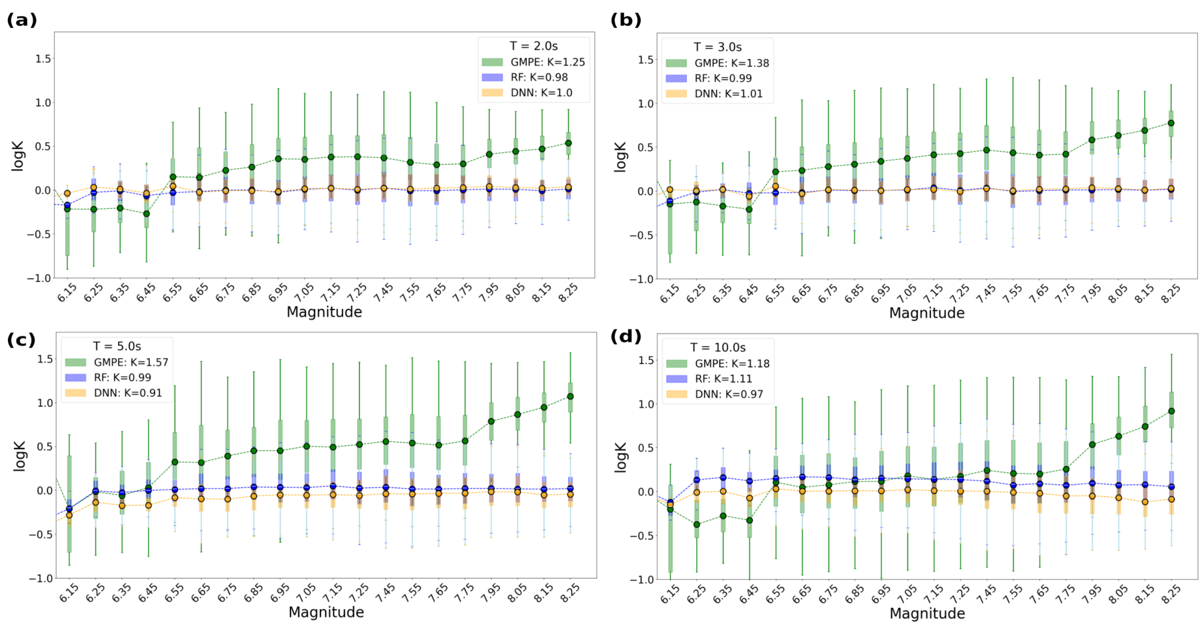
Supplementary Figure 1. Statistical distribution of ML inferences on the synthetic validation subset. Box-plots comparison for the logarithm of RotD50 coming directly from CSS-15.4 observations (dark blue) and the RF inferences (light blue) per magnitude $\log_{10}$ for the four analyzed periods. The median value of the logarithm of RotD50 is shown as a line inside each boxplot. The average value of the RMSE of the logarithm of RotD50 median for all magnitudes is 0.027 for $T = 2$s, 0.026 for $T = 3$s, 0.039 for $T = 5$s, and 0.052 for $T = 10$s, indicating a large similarity between the predicted and the expected values. Boxplot bars depict first and third quartile

	RMSE	MSE	R^2	MAE	MAPE [%]	PCC
RF	0.20 ± 0.04	0.04 ± 0.002	0.85 ± 0.007	0.15 ± 0.008	13.75 ± 3.7	0.93 ± 0.004
DNN	0.17 ± 0.04	0.03 ± 0.003	0.87 ± 0.005	0.14 ± 0.01	15 ± 7.3	0.93 ± 0.004

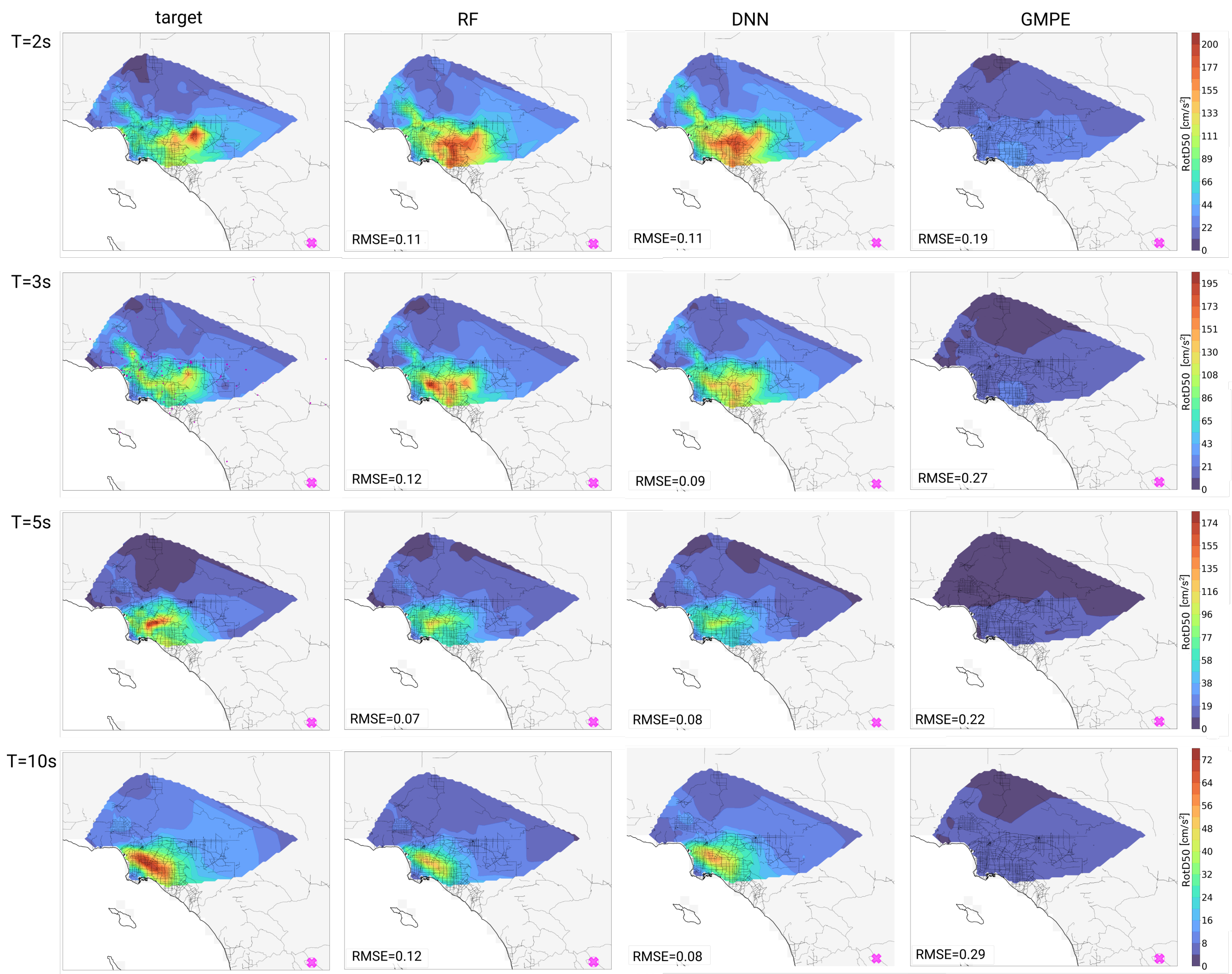
Supplementary Table 1. The average for all four periods of the six score metrics (Root Mean Square Error, RMSE; Mean Square Error, MSE; Determination coefficient, R^2; Mean Absolute Error, MAE; Mean Absolute Percentage Error, MAPE; Pearson correlation coefficient, PCC) obtained for the inferences of the 3.8M event realisations in the validation subset shown in Fig. 2 in the main text. The values for the metrics are calculated between the predicted and synthetic log(RotD50) value on one recording site for all earthquake scenarios, and then averaged. The values for all metrics are consistent for both algorithms, indicating a high predictive capacity of all models. In particular, the average of R^2 close to 1 indicates the very high likelihood of correct predictions for unseen samples, while the MAPE indicates that the average absolute percentage difference between the predictions and the actual values is below 15%.

