

<https://doi.org/10.1038/s43247-024-01436-1>

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



Marisol Monterrubio-Velasco¹✉, Scott Callaghan², David Modesto¹, Jose Carlos Carrasco¹, Rosa M. Badia¹, Pablo Pallares³, Fernando Vázquez-Novoa¹, Enrique S. Quintana-Ort³, Marta Pienkowska⁴ & Josep de la Puente¹

After large-magnitude earthquakes, a crucial task for impact assessment is to rapidly and accurately estimate the ground shaking in the affected region. To satisfy real-time constraints, intensity measures are traditionally evaluated with empirical Ground Motion Models that can drastically limit the accuracy of the estimated values. As an alternative, here we present Machine Learning strategies trained on physics-based simulations that require similar evaluation times. We trained and validated the proposed Machine Learning-based Estimator for ground shaking maps with one of the largest existing datasets (<100M simulated seismograms) from CyberShake developed by the Southern California Earthquake Center covering the Los Angeles basin. For a well-tailored synthetic database, our predictions outperform empirical Ground Motion Models provided that the events considered are compatible with the training data. Using the proposed strategy we show significant error reductions not only for synthetic, but also for five real historical earthquakes, relative to empirical Ground Motion Models.

Large earthquakes are amongst the most destructive and unpredictable natural phenomena, yet our ability to rapidly and accurately estimate their impacts remains limited. Due to its complexity, numerical wave propagation tends to be too computationally expensive for disaster mitigation purposes, even when massive High-Performance Computing (HPC) resources are available. Simulations also are sensitive to model inputs—in particular for large events where shaking is especially hard to correlate with standard inputs such as topography or wave amplification at basins—but can provide high spatial resolution of ground motions when the underlying physical models are sufficiently accurate. Recent urgent HPC workflows^{1,2}, designed for fast physics-based earthquake simulations for emergency relief measures, using top-tier HPC facilities can provide sets of synthetic solutions that aim to capture the input variability within one hour. Although such computing times are extraordinary for rapid post-event analyses, they do not meet the requirements of decision-makers for near real-time societal alerts.

Empirical ground motion models (GMMs) are the traditional solution to fast estimations of intensity measures (IMs) that circumvents the use of physics-based approaches³. Empirical GMMs, proposed in a range of model functional forms, give mean and variability values for IMs (e.g., peak ground acceleration (PGA) or pseudo-spectral acceleration (PSA)) as functions of seismic observations (e.g., earthquake magnitude or site-to-source distance⁴). An explicit evaluation of these models generates near real-time estimates of IMs. As a result, empirical GMMs serve as the fundamental technique in many software packages used for earthquake analysis (e.g., ShakeMap⁵). However, the sparsity of catalogs and datasets, together with regional differences and large variability in the characteristics of large earthquakes, compromise their predictive capacity.

Given these limitations, a functional earthquake analysis demands research on complementary strategies that retain the evaluation speed of empirical GMMs while providing physics-based precision. We propose a Machine Learning (ML) methodology that combines the best of both

¹Barcelona Supercomputing Center, Computer Application in Science and Engineering and Computer Sciences departments, Barcelona, Spain. ²Southern California Earthquake Center, Los Angeles, CA, USA. ³Universitat Politècnica de València, Depto. de Informática de Sistemas y Computadores, Valencia, Spain.

⁴ETH Zurich, Department of Earth Sciences (D-ERDW), Institute of Geophysics, Zurich, Switzerland. ✉e-mail: marisol.monterrubio@bsc.es

approaches, capturing physical information from well-curated simulations (directivity, topography, site effects) with times-to-solution analogous to empirical GMMs, helping to produce the next generation of shaking maps. The ML-based EStimator for ground-shaking maps (MLESmap) exploits two supervised ML algorithms—an ensemble learning model, Random Forest, and a connectionist method based on Deep Neural Networks—trained with data from a CyberShake Study 15.4^{6,7} (CSS-15.4) database of physics-based simulations to generate regional ML-based models that accurately estimate an IM within a few seconds of an earthquake occurrence given its location and magnitude.

With a wide range of ML applications on the rise in earthquake seismology^{8,9}, many recent studies explored ML approaches using seismic observations to train ML-based GMMs and infer IMs^{10–14} or even to synthesize acceleration time-histories¹⁵ and to predict damage states from IMs¹⁶. Even though improvements with respect to traditional regression-based GMMs are reported when sufficient data is available, the existence of such data is not guaranteed for all regions: data availability and its quality can compromise the accuracy and limit applicability. Moreover, there is a paucity of data from large-magnitude events, which have the greatest impact potential, while ML models tend to be poor at extrapolation. Finally, the data is spatially sparse, while synthetics can be generated on a dense and uniform grid. More recently, Withers et al.¹⁷ proposed an artificial neural network (ANN) GMM trained on CyberShake Study 15.12 synthetics⁶ (CSS-15.12), with promising results. The authors choose the same predictor variables as Next Generation Attenuation-West2 (NGA-West2) empirical GMMs¹⁸, including rupture, velocity, and site effect parameters (e.g., fault width, V_{S30} , $Z_{1.0}$, $Z_{2.5}$) in order to compare results and explore ML approaches to complement empirical GMMs in data-poor areas.

We present an MLESmap application in Southern California, proposing regional ML-based GMMs trained on more than 150,000 physics-based CSS-15.4 scenarios. Unlike Withers et al.¹⁷, to maximize applicability, we choose only elementary information as predictor variables, namely the event location and magnitude, which are rapidly estimated and made available by international agencies. We assume that site effects and other complexities are learned implicitly by the models and do not need to be

accounted for explicitly. In particular, uncertainty in rapid estimates of such complex predictors can become an issue at the time of application to the detriment of prediction accuracy. In the presented work, RotD50¹⁹—also typically used by empirical GMMs—is the target of the learning process. Given the well-tailored synthetic database, the algorithms used and the elementary input characteristics prove efficient in estimating RotD50 values in the region. For events that are within the parameters range of the training database, our predictions using MLESmap-generated models outperform empirical GMM solutions requiring similar evaluation times both for synthetic events (reduction up to 45% in median RMSE) and for real historical earthquakes in Southern California (reduction of 11–88% in RMSE).

Results

MLESmap training and predictive capacity

MLESmap was used to train ML-based GMMs for Southern California using a CyberShake 15.4 database^{6,7} (CSS-15.4) of 3D simulations. Figure 1a shows the region of study with the considered faults and the network of sites in CSS-15.4, Fig. 1b shows the RotD50 distribution at different periods, and Fig. 1c summarizes the magnitude distribution of all scenarios.

The dataset was subdivided into training and validation subsets, which contain synthetic IMs for 153,628 scenarios. Two ML algorithms—the Random Forest (RF) and Deep Neural Networks (DNN)—were used to generate eight independent ML models with four-period bands per algorithm ($T = 2, 3, 5$, and 10 seconds). The hypocentral location (latitude, longitude, and depth), magnitude, site latitude, and longitude, and the spatial relation between the hypocenter and the site (the Euclidean distance and the azimuth) were used as the predictors, while the logarithm of RotD50¹⁹ was chosen as the supervised learning target due to the high dynamic range in its values.

Regression metrics remain consistent for all models and imply high predictive power (Supplementary Table 1). In particular, the average R^2 of 0.86 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%. This robust predictive capacity of RF and DNN algorithms is illustrated in Fig. 2,

