Lilienkamp, H., von Specht, S., Weatherill, G., Caire, G., Cotton, F. (2022): Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach. - Bulletin of the Seismological Society of America, 112, 3, 1565-1582.

https://doi.org/10.1785/0120220008

Ground-Motion Modeling as an Image Processing Task: Introducing a Neural Network Based, Fully Data-Driven, and Nonergodic Approach

Henning Lilienkamp*[1,2], Sebastian von Specht[3], Graeme Weatherill[1], Giuseppe Caire[4], and Fabrice Cotton[1,2]

ABSTRACT

We construct and examine the prototype of a deep learning-based ground-motion model (GMM) that is both fully data driven and nonergodic. We formulate ground-motion modeling as an image processing task, in which a specific type of neural network, the U-Net, relates continuous, horizontal maps of earthquake predictive parameters to sparse observations of a ground-motion intensity measure (IM). The processing of map-shaped data allows the natural incorporation of absolute earthquake source and observation site coordinates, and is, therefore, well suited to include site-, source-, and path-specific amplification effects in a nonergodic GMM. Data-driven interpolation of the IM between observation points is an inherent feature of the U-Net and requires no a priori assumptions. We evaluate our model using both a synthetic dataset and a subset of observations from the KiK-net strong motion network in the Kanto basin in Japan. We find that the U-Net model is capable of learning the magnitude–distance scaling, as well as site-, source-, and path-specific amplification effects from a strong motion dataset. The interpolation scheme is evaluated using a fivefold cross validation and is found to provide on average unbiased predictions. The magnitude–distance scaling as well as the site amplification of response spectral acceleration at a period of 1 s obtained for the Kanto basin are comparable to previous regional studies.

KEY POINTS

- We develop a fully data driven and spatially varying ground-motion model using U-Net neural networks.
- Image processing techniques naturally capture source, site, and path effects in ground-motion prediction.
- Our extensible framework is a step toward spatially varying ground-motion modeling without prior assumptions.

Supplemental Material

INTRODUCTION

A ground-motion model (GMM) relates the probability distribution of ground-motion intensity caused by an earthquake to a set of predictive parameters, for example, earthquake magnitude, distance at which the shaking is observed, and observation site characteristics that influence the local amplification. Commonly, GMMs are equations that are expert designed to represent physical processes causing observable relations in strong motion data, such as decreasing ground-motion intensity with increasing distance from the earthquake source. The coefficients of these equations are calibrated via regression analysis using observations from the

1. Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Seismic Hazard and Risk Dynamics, Potsdam, Germany, https://orcid.org/0000-0001-9386-7532 (HL); https://orcid.org/0000-0001-9347-2282 (GW); https://orcid.org/0000-0002-9242-3996 (FC);
2. Institute of Geosciences, University of Potsdam, Potsdam, Germany; 3. Institute of Mathematics, University of Potsdam, Potsdam, Germany, https://orcid.org/0000-0003-3735-9562 (SvS);
4. Faculty of Electrical Engineering and Computer Science, Technische Universität, Berlin, Germany, https://orcid.org/0000-0002-7749-1333 (GC)

***Corresponding author:** lilienka@gfz-potsdam.de

previous earthquakes. The choice of the equation design, that is, which physical aspects are reflected and how, depends largely on the application purpose of the model, resulting in a vast number of proposed models over the last decades (Douglas, 2003). Although the amount of available strong motion observations and, therefore, also the complexity of GMMs have increased significantly since the early stages of ground-motion modeling (e.g. Esteva and Rosenblueth, 1964; Trifunac, 1976), associated uncertainties have remained stable during the past 50 yr (Strasser et al., 2009; Douglas and Edwards, 2016), indicating that the development of GMMs might benefit from the complementary incorporation of novel conceptual and methodological modeling paradigms.

One of the new conceptual model paradigms in recent years is the waiving of the ergodic assumption, which was previously a fundamental feature of GMMs. The ergodic assumption states that the distribution of ground-motion intensity values observed at a single location over time converges to the same distribution of values sampled across multiple locations (Anderson and Brune, 1999). This assumption was an indispensable feature of GMMs in the past when strong motion observations were scarce. However, in recent years, the amount of ground-motion data and computational power increased to levels that allow for the development of more regional and (partially) nonergodic models. Landwehr et al. (2016), for example, suggest a model with spatially varying coefficients for California that considers the absolute locations of earthquake source and observation sites as model parameters, leading to a significant decrease in aleatory uncertainty compared to the previous ergodic models. Following Stafford (2014), a similar strategy is followed by Kotha et al. (2016, 2020), who consider region-specific properties of ground motion as a random effect in a mixed-effects regression and, consequently, provide partially nonergodic GMMs for Europe. Recently, the problem of modeling spatial correlation of ground-motion intensity within a nonergodic framework was addressed by Kuehn and Abrahamson (2020).

On the methodological side, another trend has arisen across many fields in recent years: fully data-driven modeling through artificial neural networks (ANNs). ANNs have successfully been applied to ground-motion modeling in numerous studies. Derras et al. (2012, 2014) developed GMMs for shallow crustal earthquakes in Japan and Europe, Pozos-Estrada et al. (2014) modeled inslab and interplate earthquakes in Mexico, and Dhanya and Raghukanth (2018) derived a GMM from the Pacific Earthquake Engineering Research Center–Next Generation Attenuation-West2 Project database (Ancheta et al., 2014). The big advantage of ANNs over model-based approaches is that no a priori definition of a functional form is required. Instead, ANNs autonomously learn the relations between predictive parameters and the target ground-motion intensity measures (IMs) from ground-motion observations. Even though subjective decisions about the specific design of an ANN may have an impact on the results, the independence from a user-defined functional form opens up entirely new opportunities to reveal previously unknown relations and compare the predictive power of descriptive parameters objectively (e.g., Derras et al., 2012). Although ANNs have shown to provide excellent predictions within the data range that was used to train them, it is also clear that predictions outside this range are highly uncertain due to the lack of physical constraints that would allow such extrapolation. This becomes particularly relevant regarding the persistent lack of strong motion observations of large earthquakes at short distances (e.g., Kong et al., 2019).

Although considerable advances in both nonergodic and fully data-driven ground-motion modeling have been made in recent years, there was no attempt to merge these two, so far isolated, concepts within the ground-motion modeling community. This study aims at the assessment of opportunities and limitations of a specific ANN architecture, the U-Net (Ronneberger et al., 2015), for the task of ground-motion modeling. The U-Net is a well-established architecture that bears the potential to develop GMMs that are both fully nonergodic and fully data-

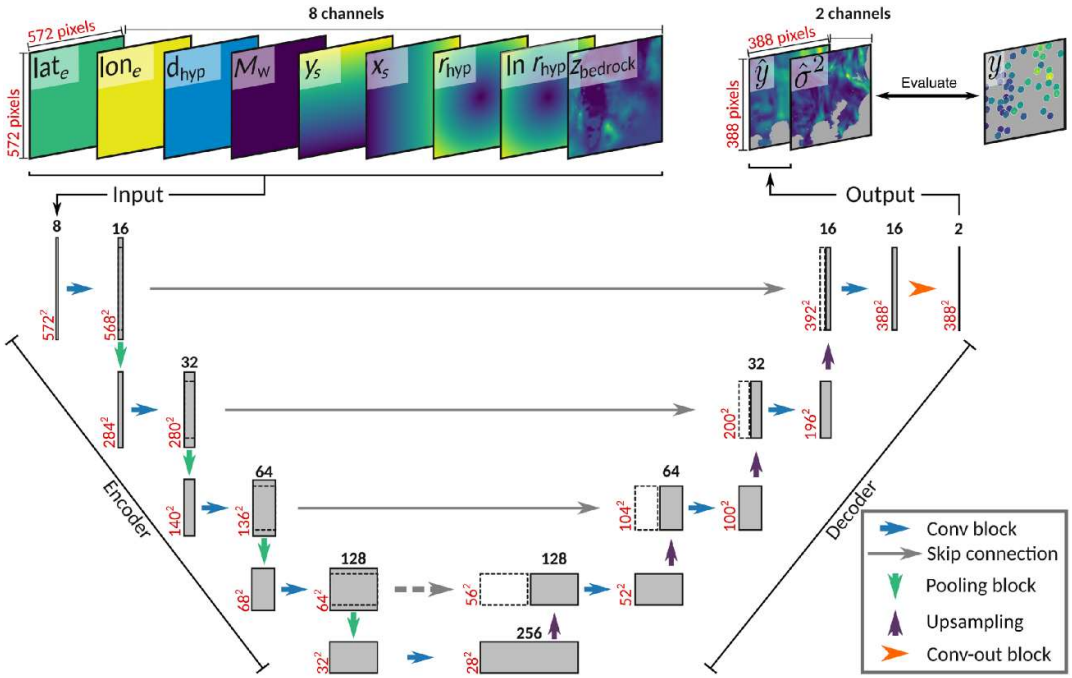

Figure 1. *The U-Net architecture used throughout this study. The predictive parameters for a single earthquake are provided to the U-Net as input in the form of a stack of maps. The input is processed through the encoder and decoder branches of the U-Net, until eventually the mean $\hat{y}$ and variance $\hat{\sigma}^2_{\epsilon,n}$ estimators of the intensity measure (IM) are provided as output. In the training phase, the output is evaluated against sparse measurements y of the IM that are available for this specific earthquake. Bold black numbers indicate the number of features at the respective locations within the U-Net. Red numbers indicate the resolution of the features. The depicted configuration of the input corresponds to the setup used for training with data from the Kanto basin (see the Application to the Kanto basin section). The latitude, longitude, and depth of the event hypocenter are denoted as lat_e, lon_e, d_{hyp} respectively, M_w denotes the moment magnitude, x_s and y_s denote the coordinates of each pixel in the input layer, r_{hyp} is the hypocentral distance, and $z_{bedrock}$ denotes the depth to seismic bedrock. A more detailed explanation of this figure is provided in the Appendix. This figure is based on figure 1 in Ronneberger et al. (2015).*

driven. The U-Net was initially developed for image processing tasks such as identifying and differentiating individual cells within tissue specimens, and has since been adopted to a vast range of applications, such as the prediction of the strength of wireless communication signals across maps (Levie et al., 2020), and pansharpening of satellite images (Yao et al., 2018). It is the inherent ability to process data in the form of 2D arrays (maps) that makes this architecture especially interesting for ground-motion modeling too, because it offers the opportunity to natively operate on map data and therefore preserves the actual spatial distribution of ground-motion observations and seamlessly links it to other geospatial information, for example, lithology and subsurface velocity structure.

We present our study as follows: First, we introduce the methodological adaption of the U-Net architecture to the task of ground-motion prediction. Subsequently, we present a proof of concept using a synthetic dataset to demonstrate the functionality of our model. We then show an exemplary application of our method to a subset of the Kiban–Kyoshin (KiK-net) (National Research Institute for Earth Science and Disaster Resilience, 2019) strong motion dataset in the Kanto basin in Japan and discuss our findings in comparison to local studies.

METHODOLOGY

The U-Net neural network architecture (Ronneberger et al., 2015) is the key component of our proposed GMM. Similar to all supervised machine learning methods, it learns the relation between predictive parameters and target parameters from a large number of examples that are provided to train the neural network. In ground-motion modeling, the target parameter is a ground-motion IM, which can be inferred from predictive parameters such as the moment magnitude M_w and the hypocentral distance r_{hyp}.

For a more detailed overview on neural networks, we refer to the review of Lecun et al. (2015).

The U-Net architecture is of particular interest for the task of ground-motion modeling, because it is designed to process data that comes in the shape of 2D numerical arrays, that is, maps. Representing ground-motion data as maps naturally preserves the absolute locations of observations and their relative position to each other—information that is required to develop a fully non-ergodic GMM. Apart from its technical suitability, the U-Net has already been approved in a similar application—the pathloss prediction of wireless telecommunication signals (Levie et al., 2020).