Periods	GMM	RF	DNN
2	0.14	0.08	0.08
3	0.16	0.09	0.09
5	0.19	0.10	0.10
10	0.19	0.10	0.10

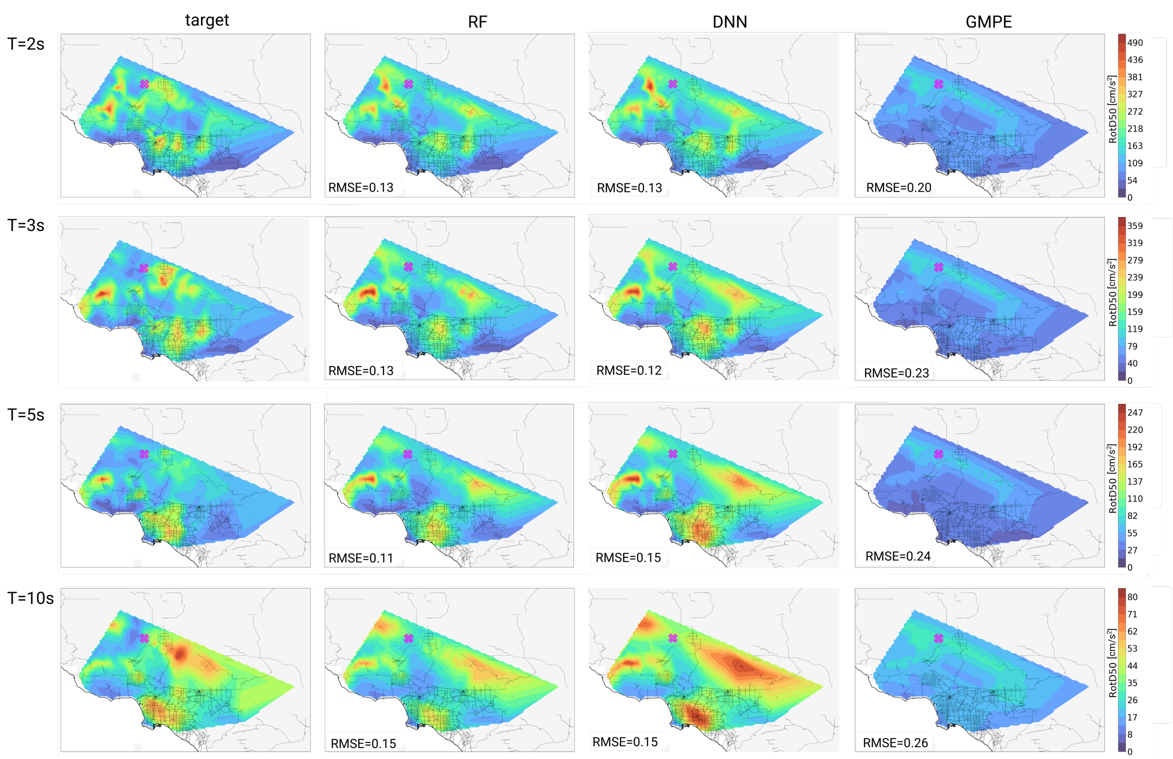
Supplementary Table 2. The mean of the median RMSE $[cm/s^2]$ values computed for each 0.1 width magnitude bin in Fig. 3 in the main text. The RMSE is calculated between the predicted and synthetic RotD50 for one earthquake scenario on all recording sites, and then averaged.



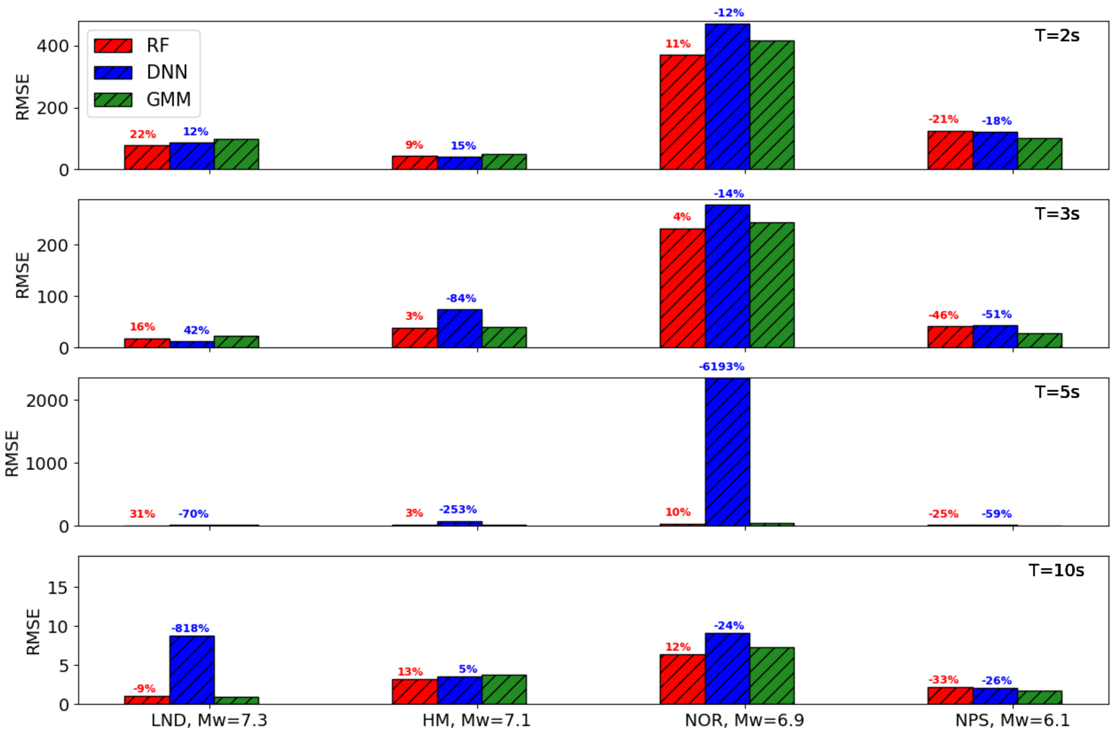
Supplementary Figure 2. log(K) distribution for the ASK-14 GMM (green), RF (blue), and DNN (orange) estimations against the synthetic results shown per magnitude bin for (a) $T = 2$s, (b) $T = 3$s, (c) $T = 5$s, (d) $T = 10$s. The median value is marked with a circle for each boxplot, while the mean of the median K value for all magnitudes is added in the legend. Boxplot bars depict first and third quartile of the distribution of values. The RF and