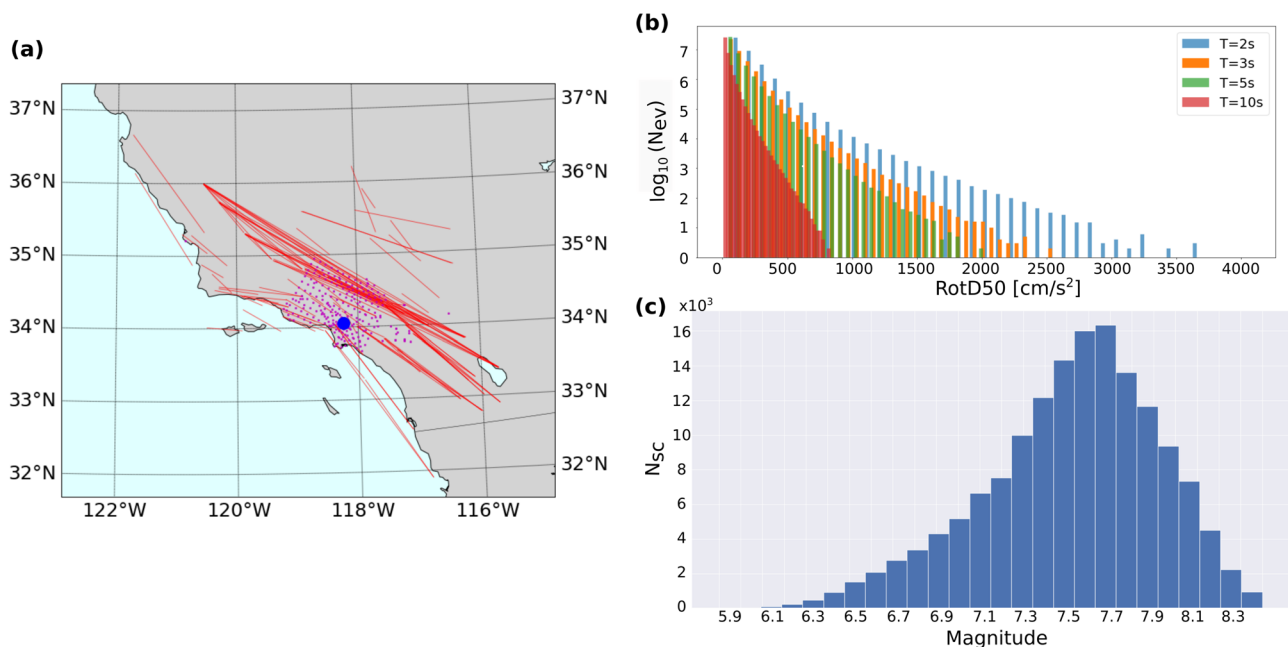


Fig. 1 | Relevant statistical characteristics of the CyberShake Study 15.4 (CSS-15.4) datasets. **a** Map of ground motion sites (magenta dots) and faults (red lines) that are accounted for the dataset, together with the Los Angeles city center (blue dot) as reference. **b** Histogram of RotD50 distribution for all events (number of events in logarithmic scale versus RotD50). **c** Histogram of magnitude distribution

for all seismic scenarios (number of seismic scenarios in thousands versus magnitude), where a predominant magnitude of 7.6 is observed. CyberShake uses a magnitude cutoff of 6.5, so only events with a median magnitude of at least 6.5 are considered, though aleatory magnitude variability implies that some events with lower magnitudes are included.

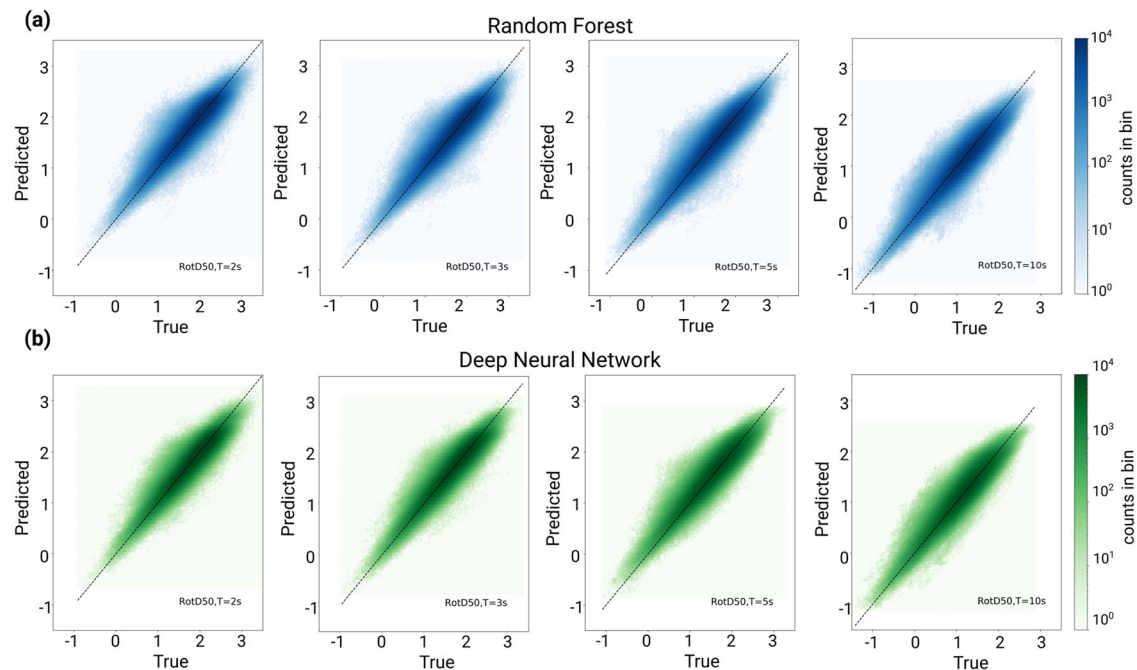


Fig. 2 | Model evaluation on the validation dataset. True vs. predicted values for the 3.8M event realizations from the validation subset, where an event realization is a single $\log(\text{RotD50})$ recording on a single site for one hypothetical scenario from the validation dataset. True refers to RotD50 values in logarithmic units directly for the synthetics, while predicted refers to ML inferences. **a** shows the results for the RF algorithm in blue and **b** for the DNN algorithm in green. Columns, from left to right, correspond to the four considered periods, namely $T = 2\text{s}$, 3s , 5s , and 10s . Given the

number of data values, a color intensity map has been used to display the density of data counts in each of the 100×100 cells of each plot, with dark hues indicating a high count. The dashed black line shows a perfect prediction for reference. The predictions are consistent for all models. The corresponding averages of the six considered score metrics for all event realizations are summarized in Supplementary Table 1.

where the $\log(\text{RotD50})$ predictions are plotted against the synthetic reference. The statistical distribution of the histograms of the RotD50 predicted values closely reflects the distribution of the reference values for each magnitude bin, with the corresponding low average RMSE of the RotD50 median for all magnitudes, further reflecting the high coherence of predictions (Supplementary Fig. 1).

Refer to the Methods section for further information on the CSS-15.4 dataset, the MLEsmap methodology, the ML algorithms, and the validation score metrics.

MLEsmap v.s. empirical GMM for synthetic earthquakes

To assess the applicability of the ML approach in Southern California, the regression score between the actual RotD50 synthetic values from CSS-15.4 is compared both with the predictions from the ML models as well as with predictions from the empirical ASK-14 Ground Motion Model (GMM)²⁰. Such GMMs, obtained by statistical regression from empirical observations in a given region, are the foundation of seismic hazard analysis, so measuring the relative performance of the two approaches is important to showcase the operational relevance of MLEsmap. ASK-14 serves as a state-of-the-art benchmark for our ML results.

Figure 3 shows the RMSE distribution in 0.1 magnitude bins. Since RMSE scales with amplitude, an increase in the RMSE values can be observed with increasing periods for all models: the values differ for similar relative errors and broadly different amplitude ranges. Both MLEsmap models show a similar predictive capacity and in this particular case outperform the ASK-14 predictions, with a reduction of up to 45% in the median RMSE. Since RMSE does not provide information on prediction bias, we also consider the geometric mean ratio of Aida's number K^{21} , with its distribution in 0.1 magnitude bins shown in Supplementary Fig. 2. Underestimation and overestimation correspond to $K > 1$ and $K < 1$, respectively. The RF and DNN predictions do not show a particular bias, whereas ASK-14 tends to underpredict RotD50 for larger events ($>M_W 6.5$) and overpredict for smaller events ($<M_W 6.5$).

The plots in Fig. 4, Supplementary Fig. 3, and Supplementary Fig. 4 show examples of RotD50 maps for three events of magnitudes $M_W = 6.85$, 7.45 , and 8.05 , respectively. As in Fig. 3, the RMSE between the inferences and the reference values indicates a better fit for the MLEsmap models than for ASK-14, with ASK-14 consistently underestimating RotD50 (as indicated by Aida's number K for events larger than $M 6.5$). Qualitatively, the spatial distribution of RotD50 in MLEsmap models reflects that of the reference maps.

MLEsmap v.s. empirical GMM for real earthquakes

We further benchmarked MLEsmap and ASK-14 prediction with observed seismological records from five significant earthquakes that occurred in the region, namely:

1. M7.2 1992 Landers (LND) with hypocenter at 116.44°W , 34.19°N , and 7.6 km deep. The LND earthquake was a right-lateral strike-slip event²². It was the largest event in Southern California in the last century, causing severe damages including over 400 casualties²³.
2. M7.1 1999 Hector Mine (HM) with hypocenter at 116.27°W , 34.57°N , and 8.05 km deep. Evidence suggests that HM was triggered by the LND earthquake²⁴. It was classified as a very strong event on the Mercalli Intensity scale²⁵.
3. M6.7 1994 Northridge (NOR) with hypocenter at 118.54°W , 34.203°N , and 17.4 km deep. NOR was close to the Los Angeles downtown on an undiscovered blind thrust fault²². It caused more than 5000 injuries and an economic loss of $50\text{b US\$}$ ²⁶.
4. M6.1 1986 North Palm Springs (NPS) with hypocenter at 116.61°W , 34.00°N and 10.9 km deep. NPS occurred along the San Andreas Fault producing very strong shaking up to 0.778 g ^{27,28}.
5. M5.9 1987 Whittier (WHI) with hypocenter at 118.08°W , 34.05°N , and 14.6 km deep. WHI mainly affected Los Angeles and Orange counties and generated economic losses of up to $400\text{M US\$}$ ^{29,30}.