The functionality of a U-Net operating as a GMM is depicted in Figure 1. The predictive parameters for a single earthquake event represent the input features to the U-Net and are provided in the form of a stack of maps covering a predefined area. The input is then processed through all the layers of the U-Net until eventually two output maps are generated. Repeated convolution of the input maps with filter masks is the main operation within the U-Net, causing boundary value loss. The output maps are, therefore, slightly smaller than the input maps and cover a smaller region. Consequently, we distinguish between the U-Net input area and output area (the latter being the actual area of interest for which IM predictions are obtained) from here on. More details on this circumstance and a more detailed description of the implemented U-Net architecture are given in the Appendix.

If the U-Net is in an untrained state, outputs are just randomly generated maps, because the U-Net just consists of a number of randomly initialized coefficients and has not learned the relation between the predictive parameters and the target IM yet. However, we want the U-Net to provide estimates of the mean $\hat{y}$ and variance $\hat{\sigma}^2$ of the target IM, and therefore need to train the U-Net on a large training dataset that contains examples of event predictive parameters and corresponding IM observations. During training, the loss L_e between the U-Net outputs and the observations y_e for an event e with N_{obs} observations is evaluated as the negative log likelihood of the normal distribution:

$$L_e = \sum_{n=1}^{N_{obs}} \ln \hat{\sigma}_{e,n}^2 + \frac{\left(y_{e,n} - \hat{y}_{e,n}\right)^2}{\hat{\sigma}_{e,n}^2}. \quad (1)$$

The U-Net learns iteratively how to relate the input predictive parameters to the IM observations, that is, the loss is iteratively minimized through the gradient descent method Adam (Kingma and Ba, 2015), with the gradient efficiently implemented by backpropagation (Rumelhart et al., 1986). By the choice of the loss function and the optimization routine, training of the U-Net is equivalent to nonlinear least-squares regression, as it is commonly used in the most mixed-effects models. After each epoch of training, that is, one episode of training during which the U-Net sees all data examples in the training set, its ability to generalize to previously unseen events in a second dataset, the validation set, is evaluated. After a number of epochs, the loss on the validation dataset does not decrease any further, so the training is assumed to be complete, and the U-Net can henceforth be used as a GMM. The exact technical configuration of the training procedure followed throughout this study is given in the supplemental material S1, available to this article.

Although the observations y are point-wise measurements, the U-Net predictions $\hat{y}$ and $\hat{\sigma}^2$ are continuous maps. This means first that the loss function is only evaluated at those locations where actual observations are available, and second that the U-Net automatically interpolates the learned attenuation relation from the observation locations across the U-Net output area. The quality of this interpolation is examined in more detail in the following sections.

A clean separation between training and validation data could be achieved by splitting a strong motion dataset strictly according to both events and sites with seismic stations. However, huge portions of the dataset would be lost, because records of "training events" on "validation stations" and vice versa would be discarded. To address this issue, we developed the following training strategy:

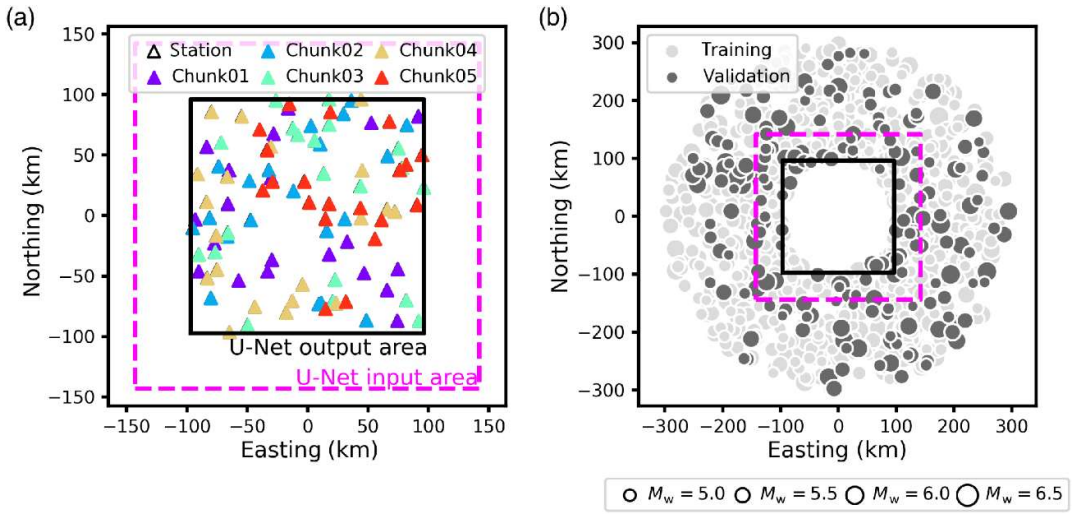

Figure 2. *Data generation setup for the synthetic experiments. (a) The geometric setup of the area of interest and the locations of randomly sampled stations. (b) The spatial distribution of the synthetic earthquake catalog with respect to the area of interest.*

---

1. The data are separated into training and validation events.
2. Stations (sites where seismic stations are located) are randomly grouped into a number Nchu of "chunks."
3. One U-Net is trained per station chunk, using the respective station chunk and validation events for validation and all other station chunks and training events for training.
4. Training of each U-Net is conducted N_{init} times to sample the variability that is caused by the random initialization of coefficients in the U-Net prior to training.

The final predictions for the mean and variance of the target IM ($\hat{Y}_e$ and $\hat{\Sigma}_e^2$, respectively) for the event e are subsequently obtained via ensemble averaging the mean and variance predictions of the separate U-Nets ($\hat{y}_e^{ij}$ and $\hat{\sigma}_e^{ij^2}$, respectively). The ensemble prediction is then a mixture distribution, the mean and variance of which are obtained based on the law of total expectation and the law of total variance, respectively (Blitzstein and Hwang, 2014):

$$\hat{Y}_e = \frac{1}{N_U} \sum_{i=1}^{N_{chu}} \sum_{j=1}^{N_{init}} \hat{y}_e^{ij}, \tag{2}$$

$$\hat{\Sigma}_e^2 = \frac{1}{N_U} \sum_{i=1}^{N_{chu}} \sum_{j=1}^{N_{init}} [\hat{\sigma}_e^{ij^2} + \hat{y}_e^{ij^2}] - \hat{Y}_e^2, \tag{3}$$

in which $N_U = N_{chu} N_{init}$ equals the total number of U-Nets. We note that the suggested procedure is not a classic example for the separation of training and validation data. However, we consider this adoption to the particular challenges of strong motion datasets an appropriate compromise between methodological accuracy and optimal data usage.

PROOF OF CONCEPT WITH SYNTHETIC DATA

Before we apply the U-Net methodology to derive a fully nonergodic GMM from real data, we demonstrate its abilities to:

- Learn the relationship between an IM and predictive parameters (such as M_w and r_{hyp}).
- Learn site-, source-, and path-specific amplification from site and event coordinates.
- Perform reasonable interpolation between observation sites.

We will present the case of a synthetic dataset including measurements of an IM that depends on hypocentral distance, magnitude, and site-specific amplification. Similar studies for datasets containing source- and path-specific amplification are conducted alike and are provided in the supplemental material S2.

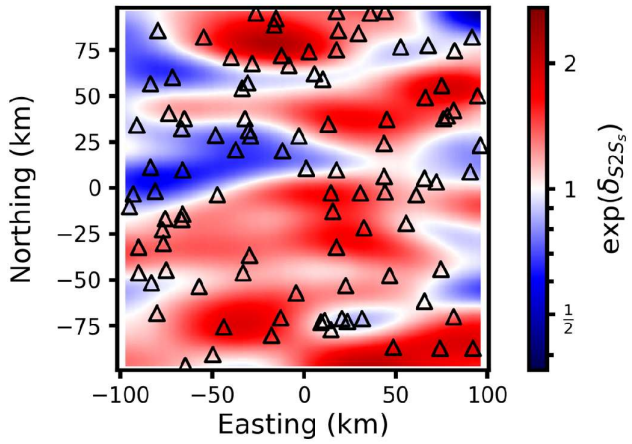

Figure 3. *The predefined spatially correlated random field representing exp δ_{S2S_s} used in the synthetic experiments. The black triangles indicate the locations of seismic stations.*

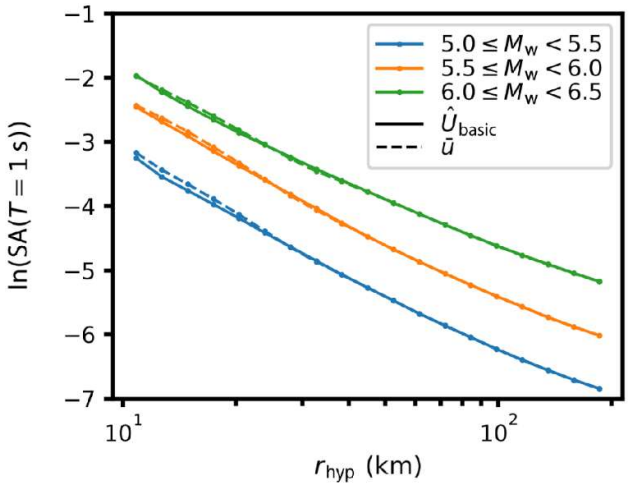

Figure 4. *Comparison of the predefined magnitude–distance scaling $\bar{u}(M_w, r_{hyp})$ (dashed lines) and the estimated relation $\hat{U}_{basic}(M_w, r_{hyp})$ (solid lines). The lines are derived via binning and averaging ln(SA(T = 1 s)) predefined and predicted values for the validation events according to r_{hyp} and M_w. For this comparison, the entire maps of $\hat{U}_{basic}$ and corresponding $\bar{u}$ values are used.*

Dataset generation

According to the input and output resolutions of the U-Net architecture (see the Appendix), we define two regular grids of sites for the U-Net input area (572 × 572 pixels) and the U-Net output area (388 × 388 pixels). We choose a grid spacing of 500 m, resulting in extents of 286 and 194 km^2 for the input and output areas, respectively (Fig. 2a). We then randomly define $N_S = 100$ sites from the output area as observation locations with seismic stations. We simulate $N_e = 1000$ events with uniformly distributed moment magnitudes $5.0 \leq M_w \leq 6.5$. The spatial distribution of events with respect to the U-Net input and output areas is presented in Figure 2b.