DNN predictions do not show a particular bias, whereas ASK-14 tents to underpredict RotD50 for larger events ($> M_W 6.5$) and overpredict for smaller events ($< M_W 6.5$). Refer to the main text for more information on Aida's number K[21]



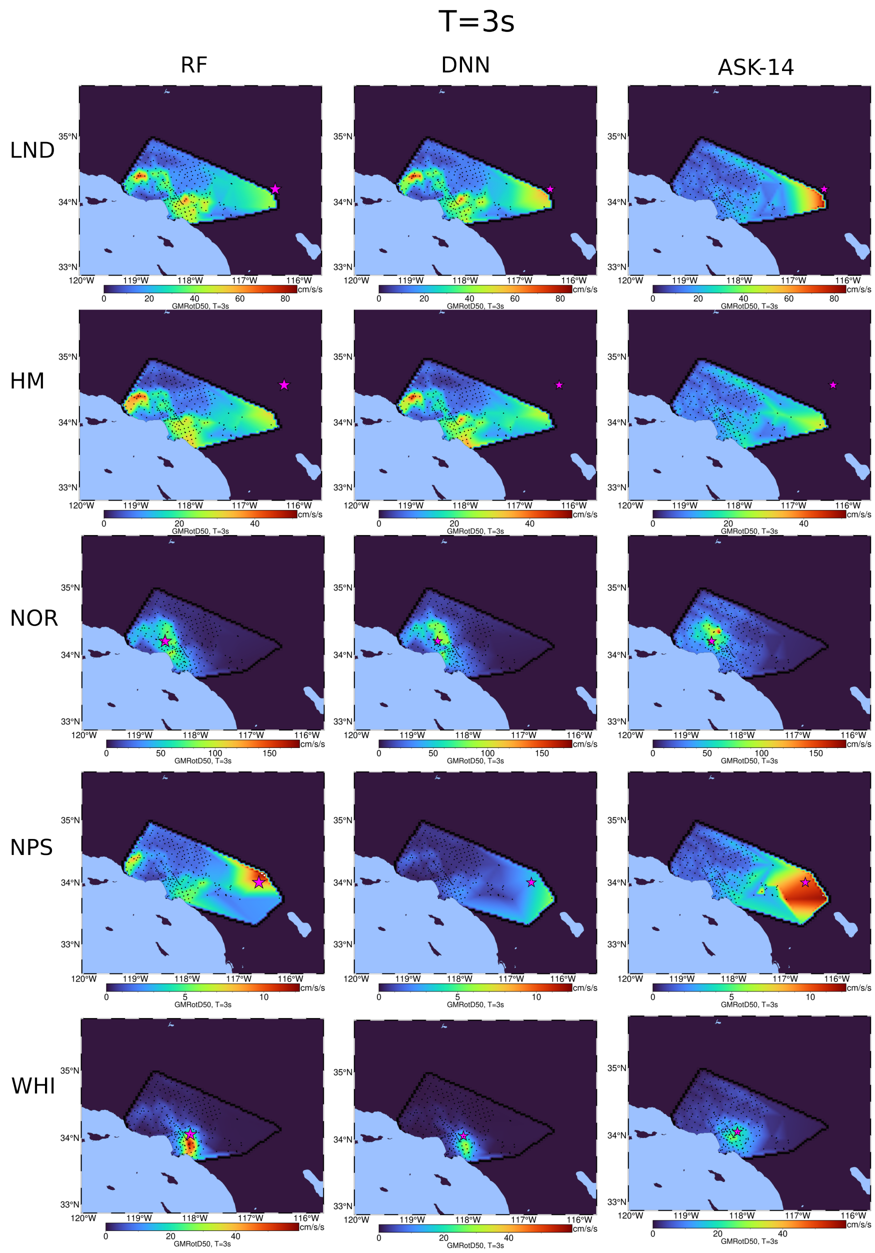
Supplementary Figure 3. RotD50 predictions on a validation event of magnitude 7.45. The reference RotD50 map (first column), as well as RF (second column), DNN (third column), and ASK-14 (fourth column) predictions given in cm/s^2 for a synthetic validation event: an $M_\mathrm{W} = 7.45$ earthquake located at 32.9979°N, 116.3737°W, and 30.73km deep pink cross). Rows show the increasing period from $T = 2$s, 3s, 5s, and 10s. The score metrics (RMSE) for each prediction are annotated in each subfigure. ASK-14 consistently underestimates RotD50, while the spatial distribution of RotD50 in MLESmap models reflects that of the reference maps.