The ground motions for the earthquakes were extracted from the SCEC broadband platform (BBP)³¹, with the stations available for each event

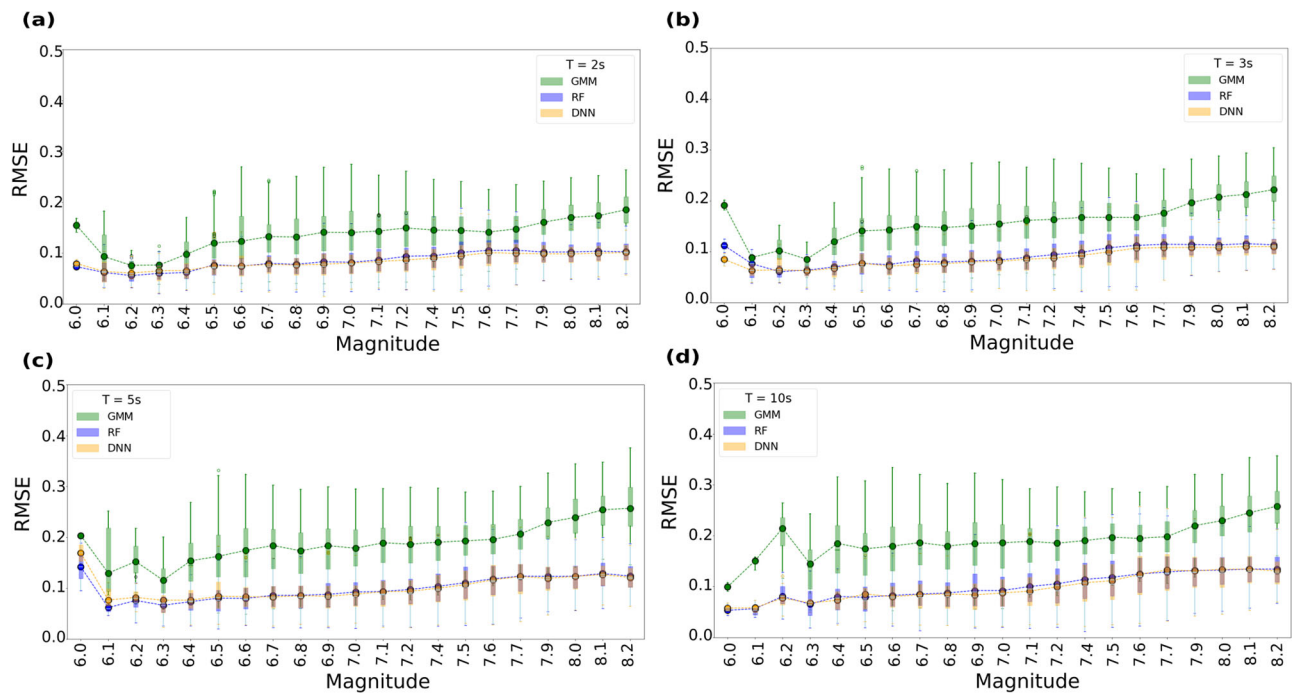


Fig. 3 | RMSE distribution comparison of log(RotD50) for ASK-14 GMM, RF, and DNN estimations on the synthetic result. RMSE distribution of log(RotD50) for the ASK-14 GMM (green boxplots), RF (blue boxplots), and DNN (orange boxplots) estimations against the synthetic results shown per magnitude bin for **a** $T = 2s$, **b** $T = 3s$, **c** $T = 5s$, **d** $T = 10s$. The median value of the RMSE, marked with a

circle for each boxplot, is reduced up to 45% for the ML estimates relative to ASK-14. The mean of the median RMSE value for all magnitudes is summarized in Supplementary Table 2. Boxplot bars represent the first and third quartiles of the metric distribution.

shown in blue in Fig. 5. As some stations are located far from the region used for training the ML models, the models need to extrapolate information to make predictions. We expect this to negatively affect performance since IMs can exhibit significant spatial variations, and the LA basin is known for its strong local site effects. Therefore, we separate station locations into two distinct groups depending on their distance to the sites used for model training: those that are further than the first quartile (Q1) value of all inter-site distances are considered as ‘outside’ stations, while the remaining ones as ‘inside’. The percentage of the ‘inside’ stations is 80%, 67%, 97%, 33%, and 100% for LND, HM, NOR, NPS, and WHI, respectively.

We evaluated the predictions of the models at the BBP station locations (rather than synthetic training sites). As in the synthetic comparisons, the RF, DNN, and ASK-14 RotD50 were evaluated for each period. The RMSE obtained for the ‘inside’ and ‘outside’ stations are shown in Fig. 6 and Supplementary Fig. 5, respectively. The percentage improvement (positive values) or deterioration (negative values) relative to the ASK-14 predictions are annotated for both RF and DNN algorithms. For the ‘inside’ stations (Fig. 6), the ML models prove to be significantly better at predicting RotD50 values for earthquakes that fall within the magnitude range of the events in the synthetic training set (14–73% improvement for RF relative to ASK-14, 19–88% improvement for DNN). The ML models, however, fail to match the ASK-14 inferences for the Whittier earthquake. As $M_W < 6.0$ events are not included in the CSS-15.4 dataset (see Fig. 1c), the models are extrapolating in this case and, as for the spatial extrapolations for ‘outside’ stations (Supplementary Fig. 5), underperform relative to ASK-14.

In Fig. 7, we show the spatial distribution of the 2s RotD50 predictions for all real events. The RF and DNN maps are consistent for LND, HM, NOR, and NPS, where the algorithms do not extrapolate beyond the bounds of the training data. The ML predictions also show more spatial variability than the ASK-14 predictions, reflecting the realistic physical assumptions embedded in the CSS-15.4 simulations. In particular, the local amplification is most pronounced for the $>M_W 7$ events (LND and HM), while ASK-14 does not capture such effects. Finally, the NPS maps reflect the very large reduction in RMSE relative to ASK-14 (in Fig. 6) for this event, as ASK-14

clearly fails to provide a realistic spatial distribution of RotD50. The corresponding maps for periods of 3, 5, and 10 s are shown in Supplementary Figs. 6–8, respectively. It should be noted that the results on real earthquakes, combined with the results on the synthetic events, are a good indication that synthetics in the database represent the physics well and can accurately predict earthquake motion in the region.

Discussion

MLESmap prediction capacity of ground motions

We generated the MLESmap methodology and associated models toward a more accurate rapid response solution by combining the accuracy of the physics-based simulations with the fast estimations given by empirical GMMs. Our methodology, applied to a high-quality simulation dataset, can predict RotD50 for real earthquakes more accurately than ASK-14, in a similar time, and using only primary information available shortly after an event. The ML models can be evaluated instantaneously and provide a reliable complement or alternative for the early assessment of the impact of a future earthquake in Southern California.

It should be noted that MLESmap models perform better than empirical GMMs as long as the earthquakes are compatible with the CSS-15.4 dataset, that is, the locations, frequencies, and magnitudes interrogated are within the bounds of the training dataset (see Fig. 1). Predicting any kind of temporal information related to the events such as travel times, shaking duration, or phases is not included in the current implementation of MLESmap, although ML models have also been employed to synthesize time-series¹⁵. Nevertheless, IMs such as RotD50 are often preferred for their direct relation to the impact of an earthquake and are a key component in rapid post-disaster analyses³².