We use the GMM by Zhao, Zhou, et al. (2016; hereafter, Zhao16ASC) as the basis to simulate values of the natural logarithm of the acceleration response spectrum at period $T = 1$ s (ln(SA($T = 1$ s))) for all pairs of events and sites on the regular grid within the U-Net output area. Thereby, we only consider the scalings with M_w and r_{hyp}, and keep all other predictive parameters required by Zhao16ASC fixed. The synthetic dataset of ground-motion observations u follows from

$$u = \bar{u}(M_w, r_{hyp}) + \delta_{S2S_s} + \varepsilon. \quad (4)$$

Here, $\bar{u}$ describes the mean prediction of Zhao16ASC, δ_{S2S_s} is a predefined site-specific amplification function, and ε is a zero-mean Gaussian with a standard deviation of 0.1 that accounts for aleatory uncertainty. The term δ_{S2S_s} is defined as the portion of the site-specific amplification that is not captured by predictive parameters such as the shear-wave velocity of the upper 30 m of the Earth's crust (V_{S30}) and can only be obtained from repeated measurements at a site (Al Atik et al., 2010). To model δ_{S2S_s}, we first initialize an uncorrelated Gaussian random field and subsequently smooth it with an anisotropic Gaussian kernel with zero mean and covariance:

$$C = \begin{pmatrix} \sigma_E^2 & 0 \\ 0 & \sigma_N^2 \end{pmatrix}, \quad (5)$$

in which the standard deviations σ_E and σ_N scale the correlation in east–west and north–south directions, respectively. The resulting δ_{S2S} field is rescaled to a standard deviation of 0.2 and presented in Figure 3. The choice of a Gaussian kernel function agrees with various semivariogram analysis studies that investigate correlation of ground-motion features (e.g., Jayaram and Baker, 2009; Loth and Baker, 2013; Markhvida et al., 2018). The choice of values of $\sigma_E = 26$ km and $\sigma_N = 8$ km agree with the findings of Sgobba et al. (2021), who report an isotropic range of correlation (the distance at which only 5% correlation remain, approximately twice the standard

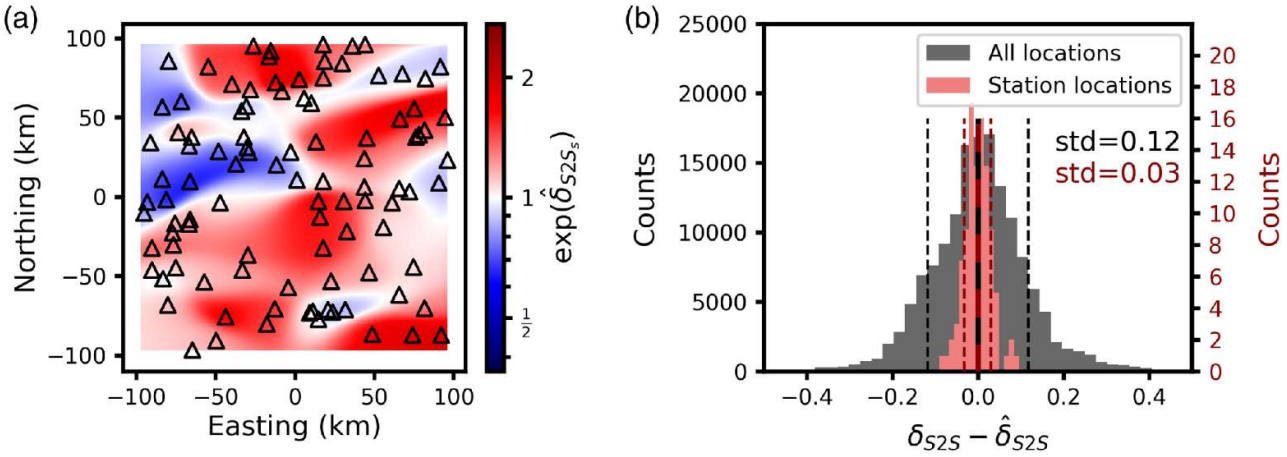

Figure 5. *(a) Estimated site amplification map $\hat{\delta}_{S2S_s}$ and (b) distributions of residuals $\delta_{S2S_s} - \hat{\delta}_{S2S_s}$ with respect to the predefined site amplification. Solid and dashed vertical lines in (b) indicate distribution means and standard deviations (std), respectively. From the visual comparison of (a) to Figure 3 and the distribution of residuals in (b) one can see that the site amplification is accurately retrieved at station locations. Predictions in interpolated areas are less accurate, but seem reasonable from visual inspection and appear to be unbiased.*

deviation) of 25 km for δ_{S2S}. Although radial isotropy, that is, $\sigma_E = \sigma_N$, is commonly assumed in ground-motion modeling for reasons of model simplicity, we choose anisotropic δ_{S2S}, because it is more realistic and to test whether the U-Net can capture such features.

The final dataset does not contain all simulated IM values on the regular grid. Instead, 25 observations are randomly selected from the 100 station locations. The synthetic dataset, therefore, consists of 25,000 records from 1000 events recorded at 100 stations.

U-Net training (synthetic)

We generally follow the training procedure described in the Methodology section. The synthetic dataset is randomly split into $N_e^{tr} = 800$ training and $N_e^{val} = 200$ validation events. The 100 stations are separated into $N_{chu} = 5$ chunks, and we consider $N_{init} = 5$ different random initializations per chunk. The derived GMM:

$$\hat{U} = f\left(M_w, r_{hyp}, \ln\left(r_{hyp}\right), x_s, y_s\right), \tag{6}$$

is an ensemble of 5 chunks × 5 initializations = 25 U-Nets. Parameters x_s and y_s denote coordinates of each grid point in the U-Net input area. We emphasize that we use no predictive parameter related to δ_{S2S_s}, which can subsequently only be learned as a function of x_s and y_s.

Evaluation of U-Net predictions (synthetic)

In a first step, we need to verify that our GMM has learned the fundamental scaling relations of u with M_w and r_{hyp}. Because of the abstract representation of information inside the U-Net, we cannot disaggregate the individual scalings with the separate parameters. However, the learned scaling $\hat{U}_{basic}$ of u with M_w and r_{hy} can be approximated by:

$$\hat{U}_{basic}\left(M_w, r_{hyp}\right) = \hat{U} - \hat{\delta}_{S2S_s} \tag{7}$$

in which $\hat{\delta}_{S2S_s}$ is the approximated site amplification

$$\hat{\delta}_{S2S_s} = \frac{1}{N_e^{val}} \sum_{e=1}^{N_e^{val}} \hat{U}_e - \bar{u}_e. \tag{8}$$

In Figure 4, we present the comparison between the predefined relationship $\bar{u}\left(M_w, r_{hyp}\right)$ and the estimated relationship $\hat{U}_{basic}\left(M_w, r_{hy}\right)$ approximated from predictions for validation events. We observe an overall good agreement indicating that the attenuation relation has been learned successfully.

In a second step, we compare the estimated site amplification $\hat{\delta}_{S2S_s}$ to the predefined δ_{S2S_s} to assess whether the site amplification was learned successfully at the station locations and

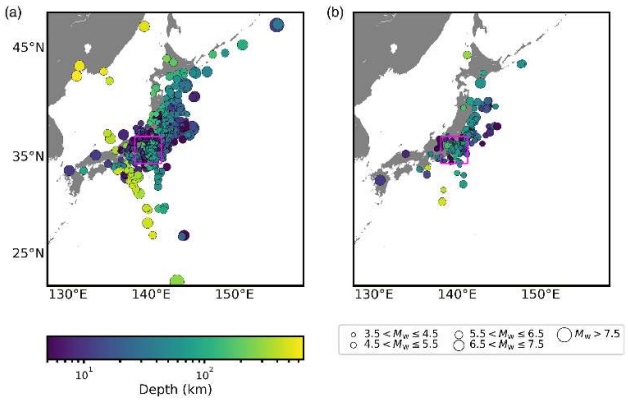

Figure 6. *(a) Training and (b) validation events selected from the Bahrampouri et al. (2021) database. The magenta square indicates the U-Net input area.*

interpolated with acceptable precision across the U-Net output area. Visual inspection of the learned site effect $\hat{\delta}_{S2S_s}$ (Fig. 5a) and comparison to the predefined site effect (Fig. 3) indicate a high similarity.

From the distribution of residuals between the predefined and the estimated site amplification depicted in Figure 5b, one can see that at those locations where observations were made, the learned site amplification closely resembles the predefined one, whereas the estimations are less precise, though unbiased, in interpolated areas. For comparison purposes, we also extract the predefined δ_{S2S_s} at station locations and perform bicubic spline interpolation to reconstruct the site amplification field. The standard deviation between this reproduction and the predefined field amounts to 0.11, very similar to the value of 0.12 that is obtained from the U-Net reproduction $\hat{\delta}_{S2S_s}$.We note that although the spatial anisotropy introduced during generation of δ_{S2S_s} clearly emerges also in the interpolated site amplification $\hat{\delta}_{S2S_s}$, no advantage over the bicubic spline interpolation can be stated in terms of misfit to the predefined δ_{S2S_s}.

Similar studies concerning the recovery of source-location-specific variations δ_{L2L_l} and path-specific amplification $\delta_{P2P_{sl}}$ have been conducted successfully and are presented in the supplemental material S2. In summary, we conclude that the proposed U-Net methodology is capable of extracting and interpolating the scaling of an IM with magnitude and distance, as well as site-, source-, and path-specific amplification from a strong motion dataset.

APPLICATION TO THE KANTO BASIN

In this section, we present an exemplary application of our U-Net GMM to the Kanto basin area, Japan. We will first describe the used dataset and the U-Net training procedure, followed by the discussion of an example prediction and a more general evaluation of the model performance on the entire dataset.

Kanto basin dataset

We chose the Kanto basin as our study area due to its high seismic activity and high station density of the regional KiK-net strong motion network. Operating for over 20 yr, KiK-net provides a substantial strong motion dataset necessary to constrain the U-Net methodology. We use the strong motion database by Bahrampouri et al. (2021) and select a subset of records from the Kanto basin, in total 46,191 records of $N_e = 2864$ events (Fig. 6) recorded at 65 stations (Fig. 7). Thereby, we use records from events of all tectonic region types. The average interstation distance of the selected KiK-net stations is about 94

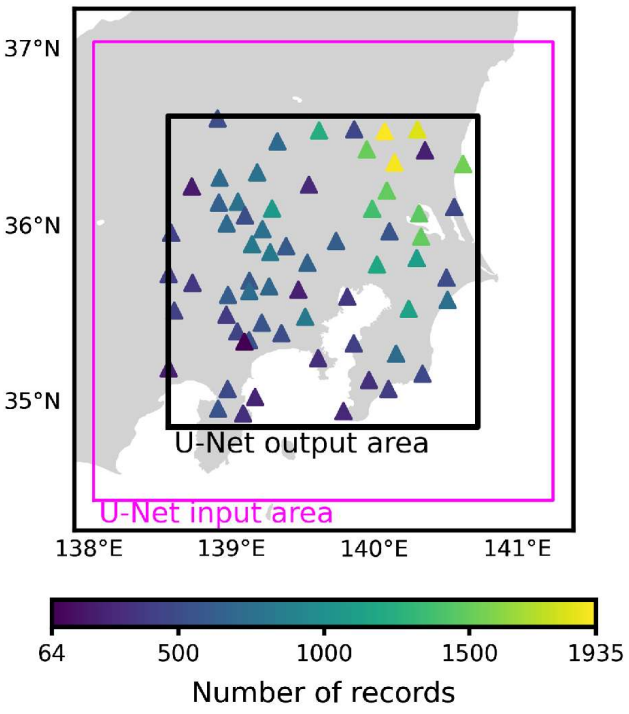

Figure 7. *Station coverage of the KiK-net network in the Kanto basin. Triangles indicate the locations of seismic stations. The color indicates the number of available strong motion records for ln(SA(T=1 s)) in the Bahrampouri et al. (2021) database.*