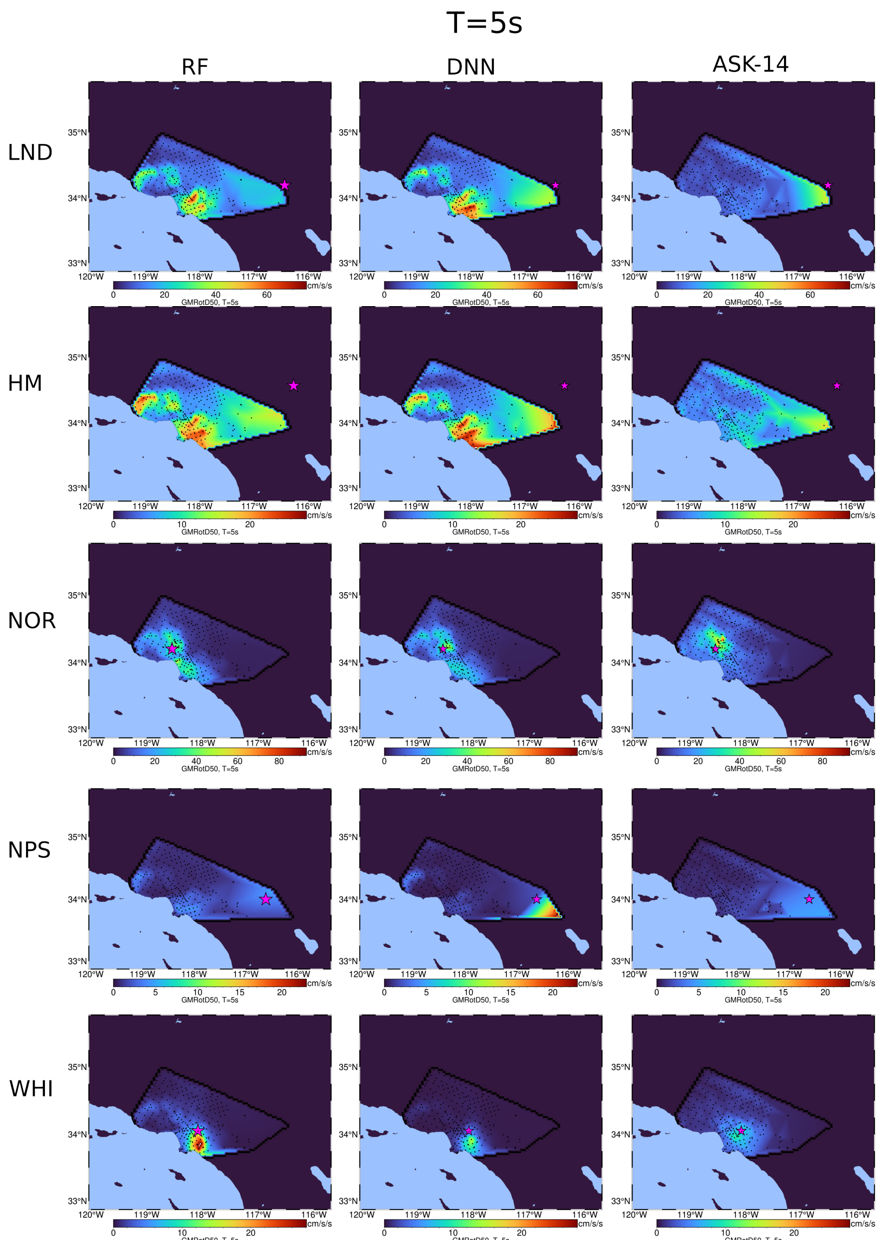
Supplementary Figure 4. RotD50 predictions on a validation event of magnitude 8.05. The reference RotD50 map (first column), as well as RF (second column), DNN (third column), and ASK-14 (fourth column) predictions given in cm/s^2 for a synthetic validation event: $M_\mathrm{W} = 8.05$ earthquake located at 34.6555°N, 118.3928°W, and 16.1 km deep (pink cross). Rows, from top to bottom, show the increasing period from $T = 2$s, 3s, 5s, and 10s. The score metrics (RMSE) for each prediction are annotated in each subfigure. ASK-14 consistently underestimates RotD50, while the spatial distribution of RotD50 in MLESmap models reflects that of the reference maps.



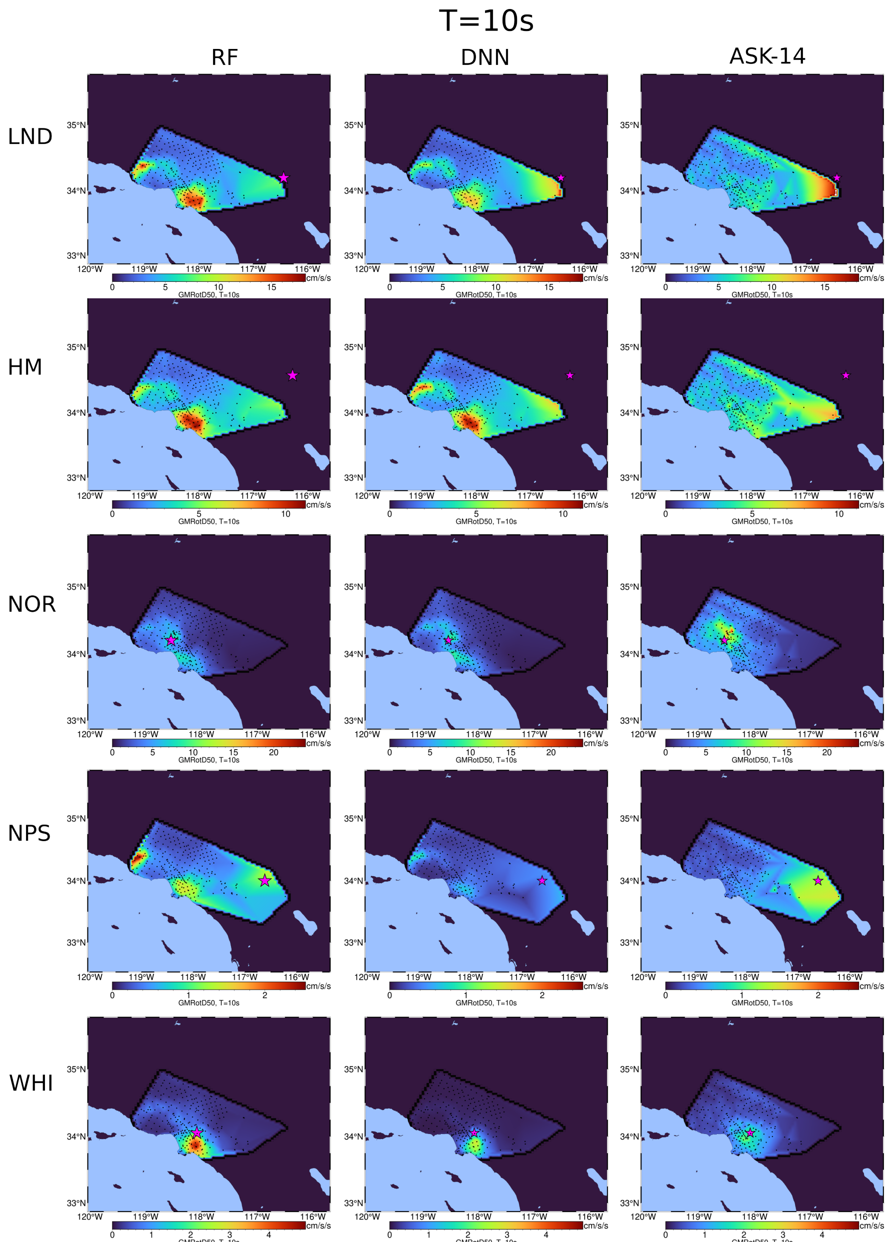