Outlooks for synthetic training databases of ground motions

MLESmap has been validated using a pre-existing high-quality synthetic database. At present, new CyberShake studies are being carried out using updated velocity models, rupture generators, and higher, stochastically simulated frequencies⁶. A direct follow-up of the present study is the

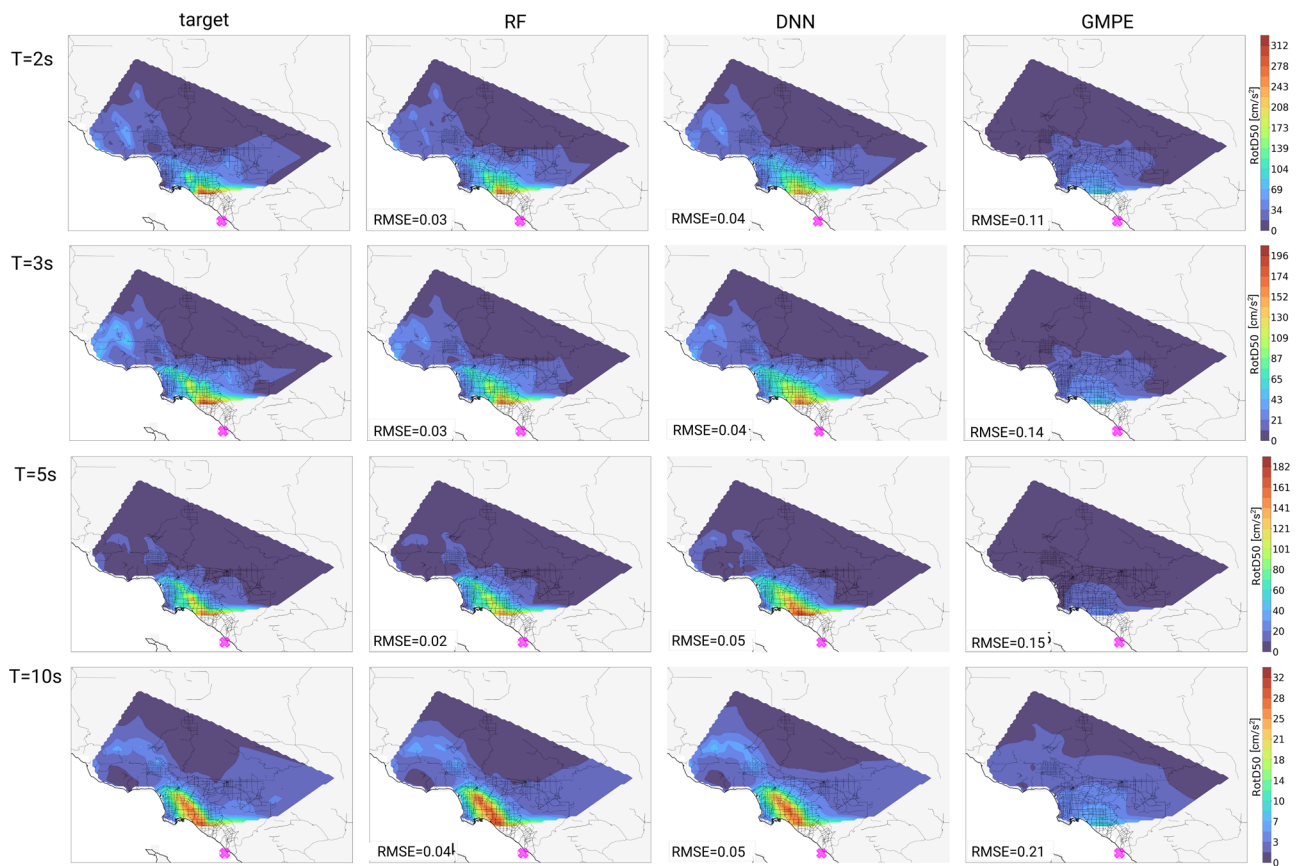


Fig. 4 | RotD50 predictions on a validation event of magnitude 6.85. The reference RotD50 map (first column), as well as RF (second column), DNN (third column), and ASK-14 (fourth column) predictions given in cm/s^2 for a synthetic validation event: an $M_W = 6.85$ earthquake located at 33.45°N , 117.73°W , and 6.1 km deep

(pink cross). Rows show the increasing period from $T = 2$ s, 3 s, 5 s, and 10 s. The RMSE score metrics 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.

application of MLESmap to such databases and the evaluation of their impact both on the training and on the predictive capabilities of the models, in particular for shorter periods relevant to seismic risk mitigation strategies. Other applications could use databases tailored specifically for MLESmap, allowing for applications in areas other than Southern California³³. Populating ad-hoc databases for use in MLESmap, however, poses problems in finding the optimal set of input parameters. Accounting for large and rare earthquakes is particularly important, as they have the highest damage potential and are poorly represented in empirical GMMs due to the scarcity of high-quality data. In a process sensitive to input parameters and strongly constrained by computing power, designing a new database requires a detailed analysis, benchmarking, and validation³⁴ prior to implementation.

Implications for ML in rapid ground motion IM estimates

Since ML models trained on synthetics can predict the IMs of real earthquakes more accurately than empirical GMMs and at a similar time, they are bound to become the next-generation tool for post-disaster analysis that complements existing data-based approaches in guiding relief efforts. Rapid hardware and software developments will progressively decrease the associated computational costs and render the generation of high-quality synthetic databases and the subsequent model training more accessible and, thus, more suitable for routine use.

As decreasing uncertainty in predicted IMs is of paramount importance, further event features, such as the focal mechanism or rupture extent, could be included to train MLESmap models, provided that such parameters can be assessed rapidly for a new earthquake. Such additional input characteristics may result in ML models that yield even more accurate inferences. Integrating the scope-limited MLESmap with the more general empirical GMMs in a hybrid approach¹¹ could also lead to significant improvements

and render the approach more generally applicable. While MLESmap, in its current implementation, will potentially perform poorly at predictions for earthquakes outside of the parameters of its training dataset, our results suggest that empirical GMMs could cover such gaps. Another way to address the scope limitations would be to explore transfer learning³⁵, a technique where learning from one task can be reused to improve the performance of a related task (with tasks here understood as region-specific predictions).

We also foresee interesting applications where massive computations of IM inference are needed, for example for uncertainty quantification. Finally, quick estimates of IMs could be used to provide fast PSHA estimates for operational earthquake forecasting, for example whenever aftershocks are expected.

Methods

Synthetic physics-based earthquake ground motions

In this work we leveraged a large dataset generated via physics-based wave propagation simulations to generate ground motions from hundreds of thousands of hypothetical earthquakes to train ML algorithms. The data was generated using the CyberShake platform and, in particular, the CSS-15.4 study for the Southern California region.

CyberShake, developed by the Southern California Earthquake Center (SCEC), is an integrated collection of scientific software that performs PSHA by using 3D physics-based modeling. It simulates ground motions for a large suite of earthquakes derived from an earthquake rupture forecast (ERF) and has been used to assess seismic hazards in California in multiple studies^{6,7}. Simulations are based on seismic reciprocity, so two unit impulses at a given ground site, one in each horizontal direction, are propagated to fault surfaces to calculate their Strain Green Tensor (SGT) response.