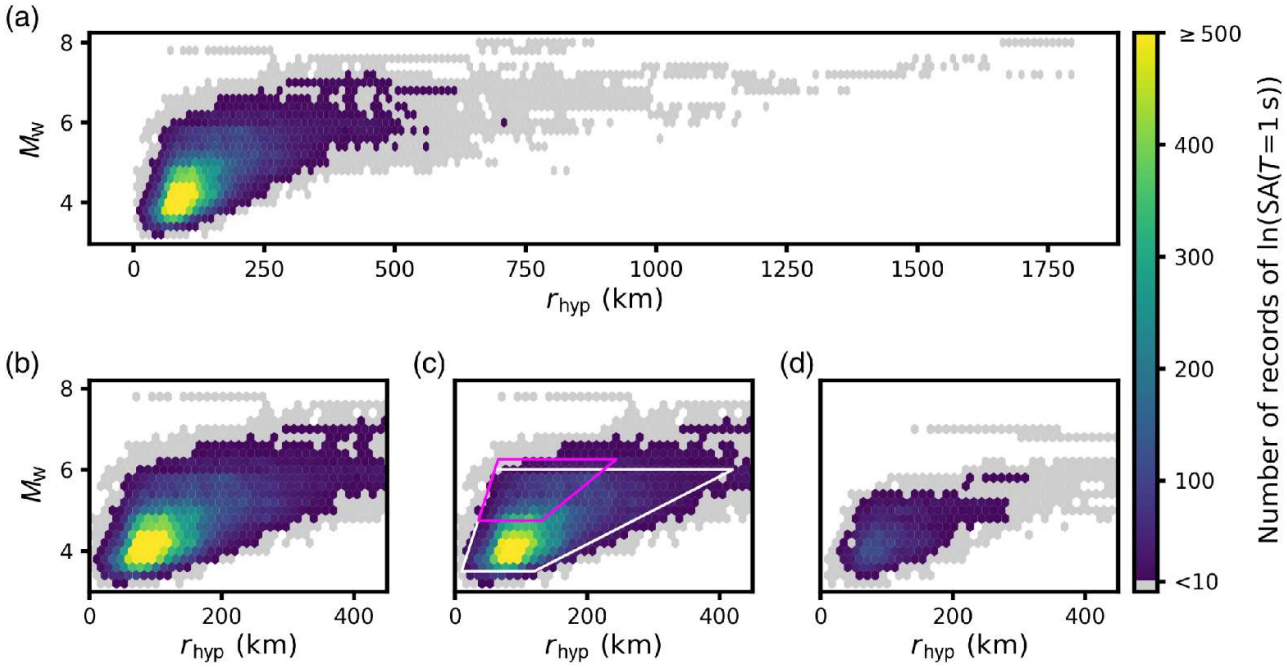

Figure 8. *Distribution of all KiK-net records from the Kanto basin according to (a,b) M_w and r_{hy} available in the Bahrampouri et al. (2021) database. The entire dataset is split into a set of (c) training events and (d) validation events. Although all available records were used for either training or validation, we assume that predictions from the ground-motion model (GMM) derived in this study for M_w, r_{hyp} combinations with less than 10 records in the training set are not reliable. The white polygon in (c) indicates our suggested range for application. The magenta polygon indicates the range for which comparison to conventional GMMs is presented in the Discussion section.*

km, with almost 20% of the interstation distances being less than 50 km. We select the geometric mean of the two horizontal components of ln(SA(T = 1 s)) as our target IM due to its wide application in probabilistic seismic hazard analysis. The distribution of the records with respect to M_w and r_{hy} is presented in Figure 8. The locations of the U-Net input (572 × 572 pixels) and output (388 × 388 pixels) areas in the Kanto basin are presented in Figure 7. The corresponding regular grids have a spacing of 500 m and encompass regions of 286 km² and 194 km², respectively.

We use a collection of nine predictive parameters: M_w, r_{hyp}, $\ln r_{hyp}$, event longitude lone, event latitude late, hypocentral depth d_{hyp}, coordinates of each grid point x_s, and y_s, and the depth to seismic bedrock $z_{bedrock}$ (compare Fig. 1). We expect the U-Net to learn site amplification as a function of x_s, y_s, and $z_{bedrock}$, source-location-specific variability from lone and late, and path-specific amplification from x_s, y_s, lon_e, and lat_e. The $z_{bedrock}$ data were downloaded from the J-SHIS map webservice (see Data and Resources).

U-Net training (Kanto basin)

We generally follow the training procedure described in the Methodology section. We split the dataset by events, in which events before the year 2015 are used for training and after for validation. The 65 stations are divided into N_{chu} = 5 chunks, and we consider N_{init} = 10 different random initializations per chunk. The arrangement of stations into chunks is illustrated in the supplemental material S3.

We derive 5 chunks × 10 initializations = 50 U-Net estimators:

$$\hat{y}^{ij} = f\left(M_w, r_{hyp}, \ln(r_{hyp}), lon_e, lat_e, d_{hyp}, x_s, y_s, z_{bedrock}\right)_{i \in \{1,2,\dots 10\}}^{i \in \{1,2,\dots 5\}} \quad (9)$$

We perform ensemble averaging over all 50 U-Nets $\hat{y}^{ij}$ according to equations (2) and (3) to derive the full ensemble estimator $\hat{Y}$ (compare Fig. 9) as our final GMM for the Kanto basin. Because $\hat{Y}$ is trained on data from all stations in the Kanto basin, it is not suited for evaluation of the quality of interpolated values. Therefore, we also derive partial ensemble estimators $\hat{Y}^i$, which only average over those respective 10 U-Nets that share the same ith station chunk for validation.

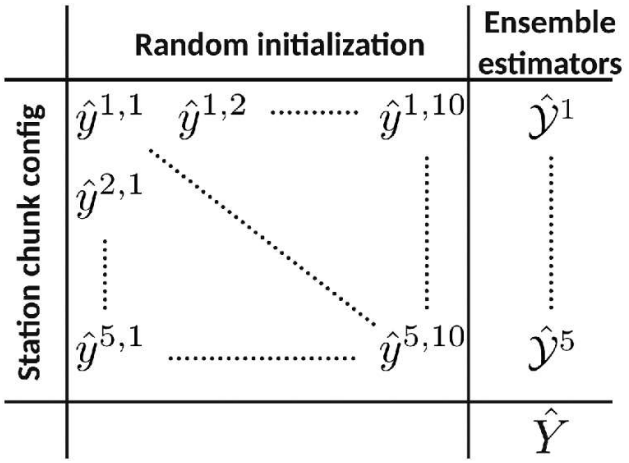

Figure 9. *Schematic representation of ensemble averaging of mean predictions. The full ensemble estimator $\hat{Y}$ is derived via averaging over all individual U-Net estimators $\hat{y}^{ij}$. Partial ensemble estimators $\hat{Y}^i$ are derived via averaging over those subsets of U-Nets that share the same station validation chunk. Partial ensemble estimators are required to evaluate predictions in interpolated regions of the area of interest. Ensemble averaging of individual standard deviation estimators $\hat{\sigma}^{ij}$ into partial ensembles $\hat{\varsigma}^i$ and the full ensemble $\hat{\Sigma}$ is performed accordingly.*

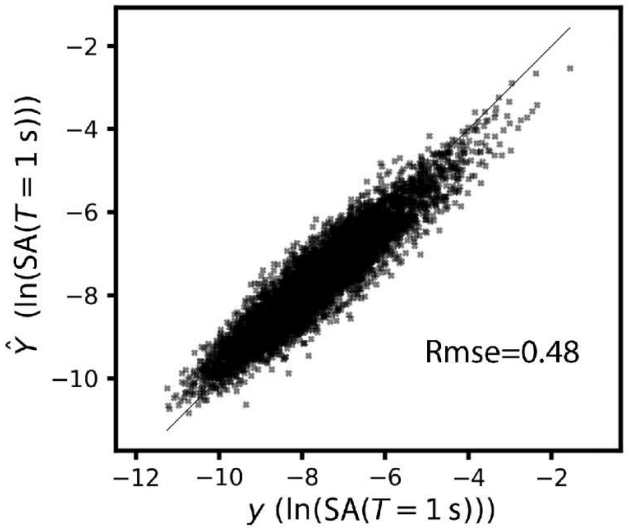

Figure 11. *Full ensemble mean predictions $\hat{Y}$ versus observations y of ln(SA(T = 1 s)) for all records from validation events.*

Evaluation of U-Net predictions (Kanto basin)

In Figure 10, we present an example prediction of our full ensemble estimator for an event from the validation set. A comparison between the mean prediction $\hat{Y}$ and the available observations y for this event is shown in Figure 10a. From visual inspection, we notice an overall satisfying agreement. The corresponding predicted standard deviation $\hat{\Sigma}$ is given in Figure 10b. One should recall that $\hat{\Sigma}$ depends on the same set of input variables as the mean prediction $\hat{Y}$ and is therefore a spatially varying and event specific quantity. One can identify the general trend that $\hat{\Sigma}$ is small in regions where station density is high and spatial variability in the mean prediction is low. This distribution indicates that the individual U-Net estimators $\hat{y}^{ij}$ learn the same scatter of observations at station locations but interpolate slightly differently due to their random initialization of coefficients. Consequently, $\hat{\Sigma}$ can, at least, be interpreted as a comprehensible measure for the relative confidence of the GMM in its prediction. More examples and comparisons to the predictions of the ergodic GMMs of Zhao, Zhou, et al. (2016), Zhao, Liang, et al. (2016), Zhao, Jiang,

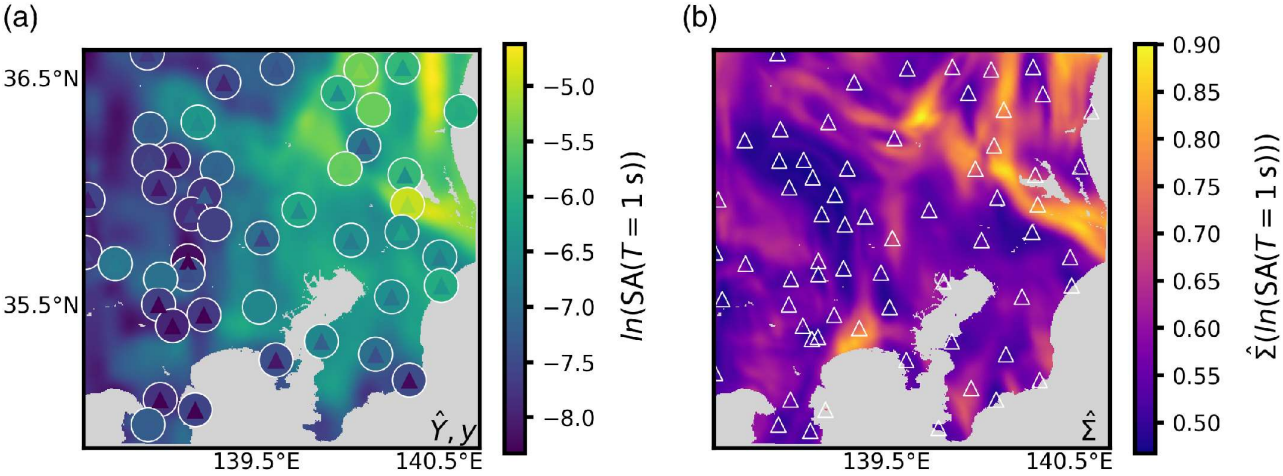

Figure 10. *Example prediction of our GMM for a magnitude 5.2 validation event at longitude 140.98° and latitude 13.86°. The plotted raster data and the circles in (a) show the mean prediction $\hat{Y}$, in which the circles represent the values at exactly those locations, where observations y (triangles) are available. The root mean square error (rmse) amounts to 0.41 in natural log units. The raster data in (b) shows the predicted standard deviation $\hat{\Sigma}$. The white triangles indicate the locations of all stations in the dataset, including those that have no record for this specific event.*

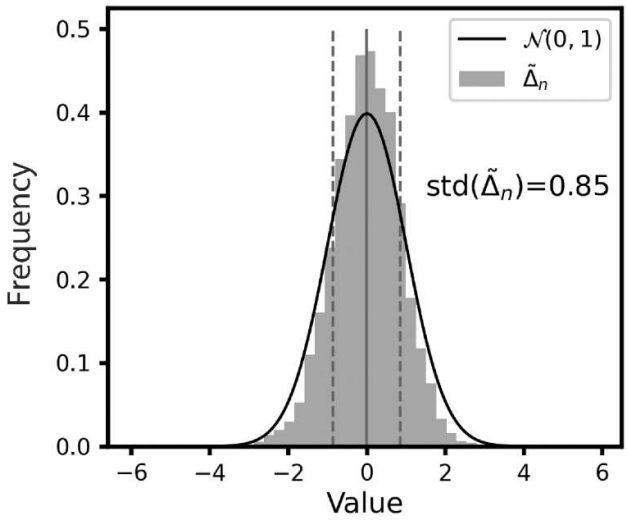

Figure 12. *Comparison of the distribution of standardized residuals $\tilde{\Delta}_n$ with the targeted standard Gaussian. Solid and dashed vertical lines indicate the empirical mean and standard deviation (std) of standardized residuals $\tilde{\Delta}_n$, respectively.*

TABLE 1
Average Root Mean Square Errors ± 1 Standard Deviation between Observations y and Predictions of the Five Partial Ensemble Estimators $\hat{\Upsilon}^i$

Configuration	Rmse
Training stations/training events	0.428 ± 0.005
Training stations/training events	0.454 ± 0.003
Validation stations/training events	0.809 ± 0.082
Validation stations/training events	0.821 ± 0.110

et al. (2016) are available in the supplemental material S4. The comparison of predictions $\hat{\Upsilon}$ with observations y from validation events is illustrated in Figure 11 and is quantified using the root mean square error ($rmse = 0.48$ in natural log units). The use of the negative log-likelihood loss function given in equation (1) provides that $\hat{\Sigma}$ should be a proxy for the scatter in observations at station locations. To verify the success of this strategy, we first calculate residuals between observations from validation events and the corresponding mean predictions of our full ensemble GMM. We then standardize these residuals by dividing each individual residual by its predicted standard deviation:

$$\tilde{\Delta}_n = \frac{y_n - \hat{Y}_n}{\hat{\Sigma}_n} n \in \{1, \ldots, N_{obs,val}\}, \qquad (10)$$

in which $N_{obs,val}$ denotes the number of observations from validation events. If $\hat{\Sigma}$ accurately described the scatter in observations, then the standardized residuals $\tilde{\Delta}_n$ should be standard normally distributed.

The actual distribution of $\tilde{\Delta}_n$, together with the targeted standard normal probability density function (PDF), is depicted in Figure 12. We notice that the actual standard deviation of 0.85 is ≈15% smaller than the targeted value of 1. Because $\hat{\Sigma}_n$ represents the denominator in equation (10), this indicates an overestimation of the uncertainty in predictions at station locations.

In addition to the performance at station locations, we also verify the model's ability to interpolate between stations. For this purpose, we use the partial ensemble estimators $\hat{\Upsilon}^i$, for which the ith station chunk was used for validation only and can thus be used to evaluate interpolated predictions. Each of the five $\hat{\Upsilon}^i$ can be evaluated in four different categories: (1) training events recorded on training stations, (2) validation

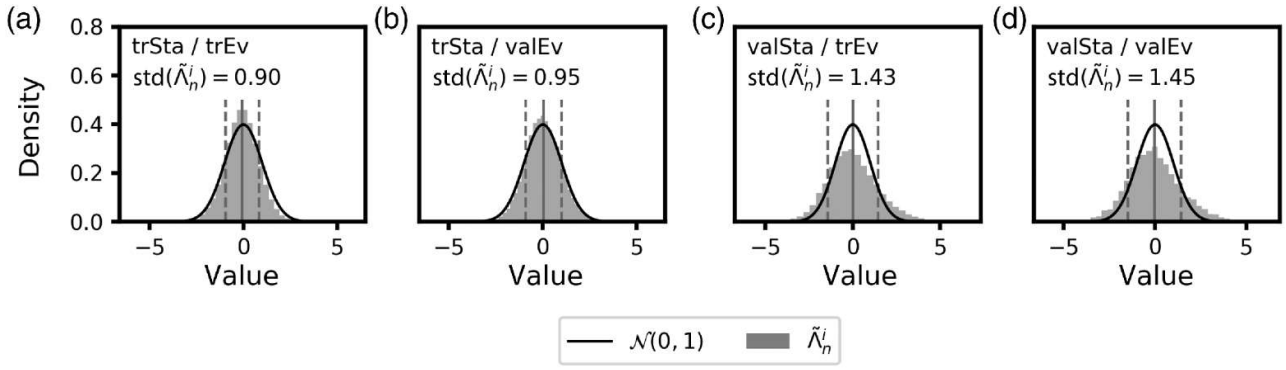

Figure 13. *Standardized residuals $\tilde{\Lambda}^i_n$ with respect to partial ensemble estimators $\hat{\Upsilon}^i$ at (a) training station locations for training events, (b) training stations for validation events, (c) validation stations for training events, and (d) validation stations for validation events. Solid and dashed vertical lines indicate the empirical mean and standard deviation (std) of standardized residuals $\tilde{\Lambda}^i_n$, respectively.*

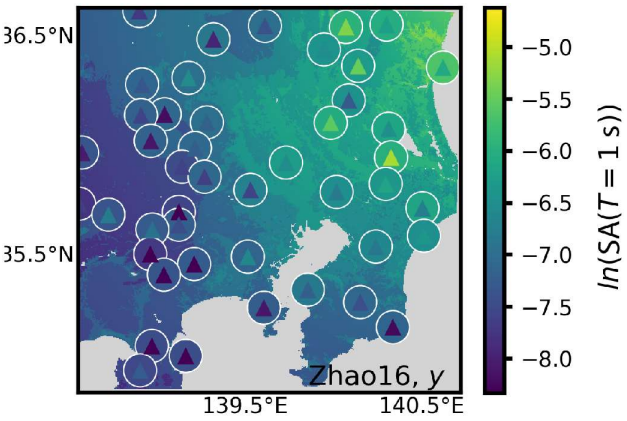

Figure 14. *The raster data and the circles show the mean IM prediction of the Zhao, Liang, et al. (2016) GMM for the same event presented in Figure 10. The circles represent the values at exactly those locations, where observations (triangles) are available. The rmse amounts to 0.65 in natural log units.*

events recorded on training stations, (3) training events recorded on validation stations, and (4) validation events recorded on validation stations. The *rmse*s between observations and predictions averaged over the five $\hat{\Upsilon}^i$ are presented in Table 1. As expected, the smallest *rmse* is obtained for those observations that were used during training. A similar value is obtained for records from validation events observed at training stations, indicating successful generalization to events outside the training dataset. The *rmse* increases significantly for validation stations, because these predictions include the additional error due to interpolation. The *rmse* at interpolated sites obtained from $\hat{\Upsilon}^i$ can be interpreted as an upper bound estimate of the corresponding *rmse* of the full ensemble estimator $\hat{\Upsilon}$, because it is trained using all stations and is, therefore, better constrained due to the higher station density.

In a final step, we evaluate the precision of the predicted standard deviations $\hat{\varsigma}^i$ from the partial ensemble estimators. As for the full ensemble estimator, we calculate standardized residuals between all observations from both training and validation events with respect to all partial ensemble estimators:

$$\tilde{\Lambda}_n = \frac{Y_n - \hat{\Upsilon}_n}{\hat{\varsigma}_n^i} n \in \{1, \dots, N_{obs}\}, \qquad (11)$$

in which N_{obs} denotes the number of observations. The comparison of the distribution of $\tilde{\Lambda}_n$ to the targeted standard normal PDF is depicted in Figure 13. From Figure 13a,b, we conclude that the predictive uncertainty at the training stations closely resembles the data scatter and, therefore, represents a reliable uncertainty measure. At interpolated sites, $\tilde{\Lambda}_n$ is zero centered and well described by a normal distribution (Fig. 13c,d). However, we observe that the standard deviation of $\tilde{\Lambda}_n$ is over 40% larger than the targeted value of 1, indicating a significant underestimation of the predicted standard deviation. This means that although the relative tendency to assign larger uncertainties to less constraint estimates is given, predictive uncertainty at interpolated sites is quantitatively only a lower bound uncertainty estimate. The precision in Figure 13b is actually higher than the one obtained for the full ensemble estimator presented in Figure 12. This is due to the fact that when averaging over all 50 U-Nets at a specific station site, those 10 U-Nets for which this site was in the validation chunk contribute less well-calibrated predictions.

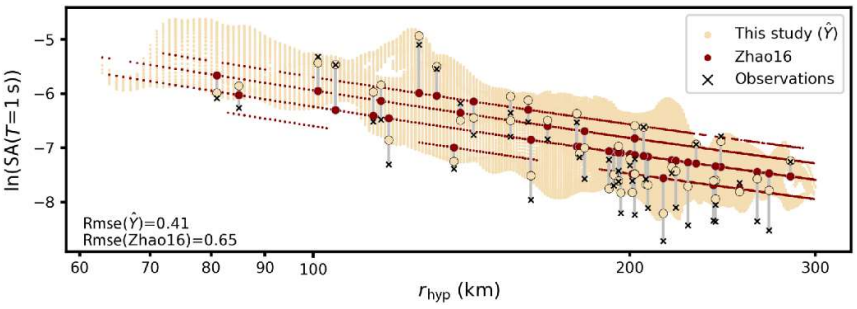

Figure 15. *Direct comparison between our model (yellow point cloud), the Zhao16 GMM (red lines), and observations (black crosses) for the event introduced in Figures 10 and 14. Because of the continuous site parameters used in our model, corresponding predictions appear as a point cloud, contrary to the four lines that represent predictions of the Zhao16 GMM for four different site classes. Gray lines indicate the affiliation of observations with corresponding Zhao16 predictions (large red markers) and $\hat{Y}$ predictions (large yellow markers).*

DISCUSSION

Comparison to ergodic GMMs

Because the presented U-Net methodology is a fully data-driven approach without any physical constraints on the

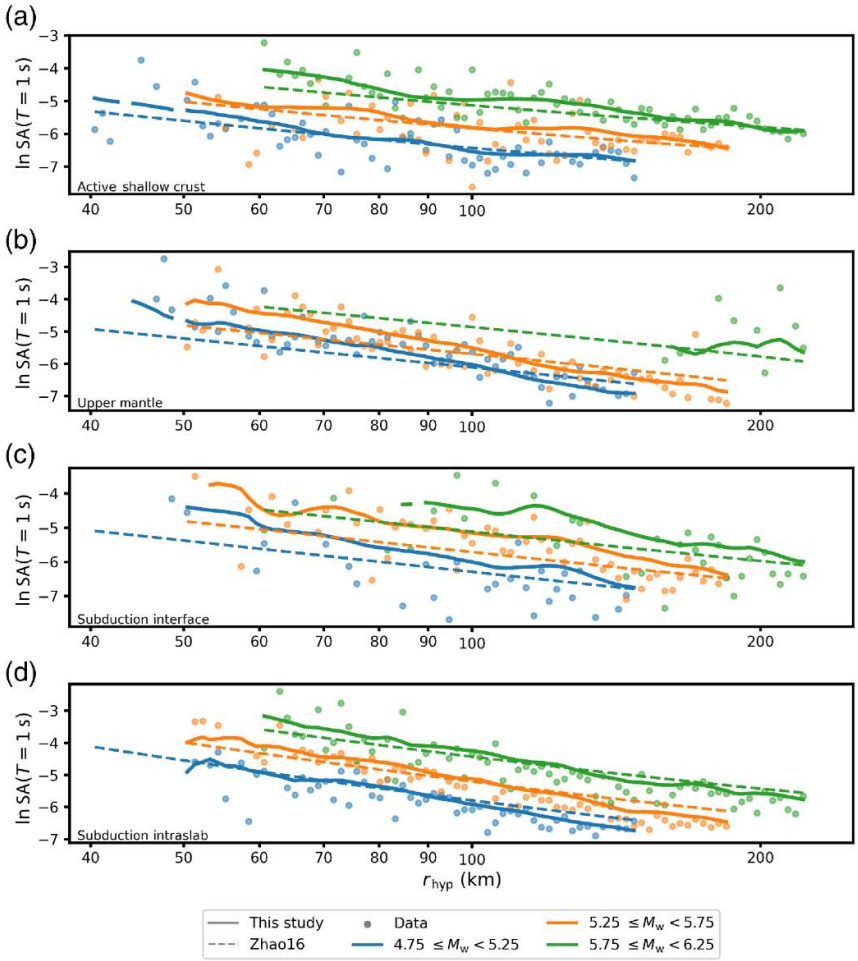

Figure 16. *Comparison between the GMM developed in this study (solid lines), observations from the Bahrampouri et al. (2021) database (dots), and Zhao16 GMMs (dashed lines), for (a) active shallow crust events, (b) upper mantle events, (c) subduction interface events, and (d) subduction intraslab events. The dots depicting the data and the curves depicting our model are derived from averaging suitable records and predictions of ln SA($T = 1$ s) along M_w and r_{hyp} bins for both training and validation events.*

model design, we want to validate that the fundamental functionality of our full ensemble estimator $\hat{Y}$ is comparable to conventional GMMs. We compare our prediction for the example event presented in Figure 10, as well as the overall learned magnitude and distance scalings to the ergodic GMMs for Japan developed by Zhao, Zhou, et al. (2016), Zhao, Liang, et al. (2016), Zhao, Jiang, et al. (2016) (hereafter, Zhao16 GMMs). We emphasize that the purpose of these comparisons is not to identify superiority of one model over the other. Instead, we want to demonstrate that the basic functionality of our model is comparable to a conventional GMM that is derived from a well-established and mature methodology. A detailed, quantitative comparison would be inappropriate for a number of reasons: first of all, there is no standardized procedure to compare nonergodic to ergodic GMMs, because they make fundamentally different assumptions on the distribution of ground-motion data. Furthermore, our model is trained on data from the Kanto basin only, whereas the Zhao16 models are calibrated using data from entire Japan. Another aspect is that in this study we do not exhaust the full power of the Zhao16 models, because we disregard its volcanic path terms and always use r_{hyp} instead of the suggested rupture plane distance r_{rup}, if 3D rupture models are available. Finally, we derive the site classes that are required as site parameters for the Zhao16 models indirectly via conversion of the V_{S30} map accessed via the J-SHIS web map service (see Data and Resources), according to Zhao et al. (2015). However, we consider this use of Zhao16 models appropriate for the qualitative comparison that we attempt.

The prediction of the Zhao16 GMM for the event presented in Figure 10 is given in Figure 14. In comparison to the prediction $\hat{Y}$ derived in this study (Fig. 10a), we note that the range of IM values and the large-scale attenuation with distance are similar. However, we also note that $\hat{Y}$ adapts better to the small-scale spatial variations in the observations of the IM for this specific event. The direct comparison between $\hat{Y}$ and the Zhao16 GMM is presented in Figure 15. Although our model appears as a point cloud and the Zhao16 model as discrete lines, due to the use of continuous and discrete site parameters, respectively, we note that the attenuation with distance of both models are comparable, with our model

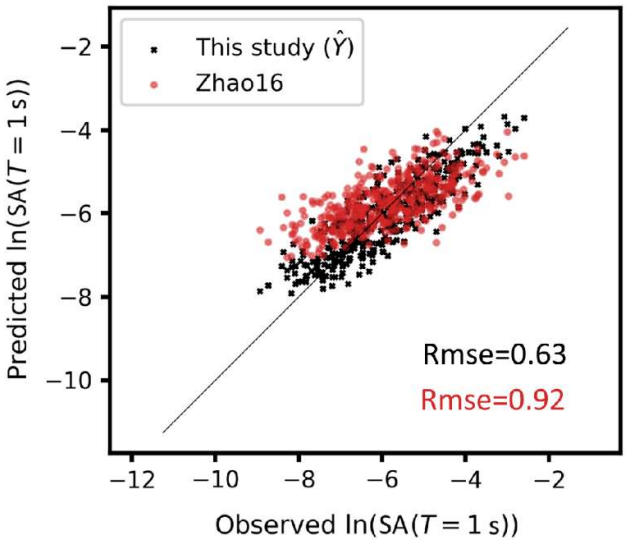

Figure 18. *Performance of our model and Zhao16 models on the fraction of observations from validation events suitable for comparison (compare Fig. 8).*

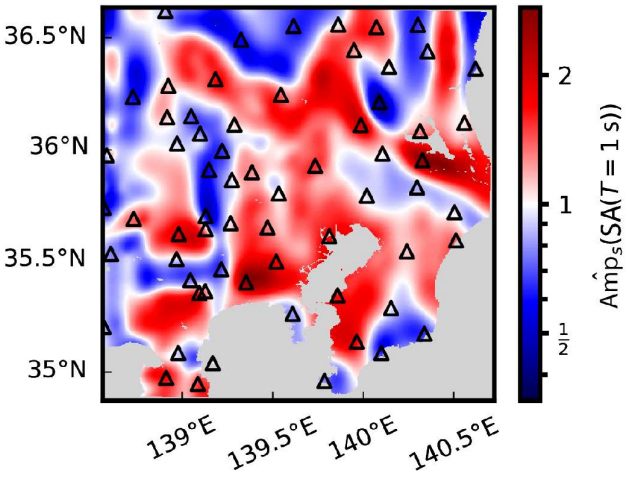

Figure 17. *Estimated site amplification of SA(T = 1 s) in the Kanto basin approximated from averaged mean predictions $\hat{Y}$.*

providing a closer fit to the observations.

In a next step, we compare the average magnitude–distance scaling learned by our model to that of the actual observations and Zhao16 GMMs. Because the individual scalings between input and target parameters cannot be disaggregated for a neural network, we need to approximate them from predictions. Only events and records within the data range that is valid for comparison in terms of magnitude, distance, and depth to top of rupture (z_{tor}) are selected for comparison. The suitable data range in terms of magnitude and distance is given in Figure 8c. According to Zhao, Zhou, et al. (2016), Zhao, Liang, et al. (2016), Zhao, Jiang, et al. (2016), limits of $z_{tor} < 25$ km for active shallow crust, $z_{tor} > 25$ km for upper mantle, $z_{tor} < 50$ km for subduction interface, and $z_{tor} < 200$ km for subduction intraslab events are selected. We average predictions of the selected training and validation events per tectonic region type along magnitude–distance bins. The results are depicted in Figure 16. Because the accuracy of this procedure depends highly on the number of predictions within a bin, we only consider bins with more than 500 predicted values. Predictions of our model in water covered areas are excluded. The dots representing the observations are derived accordingly, using the same bins. For the Zhao16 predictions, we use average magnitude and rake values obtained per tectonic region type and magnitude bin and $V_{S30} = 410$ m/s, which corresponds to site class SCII (hard soil) according to Zhao et al. (2015). More details on the configuration of the Zhao16 models are given in the supplemental material S5. For active shallow crust events, predictions of our model, Zhao16, and observations provide a consistent picture (Fig. 16a). Although magnitude scalings are similar between our model and Zhao16 for upper mantle events, distance scalings significantly differ (Fig. 16b). However, our model is closer to the observations than Zhao16, indicating that our model has learned a Kanto basin specific distance scaling that differs from the average Japanese one. For subduction interface events, scalings are slightly different between Zhao16 and our model (Fig. 16c). However, compared to the observations, our model provides reasonable results. An overall good agreement of the two models and observations can be stated for subduction intraslab events (Fig. 16d). We emphasize at this stage that the predictions of our model adapt to the tectonic region type of an event, although it was not used as a predictive parameter during training. We, therefore, conclude that the U-Net extracted the relevant information from the location of the event hypocenters. This feature is fundamentally different from common GMMs that rely on the accurate classification of events into tectonic region types. Based on the presented comparison, we conclude that our U-Net-based GMM provides predictions in

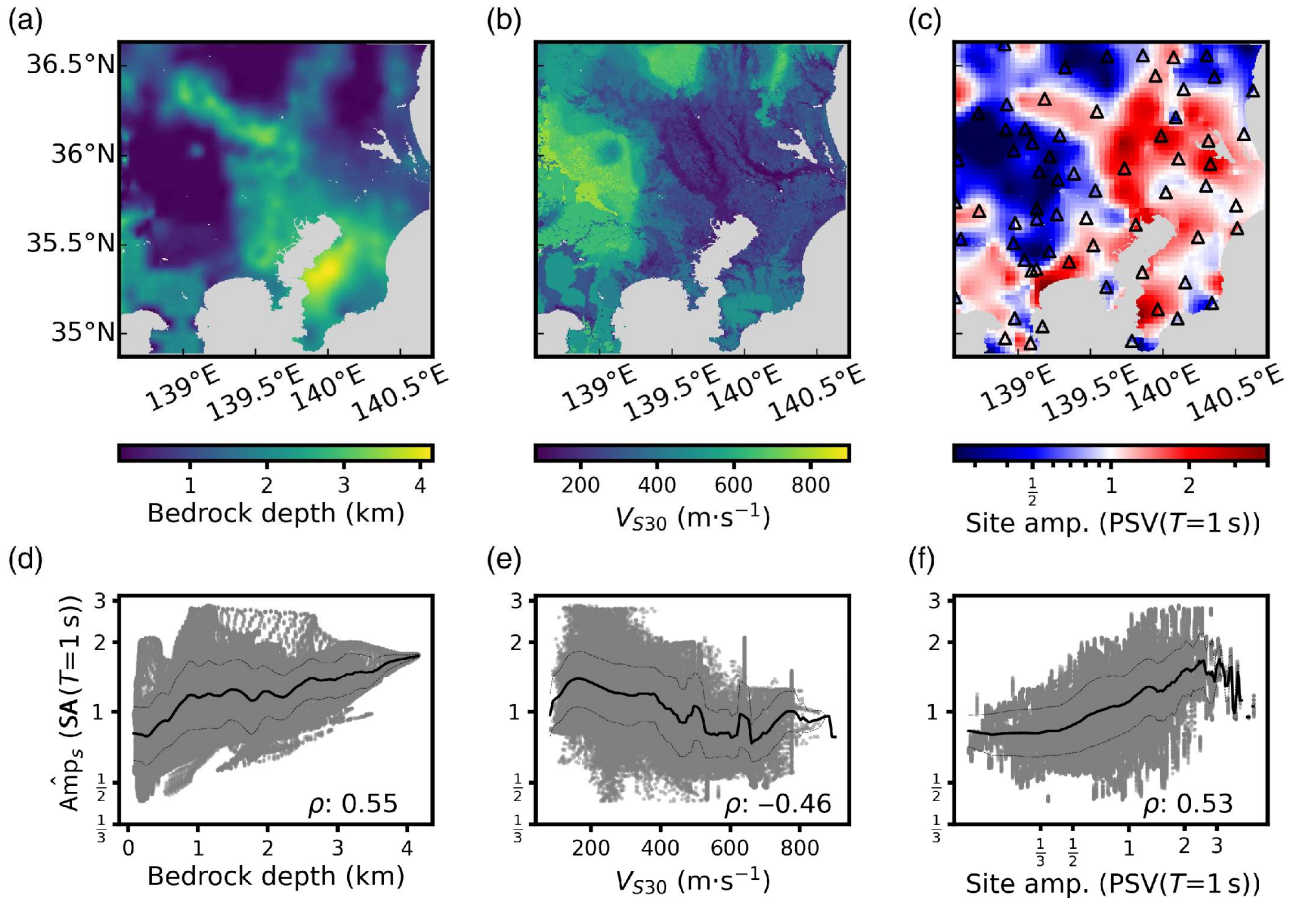

Figure 19. *Maps of (a) bedrock depth, (b) V_{S30}, and (c) site amplification of pseudospectral velocity at a period of 1 s (PSV(T = 1 s)) estimated from Nakano and Kawase (2021). The respective correlations with the approximated site amplification $A\hat{m}p_s$SA(T = 1 s)) are depicted in (d–f), quantified using the Spearman's ρ correlation coefficient. The bold and thin black lines in (d–f) represent the average and ± 1 standard deviation for the respective data bins.*

terms of magnitude and distance comparable to Zhao16 - a set of models based on a well-established and mature methodology.

In a final comparison, we evaluate the performance of our model and the Zhao16 models on those observations from validation events that are within the validity range of both the models (Fig. 17). The fact that the misfit of our model and Zhao16 are of the same order of magnitude indicate that the U-Net methodology is capable of providing reasonable GMMs.

Evaluation of learned site amplification

Our proof of concept demonstrated that our model is capable of providing fully nonergodic GMMs, including site-specific, as well as source-location and path-specific amplification. Although the latter two are difficult to evaluate on real data, because we cannot disaggregate the scaling relations within a neural network, we can approximate the average learned site amplification from the absolute site coordinates x_s, and y_s, and $z_{bedrock}$ by averaging predictions.

In a first step, we train another GMM $\hat{F}$ identically to $\hat{Y}$ but without using x_s, y_s, and $z_{bedrock}$ as predictive parameters. We then approximate the site amplification $A\hat{m}p_s$ learned by $\hat{Y}$ with

$$\ln A\hat{m}p_s = \frac{1}{N_e}\sum_{1}^{N_e} \hat{Y}_e - \hat{F}_e. \qquad (12)$$

The resulting site amplification map is depicted in Figure 18. We compare this map to the depth to bedrock map used for training, the commonly used site parameter V_{S30}, and site amplification of pseudospectral velocity at a period of 1 s (PSV(T = 1 s)) derived by Nakano and Kawase (2021) (Fig. 19). The V_{S30} map was downloaded from the J-SHIS map webservice (see Data and

Resources). The degrees of correlation are quantified using Spearman's ρ correlation coefficient.

From the correlation of $\widehat{Amp}_s$ with the depth to bedrock map (compare Fig. 19a,d), one can see that the model successfully learned a relation between depth to bedrock and the target IM. Because the outline where $\widehat{Amp}_s = 1$ systematically follows the outline of the Kanto basin (compare Fig. 19a), we conclude that this relation is successfully transferred from the observation locations across the Kanto basin. Furthermore, one can see small-scale anomalies around seismic station locations, indicating that the U-Net learned a site-specific amplification from the absolute locations of observations.

The approximated site amplification $\widehat{Amp}_s$ shows moderate correlation with the commonly used site parameter V_{S30} (Fig. 19b,e) and the site amplification of PSV($T = 1$ s) (Fig. 19c,f).

Good agreement between $\widehat{Amp}_s$ and amplification of PSV($T = 1$ s) can be stated at KiK-net sites in the central and eastern parts of the Kanto basin (compare Fig. 18). However, major differences are observable west of the Kanto basin, where the PSV($T = 1$ s) map consistently shows deamplifcation, and our map shows a more variable picture. Reasons for this difference might be the different data selection of Nakano and Kawase (2021). Furthermore, they used not only the KiK-net, but also K-net and Japan Meterological Agency (JMA) strong motion networks, and therefore provide a more detailed map. Using the same data density in our model might provide a more insightful comparison but is beyond the scope of this study.

Concerning practical application

All calculations related to this study were performed on a desktop computer with an Intel(R) Core(TM) i9-7980XE CPU (977 GFLOPS), 128 GB of RAM, and an NVIDIA RTX2080Ti GPU (13.45 TFLOPS). With this setup, training of a single U-Net takes approximately 1.5 hr. It means that the presented work can be transferred to other datasets and areas of interest with reasonable effort.

The major shortcoming of the U-Net GMM methodology at the current stage is its limited applicability to near source ground motion of large-magnitude events due to the lack of corresponding observations. Although all GMMs suffer from this problem, physical a priori information can be incorporated in equation-based GMMs to overcome this issue. An analog stabilization for the U-Net method could be the augmentation of strong motion datasets with predictions from ergodic GMMs. With such a strategy, the U-Net method would provide the same predictions as ergodic GMMs at the respective data ranges and provide nonergodic predictions where more observations are available. A similar solution is implemented for example by Landwehr et al. (2016), whose model converges toward an ergodic backbone model in regions where station density is low.

Unlike other commonly used interpolation approaches, such as Gaussian Process Regression (as used in Landwehr et al., 2016), the U-Net methodology provides data-driven spatial interpolation without the need to define the type and parameters of an interpolation kernel function. Although this functionality is convenient and grants some degree of freedom of subjective decisions by the modeler, it is yet an open question how its performance compares against conventional interpolation techniques.

Because of the open questions discussed earlier, we currently do not consider our presented methodology sufficiently mature for immediate application in seismic hazard and risk studies, yet. However, potential future research paths toward answering these questions are in sight. We are optimistic that U-Net-based fully data-driven and nonergodic GMMs can become a permanent, complementary element of hazard and risk studies in data-rich regions such as Japan, Taiwan, and California in the foreseeable future.

CONCLUSIONS

In this study, we have presented the development of a nonergodic, fully data-driven GMM prototype based on the U-Net neural network architecture. As such, the model is free of any a priori choices concerning the model design itself or the interpolation scheme.

From experiments with synthetic data we

found that the U-Net methodology is capable of learning observation site, earthquake location, and propagation path-specific amplification effects and is, thus, comparable to fully nonergodic approaches to ground-motion modeling. However, the U-Net's fully data-driven, inherent interpolation scheme significantly eases the use of such a model to predict continuous ground-motion maps, because no a priori assumption about the interpolation have to be made. In the synthetic experiments, we demonstrated that the interpolation scheme provides unbiased predictions within acceptable error bounds, whereas from the results of the application to real data from the Kanto basin we can state that interpolated predictions correlate with commonly used site parameters. However, a quantitative comparison to conventional interpolation methods is yet an open task.

From the application of our model to real observations from the Kanto basin, we have learned that ground shaking of future earthquakes can be predicted, within acceptable errorbounds, at those sites where observations from the previous earthquakes are available. We can furthermore state that the predictions of our model at observation sites and in interpolated regions agree on average with the predictions of equation-based, ergodic GMMs for Japan.

However, due to data scarcity, application of our model in the near field is currently not advised without augmenting the used strong motion dataset in this regard. Although associated predictive uncertainties were found to provide useful information about the relative confidence of the model in a prediction, we noticed that the absolute values of uncertainties are overestimated at observation sites and underestimated at interpolated sites, with respect to the data scatter.

The fact that our model can be derived on a consumer-level desktop computer with a decent GPU allows the transfer to an arbitrary area of interest with reasonable effort.

Summarizing the main findings from this study, we conclude that U-Net-based GMMs pose a worthwhile complementary tool to the already large and well-established pool of GMM methods. In our opinion, the presented results indicate that U-Net-based GMMs have the potential to become a permanent feature in seismic hazard and risk studies, complementary to conventional GMMs, in the foreseeable future.

DATA AND RESOURCES

The depth to seismic bedrock (V.3.2 ESRI shapefile "Subsurface Structure" layer 30; The Headquarters for Earthquake Research Promotion, 2021), as well as V_{S30} (ESRI shapefile "Site amplification factors," parameter "AVS"; Fujimoto and Midorikawa, 2006; Senna et al., 2013, 2019;Wakamatsu and Matsuoka, 2013, 2020) maps for the kanto basin were downloaded from the J-SHIS web map service available at https://www.j-shis.bosai.go.jp/map/?lang=en (last accessed November 2021). The Bahrampouri et al. (2021) strong motion flat file was downloaded from https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2547 (last accessed November 2021). We used NumPy (Harris et al., 2020) and SciPy (Virtanen et al., 2020). Figures were made with Matplotlib (Hunter, 2007) and Inkscape (Inkscape Project, 2020). Tensorflow (Abadi et al., 2016) was used for deep learning. We used QGis (QGIS Development Team, 2021) for data preparation. Ground-motion models (GMMs) of Zhao, Zhou, et al. (2016), Zhao, Liang, et al. (2016), and Zhao, Jiang, et al. (2016) were accessed via Openquake (Pagani et al., 2014). The supplemental material to this article encompasses the following content: S1, technical details on U-Net training; S2, documentation of complementary synthetic experiments; S3, a figure on the grouping of station chunks in the Kanto basin; S4, example predictions from the validation dataset and comparisons to Zhao16 models; and S5, technical details on the use of Zhao16 models.

DECLARATION OF COMPETING INTERESTS

The authors acknowledge that there are no conflicts of interest recorded.

ACKNOWLEDGMENTS

The first author acknowledges the support of the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS). The authors warmly thank Hiroshi Kawase and Kenichi Nakano for fruitful discussions and for willingly sharing their site amplification data.

REFERENCES

Abadi, M., P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, et al. (2016). TensorFlow: A system for large-scale machine learning, 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16), USENIX Association, 265–283.

Al Atik, L. A., N. Abrahamson, J. J. Bommer, F. Scherbaum, F. Cotton, and N. Kuehn (2010). The variability of ground-motion prediction models and its components, Seismol. Res. Lett. 81, no. 5,

794–801, doi: 10.1785/gssrl.81.5.794.

Ancheta, T. D., M. Eeri, R. B. Darragh,M. Eeri, J. P. Stewart, M. Eeri, E. Seyhan, M. Eeri,W. J. Silva,M. Eeri, et al. (2014). NGA-West2 database, Earthq. Spectra 30, no. 3, 989–1005, doi: 10.1193/070913EQS197M.

Anderson, J. G., and J. N. Brune (1999). Probabilistic seismic hazard analysis without the ergodic assumption, Seismol. Res. Lett. 70, no. 1, 19–28, doi: 10.1785/gssrl.70.1.19.

Bahrampouri, M., A. Rodriguez-Marek, S. Shahi, and H. Dawood (2021). An updated database for ground motion parameters for KiK-net records, Earthq. Spectra 37, no. 1, 505–522, doi: 10.1177/8755293020952447.

Blitzstein, J. K., and J. Hwang (2014). Introduction to Probability, First Ed., Chapman and Hall/CRC, New York, New York, 385–401, doi: 10.1201/b17221.

Derras, B., P. Y. Bard, and F. Cotton (2014). Towards fully data driven ground-motion prediction models for Europe, Bull. Earthq. Eng. 12, no. 1, 495–516, doi: 10.1007/s10518-013-9481-0.

Derras, B., P. Y. Bard, F. Cotton, and A. Bekkouche (2012). Adapting the neural network approach to PGA prediction: An example based on the KiK-net data, Bull. Seismol. Soc. Am. 102, no. 4, 1446–1461, doi: 10.1785/0120110088.

Dhanya, J., and S. T. Raghukanth (2018). Ground motion prediction model using artificial neural network, Pure Appl. Geophys. 175, no. 3, 1035–1064, doi: 10.1007/s00024-017-1751-3.

Douglas, J. (2003). Earthquake ground motion estimation using strong-motion records: A review of equations for the estimation of peak ground acceleration and response spectral ordinates, Earth. Sci. Rev. 61, nos. 1/2, 43–104, doi: 10.1016/S0012-8252(02)00112-5.

Douglas, J., and B. Edwards (2016). Recent and future developments in earthquake ground motion estimation, Earth. Sci. Rev. 160, 203–219, doi: 10.1016/j.earscirev.2016.07.005.

Esteva, L., and E. Rosenblueth (1964). Espectros de temblores a distancias moderadas y grandes, Boletin Sociedad Mexicana de Ingenieria Sísmica 2, no. 1, 1–18 (in Spanish).

Fujimoto, K., and S. Midorikawa (2006). Relationship between average shear-wave velocity and site amplification inferred from strong motion records at nearby station Pairs, Proc. Jpn Assoc. Earthq. Eng. 6, no. 1, 11–22, doi: 10.5610/jaee.6.11.

Harris, C. R, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Courapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, et al. (2020). Array programming with NumPy, Nature 585, no. 7825, 357–362.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment, Comput. Sci. Eng. 9, no. 3, 90–95.

Inkscape Project (2020). Inkscape, available at https://inkscape.org (last accessed November 2021).

Jayaram, N., and J. W. Baker (2009). Correlation model for spatially distributed ground-motion intensities, Earthq. Eng. Struct. Dynam. 38, no. 15, 1687–1708, doi: 10.1002/eqe.922.

Kingma, D. P., and J. L. Ba (2015). Adam: A method for stochastic optimization, 3rd International Conf. Learning Representations, ICLR 2015—Conference Track Proceedings, 1–15.

Kong, Q., D. T. Trugman, Z. E. Ross, M. J. Bianco, B. J. Meade, and P. Gerstoft (2019). Machine learning in seismology: Turning data into insights, Seismol. Res. Lett. 90, no. 1, 3–14, doi: 10.1785/0220180259.

Kotha, S. R., D. Bindi, and F. Cotton (2016). Partially non-ergodic region specific GMPE for Europe and Middle-East, Bull. Earthq. Eng. 14, no. 4, 1245–1263, doi: 10.1007/s10518-016-9875-x.

Kotha, S. R., G.Weatherill, D. Bindi, and F. Cotton (2020). A regionally-adaptable ground-motion model for shallow crustal earthquakes in Europe, 18, 4091–4125, doi: 10.1007/s10518-020-00869-1.

Kuehn, N. M., and N. A. Abrahamson (2020). Spatial correlations of ground motion for non-ergodic seismic hazard analysis, Earthq. Eng. Struct. Dynam. 49, no. 1, 4–23, doi: 10.1002/eqe.3221.

Landwehr, N., N. M. Kuehn, T. Scheffer, and N. Abrahamson (2016). A nonergodic ground-motion model for California with spatially varying coefficients, Bull. Seismol. Soc. Am. 106, no. 6, 2574–2583, doi: 10.1785/0120160118.

Lecun, Y., Y. Bengio, and G. Hinton (2015). Deep learning, Nature 521, no. 7553, 436–444, doi: 10.1038/nature14539.

Levie, R., C. Yapar, G. Kutyniok, and G. Caire (2020). Pathloss prediction using deep learning with applications to cellular optimization and efficient D2D link scheduling, ICASSP, IEEE Inernational Conf. Acoustics, Speech and Signal Processing, Barcelona, Spain, 4–8 May 2020, doi: 10.1109/ICASSP40776.2020.9053347.

Loth, C., and J. W. Baker (2013). A spatial cross-correlation model of spectral accelerations at multiple periods, Earthq. Eng. Struct. Dynam. 42, no. 3, 397–417, doi: 10.1002/eqe.2212.

Markhvida, M., L. Ceferino, and J. W. Baker (2018). Modeling spatially correlated spectral accelerations at multiple periods using principal component analysis and geostatistics, Earthq. Eng. Struct. Dynam. 47, no. 5, 1107–1123, doi: 10.1002/eqe.3007.

Nakano, K., and H. Kawase (2021). Spatial properties of the site amplifications evaluated by generalized inversion technique with Fourier spectra and response spectra, The 6th IASPEI/IAEE International Symposium: Effects of Surface Geology on Seismic Motion, Kyoto, Japan, 30 August–1 September 2021, GS5-P03.

National Research Institute for Earth Science and Disaster Resilience (2019). NIED K-NET, KiK-net, doi: 10.17598/NIED.0004.

Pagani, M., D. Monelli, G. Weatherill, L. Danciu, H. Crowley, V. Silva, P. Henshaw, L. Butler, M. Nastasi, L. Panzeri, et al. (2014). Openquake engine: An open hazard (and risk) software for the global earthquake model, Seismol. Res. Lett. 85, no. 3, 692–702, doi: 10.1785/0220130087.

Pozos-Estrada, A., R. Gómez, and H. P. Hong (2014). Use of neural network to predict the peak ground accelerations and pseudo spectral accelerations for Mexican Inslab and Interplate Earthquakes, Geofis. Int. 53, no. 1, 39–57, doi: 10.1016/S0016-7169(14)71489-8.

QGIS Development Team (2021). QGIS Geographical Information System, available at https://www.qgis.org (last accessed November 2021).

Ronneberger, O., P. Fischer, and T. Brox (2015). U-net: Convolutional networks for biomedical image segmentation, in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Vol. 9351, Springer, Cham, 234–241, doi: 10.1007/978-3-319-24574-4_28.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986). Learning representations by back-propagating errors, Nature 323, no. 9, 533–536, doi: 10.7551/mitpress/1888.003.0013.

Senna, S., T. Maeda, Y. Inagaki, H. Suzuki, H. Matsuyama, and H. Fujiwara (2013). Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband strong motion evaluation, J. Disast. Res. 8, no. 5, 889–903, doi: 10.20965/jdr.2013.p0889.

Senna, S., A. Wakai, A. Yatagai, K. Jin, H. Matsuyama, H. Suzuki, and H. Fujiwara (2019). Modeling of the subsurface structure from the seismic bedrock to the ground surface for a broadband strong motion evaluation in Japan, in Earthquake Geotechnical Engineering for Protection and Development of Environment and Constructions, F. Silvestri and N. Moraci (Editors), CRC

Press, London, United Kingdom.
Sgobba, S., G. Lanzano, and F. Pacor (2021). Empirical nonergodic shaking scenarios based on spatial correlation models: An application to central Italy, Earthq. Eng. Struct. Dynam. 50, no. 1, 60–80, doi: 10.1002/eqe.3362.
Stafford, P. J. (2014). Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models, Bull. Seismol. Soc. Am. 104, no. 2, 702–719, doi: 10.1785/0120130145.
Strasser, F. O., N. A. Abrahamson, and J. J. Bommer (2009). Sigma: Issues, insights, and challenges, Seismol. Res. Lett. 80, no. 1, 40–56, doi: 10.1785/gssrl.80.1.40.
The Headquarters for Earthquake Research Promotion (2021). Modeling concept of subsurface structures from seismic bedrock to ground surface, available at https://www.jishin.go.jp/evaluation/strong_motion/underground_model/integration_model_kanto_2021/ (last accessed January 2022) (in Japanese).
Trifunac, M. D. (1976). Preliminary empirical model for scaling Fourier amplitude spectra of strong ground acceleration in terms of modified Mercalli intensity and recording site conditions, Bull. Seismol. Soc. Am. 66, no. 4, 1343–1373, doi: 10.1002/eqe.4290070106.
Virtanen, P., R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python, Nat. Methods 17, no. 3, 261–272.
Wakamatsu, K., and M. Matsuoka (2013). Nationwide 7.5-arc-second Japan engineering geomorphologic classification map and Vs30 zoning, J. Disast. Res. 8, no. 5, 904–911, doi: 10.20965/jdr.2013.p0904.
Wakamatsu, K., and M. Matsuoka (2020). Update of the nation wide 7.5-arc-second Japan engineering geomorphologic classification Map and Vs30 Zoning, Bull. Jpn. Assoc. Earthq. Eng. 40, 24–27.
Yao, W., Z. Zeng, C. Lian, and H. Tang (2018). Pixel-wise regression using U-Net and its application on pansharpening, Neurocomputing 312, 364–371, doi: 10.1016/j.neucom.2018.05.103.
Zhao, J. X., J. Hu, F. Jiang, J. Zhou, Y. Zhang, X. An, M. Lu, and D. A. Rhoades (2015). Nonlinear site models derived from 1D analyses for ground-motion prediction equations using site class as the site parameter, Bull. Seismol. Soc. Am. 105, no. 4, 2010–2022, doi: 10.1785/0120150019.
Zhao, J. X., S. Zhou, J. Zhou, C. Zhao, H. Zhang, Y. Zhang, P. Gao, X. Lan, D. Rhoades, Y. Fukushima, et al. (2016). Ground-motion prediction equations for shallow crustal and upper-mantle earthquakes in Japan using site class and simple geometric attenuation functions, Bull. Seismol. Soc. Am. 106, no. 4, 1552–1569, doi: 10.1785/0120150063.
Zhao, J. X., X. Liang, F. Jiang, H. Xing, M. Zhu, R. Hou, Y. Zhang, X. Lan, D. A. Rhoades, K. Irikura, et al. (2016). Ground-motion prediction equations for subduction interface earthquakes in Japan using site class and simple geometric attenuation functions, Bull. Seismol. Soc. Am. 106, no. 4, 1518–1534, doi: 10.1785/0120150034.
Zhao, J. X., F. Jiang, P. Shi, H. Xing, H. Huang, R. Hou, Y. Zhang, P. Yu, X. Lan, D. A. Rhoades, et al. (2016). Ground-motion prediction equations for subduction slab earthquakes in Japan using site class and simple geometric attenuation functions, Bull. Seismol. Soc. Am. 106, no. 4, 1535–1551, doi: 10.1785/0120150056.

APPENDIX

A U-Net architecture

In this appendix, we provide detailed technical description of the U-Net architecture depicted in Figure 1.

From the original U-Net, we adapt the resolution of 572 × 572 pixels for the input features, which ensures even-sized features throughout the U-Net, a requirement for a smooth processing pipeline. The corresponding size of the input maps in terms of square kilometers is determined by the spacing between pixels that can be chosen by the user. A larger resolution of input features could be chosen to either investigate a larger area of interest or to increase the spatial resolution. The practical limitation for the resolution is the amount of available memory on the used graphics processing unit (GPU).

The first element of the U-Net is the encoder branch - a sequence of alternating Conv blocks and Pooling blocks through which the input is processed (compare Fig. 1). Each Conv block encompasses two convolutional layers, each followed by a tanh activation function. Features that are fed to a convolutional layer are convolved with a number of filter-masks of size (3 × 3), in which each filter mask generates one output feature. Thus, the number of features at a given point in the U-Net equals the number of filter-masks in the previously passed convolutional layer. We decided to reduce the number of filter-masks by a factor of 2 compared to the original U-Net in Ronneberger et al. (2015) due to GPU memory limitations. Subsequent passing through a nonlinear activation function (in our case the tanh) is required to learn nonlinear relations between the predictive parameters and the target parameter. Although the original U-Net uses the Rectified Linear Unit (ReLU) as default activation function, we observed that using the tanh leads to more stable convergence during training in our case. The purpose of the Max pooling blocks is to reduce the resolution of the features along the encoder branch. This is achieved via dividing each feature into subarrays of size 2 × 2 and picking the respective maximum values, thus, reducing its resolution by a factor of 2. This factor is adapted from the original U-Net. Along

the encoder branch, the size of the features continuously decreases, whereas the number of features increases. It can thus be considered a feature extractor concentrating the information given in the input layer into a number of low resolutions, highly abstract features.

The second element of the U-Net is the decoder branch—a sequence of alternating upsamplings, skip connections, and Conv blocks. Upsampling operations increase the resolution of the features via bilinear interpolation, conversely with respect to the Pooling blocks in the encoder branch. The upsampled features are subsequently concatenated with the corresponding features from the encoder branch, which is called skip connection. Via skip connection, the abstract features derived in the feature extraction are combined with the less abstract features from the encoder branch that include more spatial context. In the original U-Net architecture, the upsampling operations are followed by convolutional layers before the skip connections. In our setup, these layers cause significant artifacts in the final results, and, therefore, we removed them from the architecture.

The spatial context information and the abstract features are then combined in the Conv blocks, which eventually leads to less abstract and more practical representations of the extracted features toward the output layer. The final Convout block consists of two convolutional layers that are organized in parallel, in which the first one is followed by a linear and the second one by an exponential activation function. These two parallel layers provide two outputs that we interpret as the mean and the variance of the target intensity measure (IM), respectively. Because no padding is applied to the features during convolution operations, the resolution of the output features is smaller (388 × 388) than that of the input features (572 × 572).