Supplementary Figure 5. RotD50 prediction of real earthquakes results for the 'outside' stations. RMSE for RF (red), DNN (blue), and ASK-14 (green) for the 'outside' stations shown as blue circles in Fig. 5 in the main text. From top to bottom, each row indicates the RMSE for $T = 2$s, 3s, 5s, and 10s, respectively. The value annotated above each bar indicates the improvement/deterioration of the MLESmap predictions concerning the ASK-14 predictions. For the 'outside' stations shown here the extrapolations performed by the ML models make MLESmap underperform relative to ASK-14. The MLESmap predictions improve on the ASK-14 predictions provided that no extrapolations are necessary and the station locations fall within the area covered by the synthetic training sites (the 'outside' stations in this figure v.s. the 'inside' stations in Fig. 6 in the main text). Note also that the figure suggests that the algorithms, given their different mathematical foundations, weigh the relevant features differently. The DNN topology appears to manifest a large influence of the spatial predictors, while the RF is less influenced by those parameters. Likely the non-spatial parameters play a larger role in the RF models (such as, for example, the event magnitude) and the station location falling outside of the training range has less of an effect than for the DNN predictions.



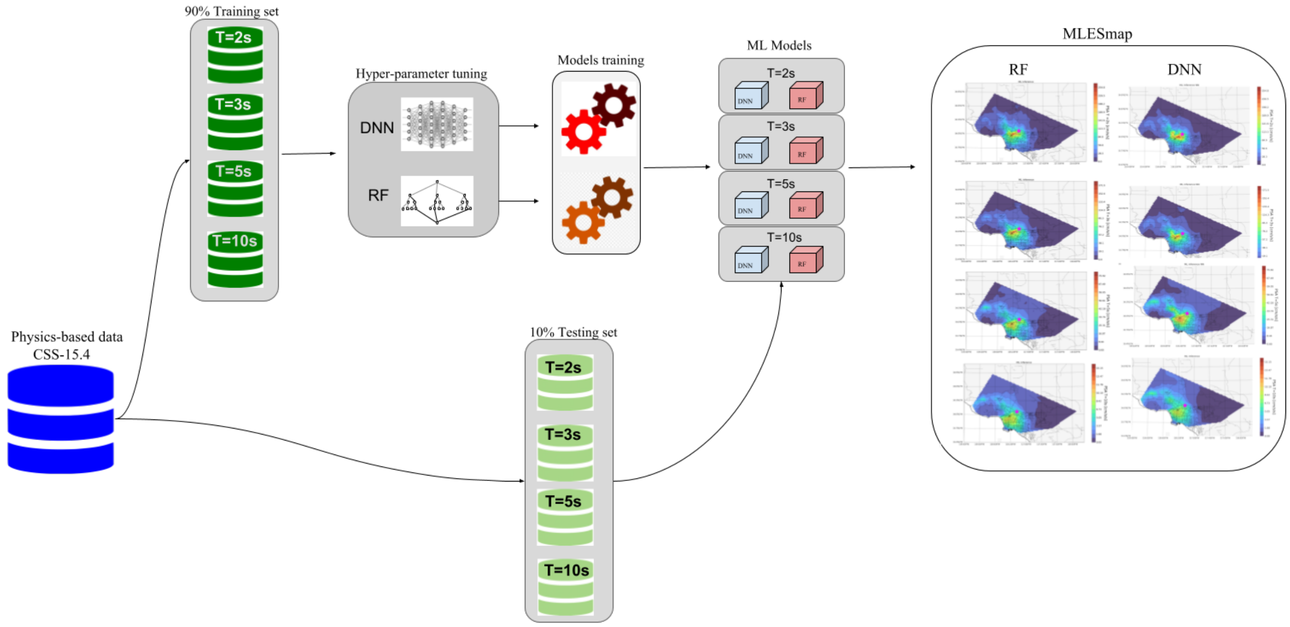
Supplementary Figure 6. RotD50 predictions for $T = 3$s for all real events in the training domain. The RotD50 maps of RF (first column), DNN (second column), and ASK-14 (third column) predictions given in cm/s^2. Each row corresponds to one of the real events (magenta star), namely Landers (LND), Hector Mine (HM), Northridge (NOR), North Palm Springs (NPS) and Whittier (WHI). See Fig. 7, and Supplementary Figure 7 and 8 for the corresponding plots of the predictions for periods of 2s, 5s and 10s.