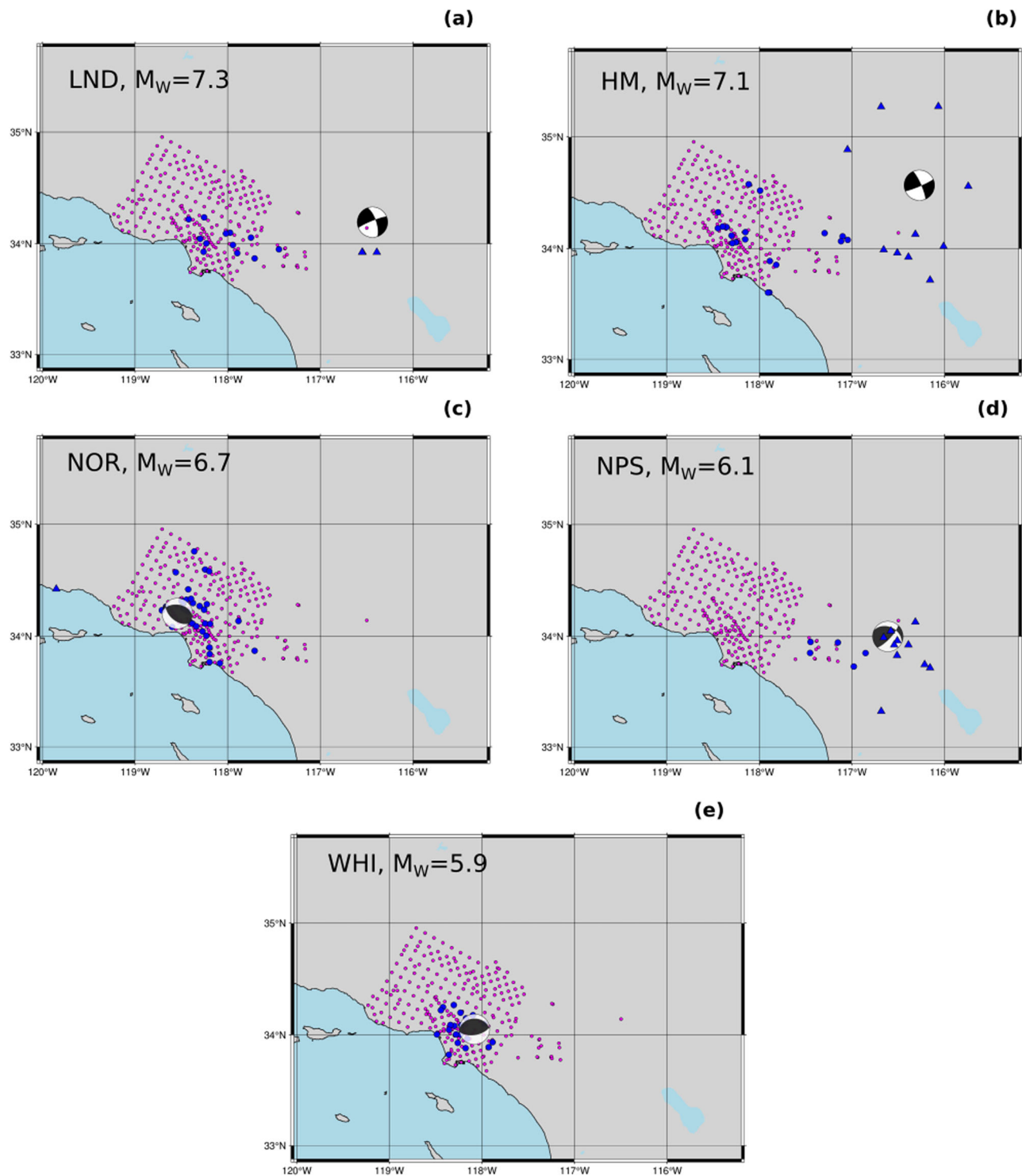


Fig. 5 | Spatial configuration of historic earthquakes and BBP stations. The epicentral location and focal mechanism are shown as a beachball for **a** Landers, **b** Hector Mine, **c** Northridge, **d** North Palm Springs, and **e** Whittier earthquakes. The BBP stations where the observations analyzed in this work were acquired are plotted in blue and separated into 'inside' (stars) and 'outside' (triangles) stations,

referring to whether the stations are within or beyond the area covered by the synthetic training sites (magenta dots). This division stems from the generally poor extrapolation performance of ML models—we expect lower-quality predictions for stations that are far from the training locations. See the main text for a full definition of 'inside' and 'outside' locations.

Then, the SGTs are convolved with slip time histories for each event to produce a seismogram at the site of interest. Thus, CyberShake simulations computationally scale with the number of sites (generally on the order of hundreds), and not with the total number of potential earthquakes (usually in the order of tens to hundreds of thousands), allowing to model an arbitrary number of earthquakes for a given area. Although this platform was

developed to perform PSHA, the rich suite of its data products makes it an ideal source of data to feed the MLESmap method.

CSS-15.4 used in this paper is a well-curated computational study to calculate a physics-based PSHA for Southern California at 1 Hz, using the tomographically-derived Community Velocity Model CVM-S4.26-M01, the GPU implementation of AWP-ODC-SGT³⁶, the GP-14 kinematic

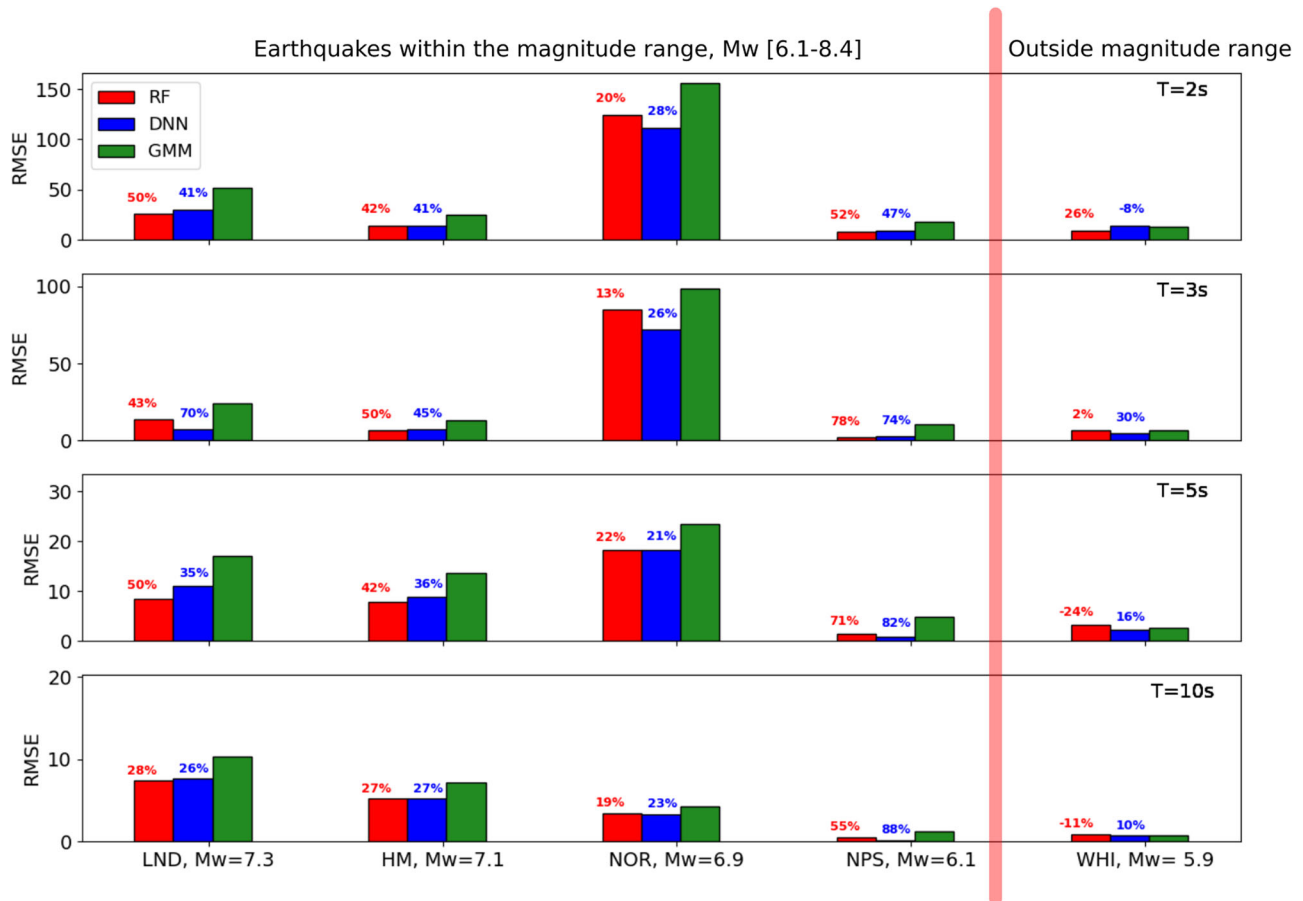


Fig. 6 | RotD50 predictions for real events for the 'inside' stations. RMSE for RF (red bars), DNN (blue bars), and ASK-14 (green bars) for the 'inside' stations shown as blue circles in Fig. 5. Each row indicates the RMSE for $T = 2s, 3s, 5s$, and $10s$, respectively. The value annotated above each bar indicates the improvement of the MLESmap predictions with respect to the ASK-14 predictions. The vertical red line divides the earthquakes that have magnitudes that fall in the range of hypothetical events included in CSS-15.4 (left of vertical line) from the event with a magnitude outside the simulated magnitude range (see Fig. 1c). The MLESmap predictions outperform the ASK-14 predictions as long as no extrapolations are necessary, that

is, the considered event is representative of the synthetic training events (left v.s. right of the vertical line), and the station locations fall within the area covered by the synthetic training sites (the 'inside' stations in this figure v.s. the 'outside' stations in Supplementary Fig. 5). Note that RMSE scales with amplitude, so the RMSE magnitude in each case depends on the event magnitude and on the location of the 'inside' stations relative to the event. The relative performance of the ML-based predictions v.s. the empirical GMM is of interest rather than absolute values for each event.

rupture method with uniform hypocenters³⁷, and the UCERF2 ERF⁷. The computation of the CSS-15.4 study required a total of 37.6M hours in tier-0 supercomputing facilities. In particular, the database contains IMs (derived from simulations) for 153628 scenarios (i.e., hypothetical earthquakes) from the UCERF2 earthquake rupture forecast. Those scenarios were recorded at a collection of sites, i.e., discrete points in space on the free surface, where seismic IMs were extracted from each scenario. Such IMs can be further analyzed to obtain discrete spectral intensity values.

Some relevant characteristics of the CSS-15.4 dataset are shown in Fig. 1. The faults are marked with red lines, the Los Angeles city center with a blue dot, and the network of ground sites where RotD50 is computed with magenta stars in Fig. 1a. Figure 1b shows the RotD50 distribution for the events at different periods. As expected, lower periods attain higher RotD50 values. The magnitude distribution of all synthetic scenarios is shown in Fig. 1c, where a predominant 7.6 magnitude is observed. CyberShake uses a magnitude cutoff of 6.5, so only events with a median magnitude of at least 6.5 are considered, though aleatory magnitude variability implies that some events with lower magnitudes are included. Since CyberShake considers multiple realizations with varying hypocenter locations and slip distributions to sample variability, the distribution of earthquakes in the event set is not directly related to actual earthquake magnitude distribution in the region of study.

Finally, it should be noted that the MLESmap methodology is not limited specifically to training on CyberShake-like databases. In CyberShake, although multiple kinematic rupture scenarios are derived for each rupture in the ERF by varying slip distributions and hypocenter locations across an input fault system, the rupture speeds and slip distributions in each simulation are pre-set. Having a richer set of rupture conditions by considering a synthetic database of dynamic rupture simulations could further improve the applicability of the method.

MLESmap: methodology

The goal of the MLESmap methodology is to rapidly predict IMs with high spatial resolution given, as inputs, the earthquake's magnitude, hypocentral location, and the relationship between the earthquake hypocentre and the site of ground motions recordings (see Supplementary Fig. 9 for a graphical summary of the MLESmap methodology). In particular, the MLESmap methodology provides a framework for training region-specific ML models of ground motion on databases of synthetic IMs with two algorithms: Random Forest (RF) and Deep Neural Networks (DNN). The resulting ML models, the RF-based GMM and DNN-based GMM predict the distribution of a selected IM at a given period for given source characteristics of a new event in the region.

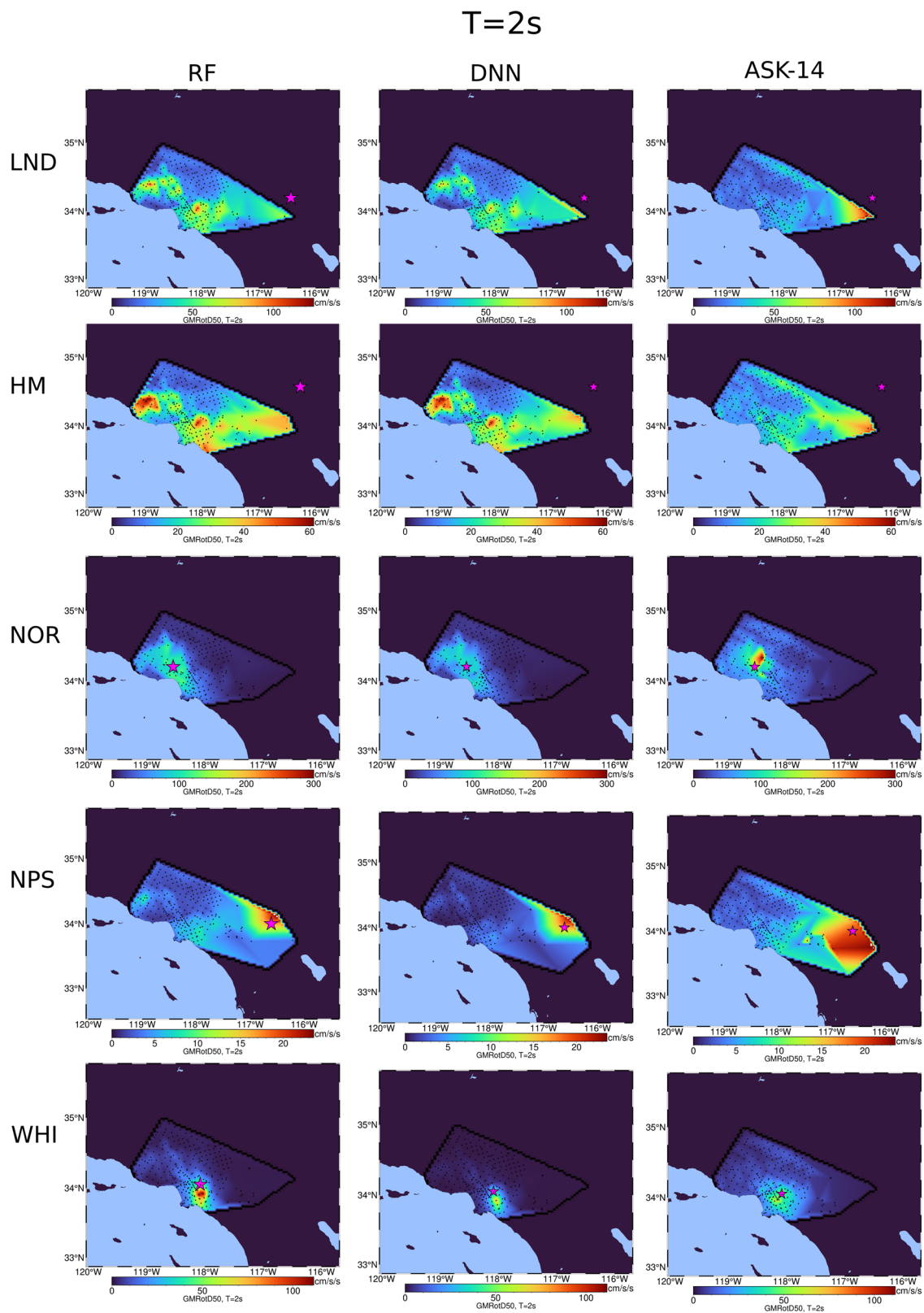


Fig. 7 | RotD50 predictions for $T=2s$ for all real events in the training domain. The RotD50 maps of RF (first column), DNN (second column), and ASK-14 (third column) predictions are given in cm/s^2 . Each row corresponds to one of the real

events (magenta star), namely Landers (LND), Hector Mine (HM), Northridge (NOR), North Palm Springs (NPS) and Whittier (WHI). See Supplementary Fig. 6–8 for the corresponding plots of the predictions for periods of 3 s, 5 s, and 10 s.

Given our objective and the associated time constraints, we choose parameters that can be quickly inferred from early data records and are readily provided by international agencies immediately after an event. It should be noted that despite the simple inputs, the methodology takes into account the complex relationship between the finite fault models used in the ground motion simulations that generate the synthetics, the associated hypocentral location, and the resulting spatial distribution of the chosen IM.

In this paper, the MLEsmap predictions are made by means of ML models for Southern California trained with CSS-15.4 events only³⁸. The CSS-15.4 database is one of the largest and best-calibrated datasets of synthetics worldwide, thus serving as an ideal first implementation of MLEsmap. For both RF and DNN algorithms, four independent ML models considering the data for a single maximum period ($T = 2, 3, 5$, and 10 s) were built. The logarithm of the RotD50—which represents the median pseudo-spectral acceleration (PSA) value selected from all azimuth directions at each recording station³⁹—was used as a target, as the logarithmic scale was shown to increase the score performance for the four periods considered (see Supplementary Fig. 10). It should be noted that other IMs could serve targets without significant impact on the performance of the MLEsmap models.

We refer to the collection of all scenarios recorded at all sites for a discrete spectral intensity, or in general for all spectral intensities, as *events*. The dataset of 153,628 scenarios was split into 90% for training and 10% for testing and validation. Therefore, considering 253 stations in our dataset (Fig. 1a), a total of 155M events were used for training the models generated with the MLEsmap methodology, involving 4 discrete values of RotD50 (at different periods), which were used independently from each other. A total of 3.8M events belonging to the validation subset were used to compute the score metrics and evaluate the MLEsmap models' accuracy. The training dataset constitutes 3.8 GB on disk, while the validation subset is 0.5 GB.

MLEsmap: RF regression algorithm

The training and model generation for the RF regressor has been performed using the Distributed Computing Library (dislib)⁴⁰, a Python software built on top of PyCOMPSs⁴¹. dislib is inspired by NumPy and scikit-learn, providing various supervised and unsupervised learning algorithms through an easy-to-use API. dislib has been used due to its efficiency to handle models with a large number of events.

To find the best hyperparameters for a model with the selected input characteristics (magnitude, hypocentral latitude, longitude, and depth, latitude and longitude coordinates of the site, and the Euclidean distance and azimuth that define the spatial relationship between the hypocentre and the site) we use a grid search on the training set, and the k-folds cross-validation functions over the three available parameters in dislib for the algorithm:

- maximum tree (d_{\max}): number of levels in each decision tree,
- number of estimators (n_{est}): number of trees in the forest, and
- try-features (t_f): maximum number of features considered for splitting a node.