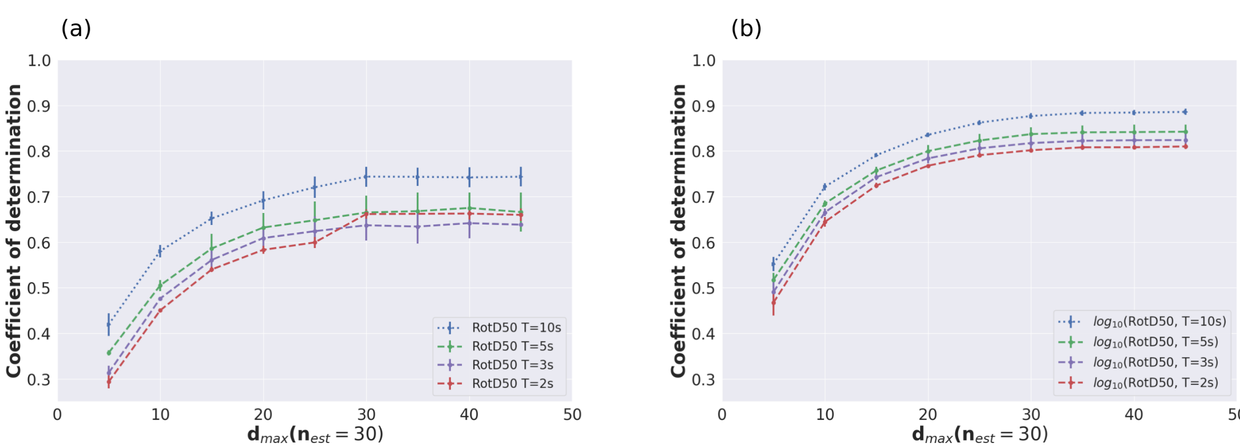
Supplementary Figure 7. RotD50 predictions for $T = 5$s for all real events in the training domain. The RotD50 maps of RF (first column), DNN (second column), and ASK-14 (third column) predictions given in cm/s^2. Each row corresponds to one of the real events (magenta star), namely Landers (LND), Hector Mine (HM), Northridge (NOR), North Palm Springs (NPS) and Whittier (WHI). See Fig. 7, and Supplementary FigureFig. 6 and 8 for the corresponding plots of the predictions for periods of 2s, 3s and 10s.



Supplementary Figure 8. RotD50 predictions for $T = 10$s for all real events in the training domain. The RotD50 maps of RF (first column), DNN (second column), and ASK-14 (third column) predictions given in cm/s^2. Each row corresponds to one of the real events (magenta star), namely Landers (LND), Hector Mine (HM), Northridge (NOR), North Palm Springs (NPS) and Whittier (WHI). See Fig. 7, and Supplementary FigureFig. 6 and 7 for the corresponding plots of the predictions for periods of 2s, 3s and 5s.



Supplementary Figure 9. Schematic representation of the MLESmap methodology. In the first step the physics-based simulations need to be provided via the CyberShake platform. Then the data is pre-processed considering four periods and divided into training (90%) and test (10%) subsets. The training set is used to fit the hyperparameters for both algorithms – (Deep Neural Network (DNN), and Random Forest (RF) – and afterwards to train the models for each period. Once the eight models are built, the test set is used for validation.



Supplementary Figure 10. The score of the coefficient of determination (R^2) for different Random Forest models by varying the values of d_{max} and taking constant $n_{est} = 30$ and $t_f =$ 'third'. **(a)** Shows the R^2 for the models trained by using the RotD50 target on linear scale, and **(b)** on a logarithmic scale.