The metric to measure the performance of the models given a specific set of hyper-parameters is the coefficient of determination (R^2) score. Supplementary Fig. 10 shows the R^2 -score values for different d_{\max} values using two different target scales: (a) logarithmic and (b) non-logarithmic. The results show that the treatment of the target on a logarithmic scale increases the score performance for the four periods considered. Moreover, the best parameters are $d_{\max} = 30$, $n_{\text{est}} = 30$, and $t_f = \text{'third'}$ for all periods.

MLEsmap: Deep Neural Network topology

MLEsmap uses a fully connected neural network. Many factors were taken into account in determining the DNN architecture, including the nature of the application. Specifically, it is a regression problem with only 8 inputs and

a single output. Tackling it with a “sophisticated” neural network model, such as a Convolutional Neural Network (CNN), a Recurrent Neural Network (RNN), or even a Transformer, was considered unnecessary.

In order to select an appropriate network topology for the target problem, we started from the most basic multilayer perceptron (MLP) with eight neurons in the input layer, corresponding to the eight features of the earthquake that are taken into account (magnitude, hypocentral latitude, longitude, and depth, latitude and longitude coordinates of the site, and the Euclidean distance and azimuth that define the spatial relationship between the hypocentre and the site), and one neuron in the output layer which determines the RotD50 component of the earthquake. The exploration of the topology space was carried out with different numbers of hidden layers and units (neurons) on those layers.

The classical “non-generalization” problem in neural networks was tackled by applying regularization, data normalization, and batch normalization. In addition, distinct learning rate schedulers were evaluated to deal with the local minimum deadlock optimization problem. Finally, to avoid problems due to the vanishing and explosion of gradients, we tested different dropout and activation functions.

After concluding this phase of experimentation with the different periods, the best results were obtained with MLPs consisting of either seven or nine hidden layers, respectively, with 32, 64, 128, 256, 128, 64, and 32 or 16, 32, 64, 128, 256, 128, 64, 32 and 16 units per layer. The MLPs, in addition, integrate pre-batch normalization (before the activation function) and a warm anneal learning rate scheduler. Softplus was adopted as the activation function for the hidden layers and sigmoid for the output layer since the IM values were previously normalized between 0 and 1.

The different experiments with neural networks were carried out in Python, making use of the open-source TensorFlow library⁴², and the high-level framework Keras⁴³. These libraries rely on other auxiliary Python libraries such as os, NumPy, matplotlib, etc.

Validation score metrics

In this work, we used different metrics to validate the accuracy of the MLEsmap inference on synthetics and real events (see Results section).

To quantify the accuracy of the ML predictions on the synthetic data in Fig. 2, we use common regression metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), coefficient of determination R^2 and Pearson's coefficient⁴³. The MAE is calculated as the mean or average of the absolute differences between predicted and expected target values, so the units of the error score correspond to the units of the predictions. The MSE is the mean of the squared differences—the units of the error and of the prediction do not match, but the metric is useful to emphasize and penalize large errors, as the errors are squared before they are averaged. RMSE is the square root of the MSE, so it is measured in the same units as the target variable, yet it still gives a relatively high weight to large errors. MAPE is sensitive to relative errors and so remains insensitive by the scaling of the target variable. R^2 represents the proportion of variance explained by the independent variables in the model and indicates how well-unseen samples are likely to be predicted by the model. Finally, Pearson's correlation coefficient measures the strength of the linear association between two variables. Note that all the metrics summarize performance in ways that disregard the direction of over- or under-prediction.

In particular, to evaluate the ML predictions on the validation subset against empirical GMMs, we focus on the RMSE as score metrics, as it is widely used in the community to compare and validate empirical GMMs with observations^{13,44–46}. RMSE ranges from 0 to infinity, is in the same units as the target variable and is most useful when large errors are particularly undesirable, as in the case of the prediction of ground motion IMs for assessing associated risks. However, since RMSE does not provide any information on prediction bias, we also consider the geometric mean ratio of Aida's number K^{21} , where underestimation and overestimation correspond

to $K > 1$ and $K < 1$, respectively. We compute the logarithm of K as

$$\log(K) = \frac{1}{N} \sum_{i=1}^N \log\left(\frac{O_i}{S_i}\right) \quad (1)$$

where O_i is the observed (target) value, and S_i is the simulated (predicted) value.

It should be noted that multiple other empirical GMMs were proposed for Southern California (e.g., BSSA⁴⁷, CB⁴⁸, and CY⁴⁹) that generate similar results and thus result in comparable RMSE metrics. We deem ASK-14²⁰ the most suitable for our validation, as it was constructed using the NGA-West2 database⁵⁰ that contains worldwide ground motion data recorded from shallow crustal earthquakes in active tectonic regimes post-2000, as well as a set of small to moderate magnitude earthquakes in California between 1998 and 2011. The ASK-14 values were computed using OpenSHA software⁵¹ with the V_{S30} values from the Thompson model⁵².

Data availability

The authors declare that the data supporting the findings of this study are available in the [Zenodo](https://doi.org/10.5281/zenodo.10812284) repository⁵³. Random Forest and Deep Neural Network Models outputs are find in <https://doi.org/10.5281/zenodo.10812284>⁵⁴.

Code availability

Readers can access the training ML codes used in this work to provide the reported inferences in the [Zenodo](https://doi.org/10.5281/zenodo.10812284) repository⁵⁴.

Received: 13 September 2023; Accepted: 1 May 2024;

Published online: 16 May 2024

References

- de la Puente, J., Rodriguez, J. E., Monterrubio-Velasco, M., Rojas, O. & Folch, A. Urgent supercomputing of earthquakes: use case for civil protection. In *Proceedings of the Platform for Advanced Scientific Computing Conference*, 1–8 (2020).
- Ejarque, J. et al. Enabling dynamic and intelligent workflows for HPC, data analytics, and AI convergence. *Future Gener. Comput. Syst.* **134**, 414–429 (2022).
- Douglas, J. Ground motion prediction equations 1964–2021. Department of Civil & Environmental Engineering, University of Strathclyde, Glasgow, UK 670 (2020).
- Petersen, M. D. et al. Documentation for the 2008 update of the United States National Seismic Hazard Maps. Technical Report, US Geological Survey Open-File Report (2008).
- Wald, D. J., Worden, C. B., Thompson, E. M. & Hearne, M. Shakedown operations, policies, and procedures. *Earthq. Spectra* **38**, 756–777 (2022).
- Graves, R. et al. Cybershake: A physics-based seismic hazard model for Southern California. *Pure Appl. Geophys.* **168**, 367–381 (2010).
- Jordan, T. H. & Callaghan, S. Cybershake models of seismic hazards in Southern and Central California. In *Proceedings of the US National Conference on Earthquake Engineering* (2018).
- Li, Y. E., O'Malley, D., Beroza, G., Curtis, A. & Johnson, P. Machine learning developments and applications in solid-earth geosciences: fad or future? *J. Geophys. Res.* **128**, e2022JB026310 (2023).
- Mousavi, S. M. & Beroza, G. C. Machine learning in earthquake seismology. *Annu. Rev. Earth Planet. Sci.* **51**, 105–129 (2023).
- Kotha, S. R., Weatherill, G., Bindi, D. & Cotton, F. A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe. *Bull. Earthq. Eng.* **18**, 4091–4125 (2020).
- Kubo, H., Kunugi, T., Suzuki, W., Suzuki, S. & Aoi, S. Hybrid predictor for ground-motion intensity with machine learning and conventional ground motion prediction equation. *Sci. Rep.* **10**, 1–12 (2020).
- Khosravikia, F. & Clayton, P. Machine learning in ground motion prediction. *Comput. Geosci.* **148**, 104700 (2021).
- Mori, F. et al. Ground motion prediction maps using seismic-microzonation data and machine learning. *Nat. Hazards Earth Syst. Sci.* **22**, 947–966 (2022).
- Zhu, C. et al. How well can we predict earthquake site response so far? Site-specific approaches. *Earthq. Spectra* **38**, 1047–1075 (2022).
- Florez, M. A. et al. Data-driven synthesis of broadband earthquake ground motions using artificial intelligence. *Bull. Seismol. Soc. Am.* **112**, 1979–1996 (2022).
- Xu, Y., Lu, X., Tian, Y. & Huang, Y. Real-time seismic damage prediction and comparison of various ground motion intensity measures based on machine learning. *J. Earthq. Eng.* **26**, 4259–4279 (2022).
- Withers, K. B., Moschetti, M. P. & Thompson, E. M. A machine learning approach to developing ground motion models from simulated ground motions. *Geophys. Res. Lett.* **47**, e2019GL086690 (2020).
- Petersen, M. D. et al. The 2014 United States National Seismic Hazard Model. *Earthq. Spectra* **31**, S1–S30 (2015).
- Boore, D. M., Watson-Lamprey, J. & Abrahamson, N. A. Orientation-independent measures of ground motion. *Bull. Seismol. Soc. Am.* **96**, 1502–1511 (2006).
- Abrahamson, N. A., Silva, W. J. & Kamai, R. Summary of the ask14 ground motion relation for active crustal regions. *Earthq. Spectra* **30**, 1025–1055 (2014).
- Aida, I. Reliability of a tsunami source model derived from fault parameters. *J. Phys. Earth* **26**, 57–73 (1978).
- Trabant, C. et al. Data products at the iris dmc: stepping stones for research and other applications. *Seismol. Res. Lett.* **83**, 846–854 (2012).
- Hauksson, E., Jones, L. M., Hutton, K. & Eberhart-Phillips, D. The 1992 landers earthquake sequence: seismological observations. *J. Geophys. Res.* **98**, 19835–19858 (1993).
- Freed, A. M. & Lin, J. Delayed triggering of the 1999 Hector mine earthquake by viscoelastic stress transfer. *Nature* **411**, 180–183 (2001).
- Behr, J. et al. Preliminary report on the 16 October 1999 m 7.1 Hector mine, California, earthquake. *Seismol. Res. Lett.* **71**, 11–23 (2000).
- Peek-Asa, C. et al. Fatal and hospitalized injuries resulting from the 1994 Northridge earthquake. *Int. J. Epidemiol.* **27**, 459–465 (1998).
- Stover, C. W. & Coffman, J. L. *Seismicity of the United States, 1568–1989 (revised)* (US Government Printing Office, 1993).
- Jones, L. M., Hutton, L. K., Given, D. D. & Allen, C. R. The July 1986 North Palm Springs, California, earthquake-the North Palm Springs, California, earthquake sequence of July 1986. *Bull. Seismol. Soc. Am.* **76**, 1830–1837 (1986).
- Shepherd, R. The October 1, 1987 Whittier narrows earthquake. *Bull. N.Z. Soc. Earthq. Eng.* **20**, 255–263 (1987).
- Hauksson, E. & Jones, L. M. The 1987 Whittier narrows earthquake sequence in Los Angeles, Southern California: seismological and tectonic analysis. *J. Geophys. Res.* **94**, 9569–9589 (1989).
- Maechling, P. J., Silva, F., Callaghan, S. & Jordan, T. H. Scac broadband platform: system architecture and software implementation. *Seismol. Res. Lett.* **86**, 27–38 (2015).
- Poggi, V. et al. Rapid damage scenario assessment for earthquake emergency management. *Seismol. Res. Lett.* **92**, 2513–2530 (2021).
- Rojas, O. et al. Insights on physics-based probabilistic seismic hazard analysis in south Iceland using cybershake. In *AGU Fall Meeting Abstracts*, vol. 2021, NH31A–05 (2021).
- Rezaei, S., Stewart, J. P., Luco, N. & Goulet, C. A. Findings from a decade of ground motion simulation validation research and a path forward. *Earthq. Spectra* **40**, 346–378 (2024).
- Torrey, L. & Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*, 242–264 (IGI Global, Hershey, 2010).
- Cui, Y. et al. Accelerating cybershake calculations on the xe6/xk7 platform of blue waters. In *2013 Extreme scaling workshop (XSW 2013)*, 8–17 (IEEE, 2013).

37. Graves, R. & Pitarka, A. Refinements to the graves and pitarka (2010) broadband ground-motion simulation method. *Seismol. Res. Lett.* **86**, 75–80 (2015).
38. Cybershake study 15.4. https://strike.scec.org/scecpedia/CyberShake_Study_15.4. Accessed 2023-05-16.
39. Boore, D. M. Orientation-independent, nongeometric-mean measures of seismic intensity from two horizontal components of motion. *Bull. Seismol. Soc. Am.* **100**, 1830–1835 (2010).
40. Cid-Fuentes, J. A., Solà, S., Álvarez, P., Castro-Ginard, A. & Badia, R. M. dislib: Large scale high performance machine learning in python. In *2019 15th International Conference on eScience (eScience)*, 96–105 (IEEE, 2019).
41. Tejedor, E. et al. Pycompss: Parallel computational workflows in python. *Int. J. High. Perform. Comput. Appl.* **31**, 66–82 (2017).
42. Abadi, M. et al. TensorFlow: Large-scale machine learning on heterogeneous systems (2015). <https://www.tensorflow.org/>. Software available from tensorflow.org.
43. Chicco, D., Warrens, M. J. & Jurman, G. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Comput. Sci.* **7**, e623 (2021).
44. Feng, T. & Meng, L. A high-frequency distance metric in ground-motion prediction equations based on seismic array backprojections. *Geophys. Res. Lett.* **45**, 11–612 (2018).
45. Kuehn, N. M., Kishida, T., AlHamaydeh, M., Lavrentiadis, G. & Bozorgnia, Y. A Bayesian model for truncated regression for the estimation of empirical ground-motion models. *Bull. Earthq. Eng.* **18**, 6149–6179 (2020).
46. Sabermahani, S. & Ashjanas, P. Sensitivity analysis of ground motion prediction equation using next generation attenuation dataset. *Geod. Geodyn.* **11**, 40–45 (2020).
47. Boore, D. M., Stewart, J. P., Seyhan, E. & Atkinson, G. M. Nga-west2 equations for predicting PGA, PGV, and 5% damped PSA for shallow crustal earthquakes. *Earthq. Spectra* **30**, 1057–1085 (2014).
48. Campbell, K. W. & Bozorgnia, Y. Nga-west2 ground motion model for the average horizontal components of PGA, PGV, and 5% damped linear acceleration response spectra. *Earthq. Spectra* **30**, 1087–1115 (2014).
49. Chiou, B. S.-J. & Youngs, R. R. Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. *Earthq. Spectra* **30**, 1117–1153 (2014).
50. Ancheta, T. D. et al. Nga-west2 database. *Earthq. Spectra* **30**, 989–1005 (2014).
51. Field, E. H., Jordan, T. H. & Cornell, C. A. Opensha: a developing community-modeling environment for seismic hazard analysis. *Seismol. Res. Lett.* **74**, 406–419 (2003).
52. Thompson, E. M., Wald, D. J. & Worden, C. B. A VS30 map for California with geologic and topographic constraints. *Bull. Seismol. Soc. Am.* **104**, 2313–2321 (2014).
53. Monterrubio-Velasco, M. Source data for graphs and charts used in the paper “a machine learning-based estimator for real-time earthquake ground-shaking predictions in Southern California” <https://doi.org/10.5281/zenodo.10640493> (2024).
54. Monterrubio-Velasco, M. Model output and training codes used in the paper “a machine learning-based estimator for real-time earthquake ground-shaking predictions in Southern California” <https://doi.org/10.5281/zenodo.10812284> (2024).
- (PCI2021-121957), project eFlows4HPC. This research has been supported by the European High-Performance Computing Joint Undertaking (JU) as well as Spain, Italy, Iceland, Germany, Norway, France, Finland, and Croatia under grant agreement no. 101093038, (ChEESE-CoE), project PCI2022-134980-2 funded by MCIN/AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR). The authors acknowledge the Center for Advanced Research Computing (CARC) at the University of Southern California for providing computing resources that have contributed to the research results reported within this publication. URL: <https://carc.usc.edu>. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, and the National Center for Supercomputer Applications under the Blue Waters Sustained-Petascale Computing Project. SCEC is funded by USGS Cooperative Agreement G17AC00047 and NSF Cooperative Agreement EAR-1600087. The authors also thank Dr. Kevin Milner, who provided the ASK-14 estimations needed for the comparisons presented in this paper. The authors thank Dr. Arnau Folch who reviewed the draft of the paper, improving it with his comments and suggestions.

Author contributions

M.M.V. and J.C.C. conceived the experiments; S.C. provided the data and support to carry out the experiments; M.M.V. and P.P. conducted the experiments; F.V.N, R.M.B. and E.Q.O. provided the technical support on the experiments; D.M, J.P, M.P. and M.M.V analyzed the results. M.M.V, D.M, J.P, E.Q.O, J.C.C., F.V.N, and R.M.B. wrote the first version of the paper, while M.P. rewrote the text in its final form. All authors reviewed the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s43247-024-01436-1>.

Correspondence and requests for materials should be addressed to Marisol Monterrubio-Velasco.

Peer review information *Communications Earth and Environment* thanks Elif Oral and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Joe Aslin. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024, corrected publication 2025

Acknowledgements

This work has been funded by the European Commission's Horizon 2020 Framework program and the European High-Performance Computing Joint Undertaking (JU) under grant agreement No 955558 and by MCIN/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR