Deep learning-based retrofitting and seismic risk assessment of road networks

Rodrigo Silva-Lopez[1], Jack W. Baker[2], and Alan Poulos[3]

[1]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, 94305, California. Email: rsilval@stanford.edu
[2]Professor, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, 94305, California
[3]Ph.D. Candidate, Dept. of Civil and Environmental Engineering, Stanford University, Stanford, 94305, California

ABSTRACT

Seismic risk assessment of road systems involves computationally expensive traffic simulations to evaluate the performance of the system. To accelerate this process, this paper develops a neural network surrogate model that allows rapid and accurate estimation of changes in traffic performance metrics due to bridge damage. Some of the methodological aspects explored when calibrating this neural network are defining sampling protocols, selecting hyperparameters, and evaluating practical considerations of the model. In addition to the neural network, a modified version of the Local Interpretable Model-Agnostic Explanation (LIME) is proposed as a retrofitting strategy that minimizes earthquakes' impact on the system. The modified version (LIME-TI) uses Traffic Impacts and rates of occurrence to aggregate the importance of individual damage realizations during the computation of variable importance. This study uses the San Francisco Bay Area road network as a testbed. As a conclusion of this study, the neural network accurately predicts the system's performance while taking five orders of magnitude less time to compute traffic metrics, allowing decision-makers to evaluate the impact of retrofitting bridges in the system quickly. Moreover, the proposed LIME-TI metric is superior to others (such as traffic volume or vulnerability) in identifying bridges whose retrofit effectively improves network performance.

INTRODUCTION

Road networks are fundamental for the normal functioning of communities, allowing the flow of goods and people throughout different geographical locations. Unfortunately, road networks elements, such as bridges and roads, are vulnerable to damage from natural disasters such as earthquakes. In the face of this challenge, decision-makers are confronted with using scarce resources to minimize the consequences of disruptive events, improving the ability of communities to recover from them.

In addition to being vulnerable, distributed systems, such as road networks, comprise thousands of elements that interact with each other in a highly complex manner. Moreover, given that the occurrence and characteristics of future events have large uncertainties, using a single scenario for risk analysis is inadequate (Bommer et al. 2000), and properly informed decision-making requires simulating the impact of several scenarios. Considering the complexities of evaluating the response of the system to a single scenario, which must be repeated several times, the computational time required to analyze distributed systems can impede decision-makers from evaluating actions to minimize risk. One common approach to decrease these computational challenges has been to select subsets of scenarios (Miller and Baker 2015; Han and Davidson 2012). However, the use of fewer scenarios has implied an increase in the uncertainty of the predictions.

Motivated by the previous problems and the limitations of existing approaches, the first goal of this study is to propose a neural network surrogate model to minimize computational times required to analyze complex road networks. The focus of the proposed neural network is to help decision-makers evaluate the effectiveness of different strategies to retrofit vulnerable bridges in seismic regions.

As the second goal of this study, taking advantage of the neural network structure, a bridge retrofitting strategy is proposed. This strategy uses a modified Local Interpretative Model-Agnostic Explanation considering Traffic Importance (LIME-TI) algorithm as a variable importance model to detect bridges in the network that contribute significantly to disrupting traffic.

The main contribution of this study is to prove that the use of neural networks allows rapid and accurate surrogate models that can be instrumental for optimal decision-making for seismic risk management of road networks. In addition, this work proposes a modified LIME-TI variable importance algorithm that efficiently decreases the potential impacts of earthquakes.

RELATED WORK

Retrofitting bridges is the most common strategy used by decision-makers as a preparedness policy against earthquakes. As limited resources are assigned to invest in the improvement of road networks, the optimal allocation has been the motivation of several previous studies. One popular way to solve this optimization problem is to use a two-staged stochastic optimization (Gomez and Baker 2019; Miller-Hooks et al. 2012; Liu et al. 2009). Although these studies consider the best way to retrofit bridges in road networks, the underlying computational costs in assessing these systems added to the costs of performing the proposed optimizations limit the versatility of these approaches and their ability to quantify the uncertainty in their results. Another popular technique used to identify bridges for retrofitting is to rank them according to an importance metric. This ranking system is the current approach used by the California Department of Transportation (Caltrans) (Caltrans 2019). Other studies that rank bridges are the ones developed by Rokneddin et al. (2013), and Basoz and Kiremidjian (1995). Using a ranking heuristic to retrofit bridges is intuitive and used in practice; however, it limits the ability of the selected bridge set to reflect complex characteristics of the road network, such as the interdependence of its components.

In order to evaluate retrofitting policies, decision-makers need to quantify the seismic risk of road networks. This process is commonly performed by using Monte Carlo simulations in which traffic performance of the network is obtained for several realizations of bridge damage, computed from a set of seismic scenarios (Bommer et al. 2002). A limitation of these Monte Carlo methods is that computing traffic performance is computationally challenging; therefore, the number of seismic scenarios included in the seismic risk analysis is limited. Motivated by this, Han and Davidson (2012) proposed a methodology to select a limited set of earthquake scenarios combining sampling importance and optimization techniques. This concept was extended by Miller and Baker (2015) to consider a proxy for damage as part of the optimization, and by Tomar and Burton (2021) to use an active learning method for selection. Another technique to speed the road network seismic risk assessment is to develop surrogate models for which traffic calculations are efficient. For instance, Özdamar and Demir (2012) and Gomez et al. (2013) use hierarchical techniques to change the resolution of the network while keeping consistency in relevant properties. Lim et al. (2015) use a spectral clustering algorithm to develop a surrogate of the original network, computing zones of greater importance connected by superlinks.

Deep learning techniques have become popular throughout earthquake engineering (Xie et al. 2020) and other areas of civil engineering. Previous efforts have used artificial intelligence, and deep learning techniques tools to quantify road network seismic risk in an accelerated way (Nabian and Meidani 2018b; Elhag and Wang 2007). Different studies have trained trained neural networks to rapidly predict complex traffic metrics such as network connectivity or vehicle travel speeds (Nabian and Meidani 2018a; Yu et al. 2017). Moreover, Alemzadeh et al. (2020) used neural networks to optimize recovery of interdependent distributed systems, showing the utility of this kind of tool for infrastructure system management.

Calibrated neural networks can also inform what drives road network performance. However, the internal mechanisms of neural network predictions are complicated, and a lack of understanding of causality between inputs and predicted values can limit the utility of these models. Olden et al. (2004) explored techniques to assess the importance of variables within a neural network, but the considered techniques were all computationally expensive. SHapley Additive exPlanations (Lundberg and Lee 2017) and Local Interpretable Model-Agnostic Explanations (Ribeiro et al. 2016), are popular and efficient methods for determining feature importance. LIME has also been extended to aggregate several realizations to explain the behavior of the whole model, which can be used to detect general significant features in a neural network (Ribeiro et al. 2016).

Relative to prior work, this study extends surrogate models to substitute a more complex traffic computation in large road networks. Furthermore, recent algorithms for determining feature importance are adopted and extended to identify bridges that contribute most to network disruption risk.

METHODOLOGICAL OVERVIEW

The study methodology consists of four main steps: (1) seismic risk analysis and data generation, (2) traffic performance estimation using neural networks, (3) development of a modified LIME algorithm (LIME-TI) to identify bridges for retrofit, and (4) evaluation of the performance of the proposed retrofitting strategy. This process is presented in Figure 1.

The first step involves data generation through seismic risk assessment of the road network. The data consist of realizations of traffic performance associated with damaged versions of the road network due to seismic scenarios, generated through four sub-steps: (1) selecting seismic scenarios; (2) obtaining realizations of the damaged road network; (3) using a traffic model, traffic demand is assigned over the network; and (4) computing a traffic performance metric.

The second step is to build a neural network surrogate model to compute traffic performance. The neural network architecture is defined by selecting hyperparameter values and evaluating neural network prediction accuracy. This architecture definition is performed iteratively until accuracy is maximized. Once hyperparameters have been selected, a calibrated neural network has been defined.

The third step uses the neural network and a modified LIME variable importance algorithm (LIME-TI) to predict which bridges have the greatest contribution to the system's performance. This ranking is used to identify bridges that could be retrofit to reduce seismic risk.

Finally, the fourth step evaluates the road network's performance when subject to the retrofitting actions from the third step.

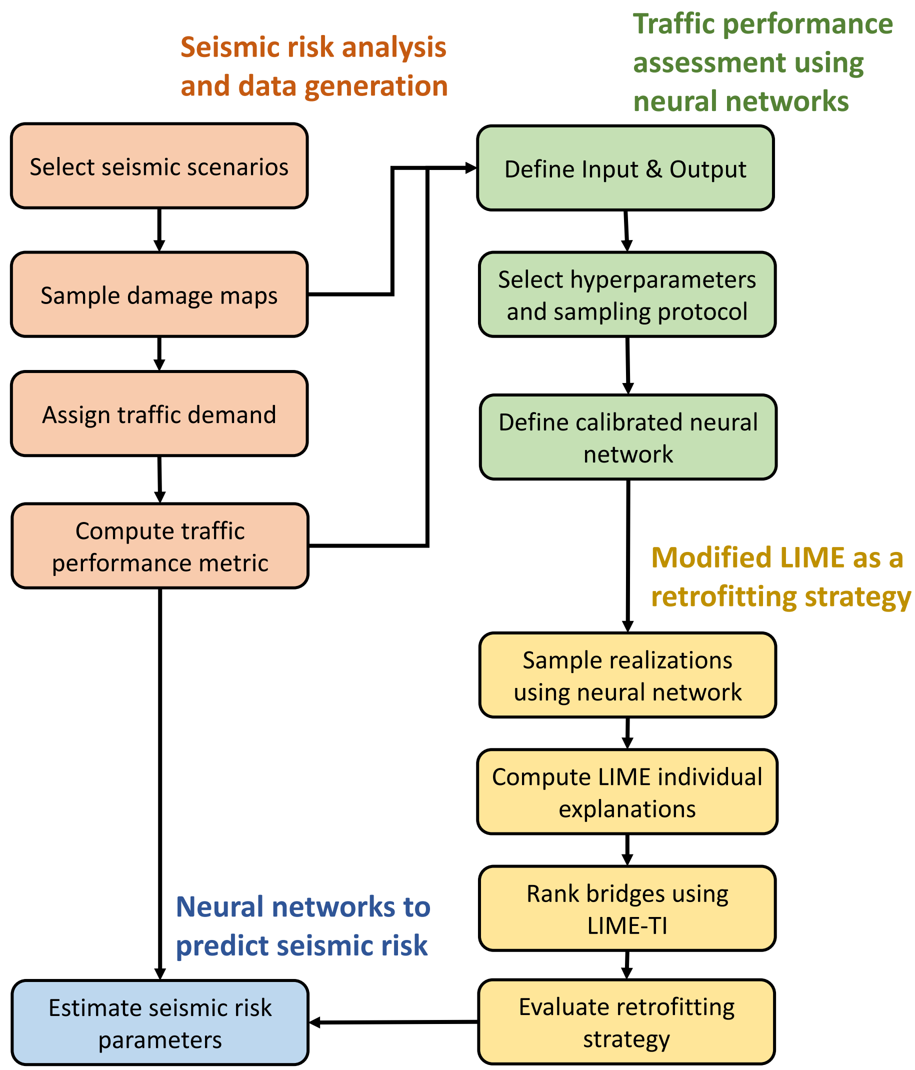

Fig. 1. Diagram of the process used in this study. Each part of the process matches a section in this paper.

To demonstrate the process proposed in this study, the road network of the San Francisco Bay Area is used as a testbed. This network is modeled by a directed graph containing 32,858 edges, 11,921 nodes, and 1743 bridges, as shown in Figure 2. This model was developed by Miller (2014), and it was modified for this study by updating and correcting properties of some links.

SEISMIC RISK ANALYSIS AND DATA GENERATION

The first step of the methodology, as shown in Figure 1, is to generate training data by performing a seismic risk analysis of the road network. Seismic risk assessment starts by selecting hazard consistent seismic scenarios, which means sampling the scenarios considering their probability of

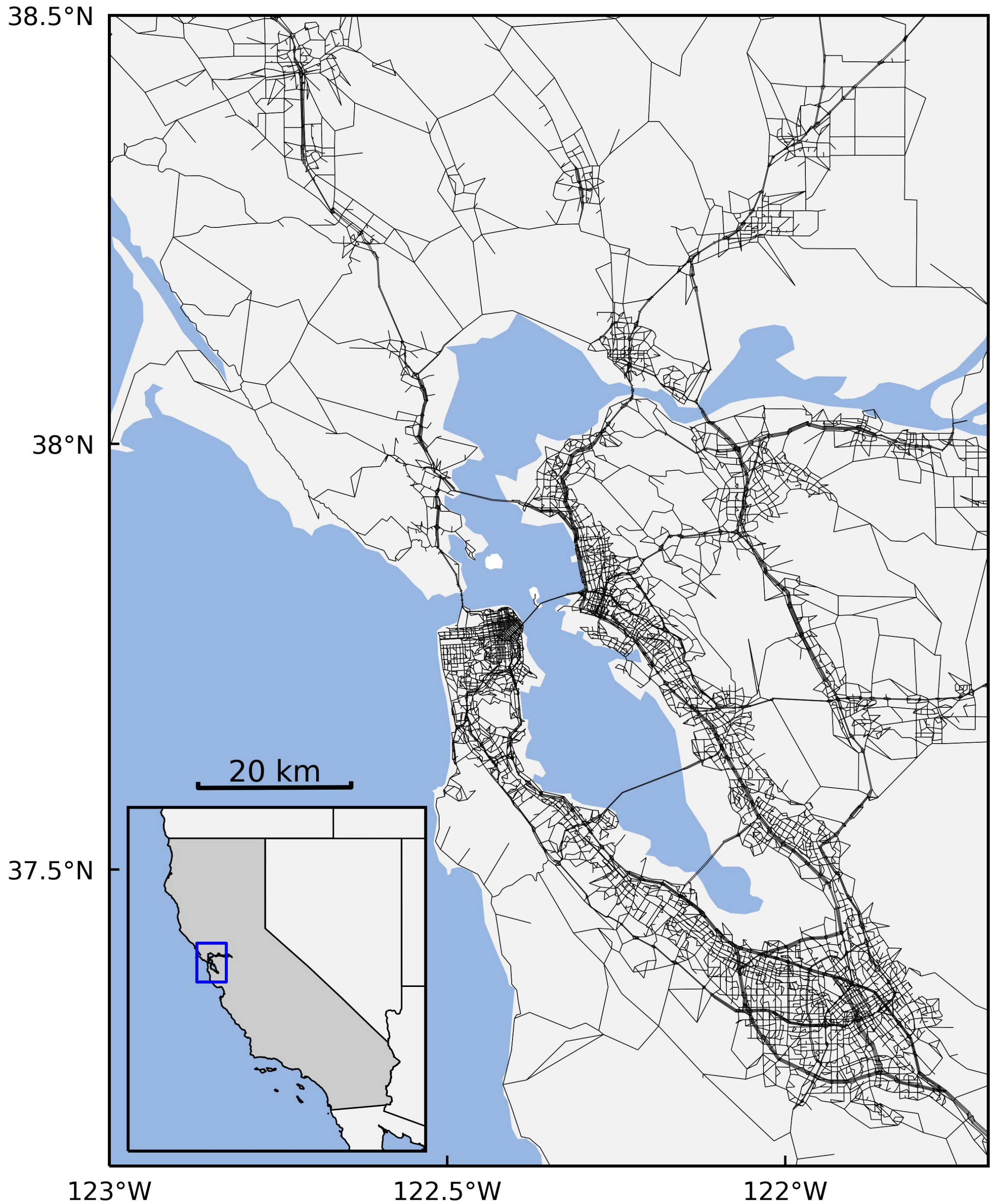

Fig. 2. Road network of the San Francisco Bay Area used in this study.

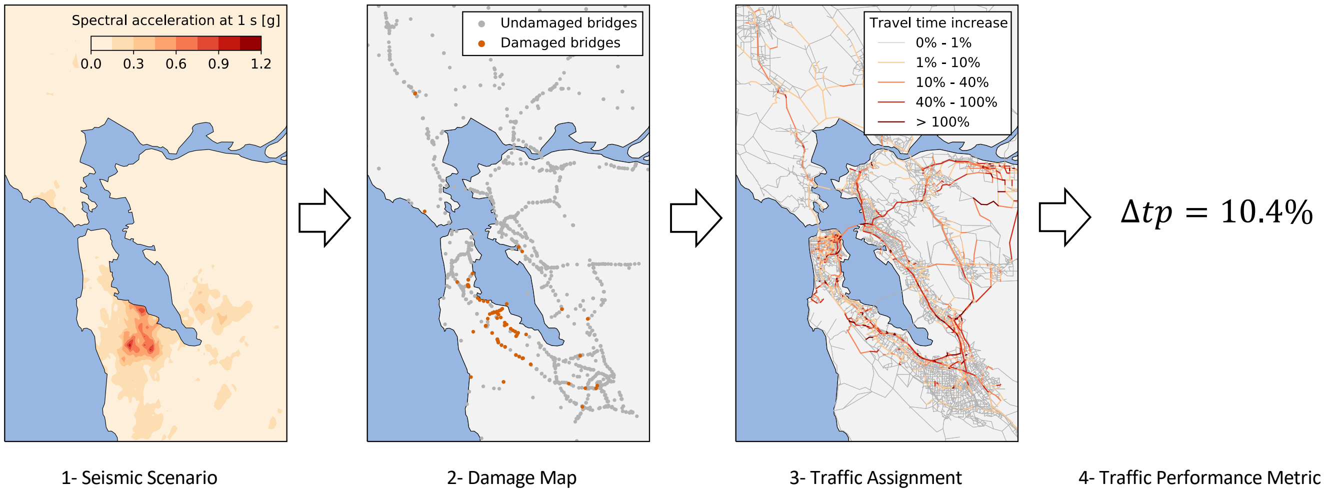

Fig. 3. Four steps of seismic risk assessment of road systems for a given seismic scenario.

occurrence according to the geological and seismological characteristics of the region. For each scenario, the computation of network performance follows the four sub-steps illustrated in Figure 3.

The location and magnitude of the seismic scenario, along with local site information, are used to feed a ground motion model (GMM) and spatial correlation model to obtain values of ground motion intensity measure (IM) at the locations of all bridges are applied. For this study (i.e., the San Francisco Bay Area road network), the IM used is the spectral acceleration at a period of 1 s. The seismic scenarios are obtained using the OpenSHA Event Set Calculator (Field et al. 2003) with the seismic source model developed by Field et al. (2009), near-surface shear-wave velocity (V_{s30}) obtained by the method from Wald and Allen (2007), the Boore and Atkinson (2008) GMM, and the model of Jayaram and Baker (2009) for spatial correlation in the IM values. These parameters and models are consistent with the study of Miller (2014), upon which this work builds.

For a given sampled value of intensity measure, the following fragility function is used to compute the probability of damage for each bridge in the system.

$$P(DS_i \geq ds_k | Y_i = y) = \Phi\left(\frac{\ln\left(y/(\alpha\lambda_{k,i})\right)}{\beta_{k,i}}\right) \quad (1)$$

where:
DS_i = Damage state of component i.
ds_k = Damage state k.
Y_i = Ground motion intensity measure value at the location of component i.
$P(DS_i \geq ds_k | Y_i = y)$ = Probability of component i's damage state being ds_k or greater, given $Y_i = y$.
Φ = Standard normal cumulative distribution function.
α = Factor that indicates retrofitting action: $\alpha = 1$ for no retrofit and $\alpha = 1.2$ for retrofit.
$\lambda_{k,i}$ = Median of Y_i causing damage state k or greater.
$\beta_{k,i}$ = Standard deviation of $ln(Y_i)$ causing damage state k or greater.

Damage states of each bridge are sampled with the above probabilities to obtain a *damage map*. For this study, the relation between damage state of bridges and their traffic functionality is consistent with Werner et al. (2006), who assigned functionality to the damage states proposed by HAZUS (MR 2003). Specifically, no traffic is allowed on bridges in an extensive or complete damage state, and bridges with no, slight, or moderate damage are still functional. For each seismic scenario, several realizations of damage maps are obtained. For the San Francisco Bay Area, the fragility curve parameters (i.e., $\lambda_{k,i}$ and $\beta_{k,i}$) were provided by Caltrans (Miller 2014).

On the damaged version of the network, traffic demand is assigned using an Iterative Traffic Assignment Algorithm (Chen and Alfa 1991). This algorithm takes fractions of the total demand for each pair of origins and destinations and assigns it to the current shortest path between them. Then, after the traffic has been assigned, the flow on each road of the network is updated, which modifies the travel time on the road segment according to Equation (2), (Beckmann et al. 1956).

$$t_a = t_f \left(1 + 0.15 \left(\frac{q_a}{c_f}\right)^4\right) \tag{2}$$

Where:
t_a = Travel time of the road after flow assignment.
t_f = Free flow travel time of the road.
q_a = Current flow on the road.
c_f = Capacity of the road.
This process is performed for each iteration and each pair of origins and destinations. Once all users have been assigned a path, the aggregated travel time of the network is computed by adding all travel times. The iterative traffic assignment is computationally expensive and imposes challenges in evaluating seismic risk on these systems, especially when this process is performed to assess the effectiveness of each retrofitting strategy.

Bridge damage may also disconnect some areas and prevent some trips from being completed. To account for these lost trips, the following traffic performance metric (tp) is used.

$$tp_k = tt_k + \gamma n_{lt_K} \tag{3}$$

Where:
tp_k = Traffic performance metric for damage map k. This metric has units of time.
tt_k = Aggregated travel time for users for damage map k.
n_{lt_k} = Number of lost trips due to lack of connectivity for damage map k.
γ = Penalty factor for lost trips. This study considers γ = 4 hours.

The percent change in the traffic performance metric Δtp for damage map k, with respect to the undamaged condition UD, is defined as:

$$\Delta tp_k = \frac{tp_k - tp_{UD}}{tp_{UD}} \times 100 \tag{4}$$

Where tp_k is the traffic performance metric shown in Equation (3) for damage map k, and tp_{UD} is the traffic performance metric in the undamaged condition which is equivalent to the aggregated travel time for all users.

The above process of computing network performance is performed for many scenarios. The results are used as training data in the following section. They can also be combined with the occurrence rates of the scenarios to compute Loss Curves, which measure the annual rate of exceedance of a given network performance metric value:

$$\lambda_{\Delta tp \geq \Delta tp'} = \sum_{k=1}^{n} w_k \mathbf{I}(\Delta tp_k \geq \Delta tp') \tag{5}$$

Where:
$\Delta tp'$ = Some level of increase in traffic performance metric.
$\lambda_{\Delta tp \geq \Delta tp'}$ = Annual rate of events with $\Delta tp \geq \Delta tp'$.
w_k = Annual rate of occurrence of damage map k (based on the occurrence rate of the associated seismic scenario).
n = Number of damage maps considered.
$\mathbf{I}(\,)$ = an indicator function equal to 1 if the argument is true and 0 otherwise.

Another indicator of network performance is the expected annual increase in the traffic performance metric:

$$E[\Delta tp] = \sum_{k=1}^{n} w_k \, \Delta tp_k \tag{6}$$

Where:
Δtp_k = Level of increase in traffic performance metric for damage map k.
w_k = Annual rate of occurrence of damage map k (based on the occurrence rate of the associated seismic scenario).
$E[\Delta tp]$ = Expected annual increase in traffic performance metric.

This aggregated measure allows comparing the effects of different retrofitting strategies. In this study, loss curves and expected annual increases in travel time are called "seismic risk parameters" and are used to evaluate the accuracy of the neural network.

TRAFFIC PERFORMANCE ASSESSMENT USING NEURAL NETWORKS

The second step of Figure 1 is to develop a neural network as a surrogate model for the traffic model. This model will take bridge damage states as input and output the traffic performance metric of Equation 3, enabling a fast calculation of traffic parameters in a damaged road network. That is, the neural network replaces Step 3 of the seismic risk assessment process (see Figure 3), which is the most computationally expensive step.

Note that a neural network could be formulated to develop an End-to-End surrogate (Nabian and Meidani 2018b). However, a neural network that skips more steps of the process will increase the uncertainty in the predicted results and decrease the model's accuracy. In this case, generating realizations of damage given values of intensity measures is not computationally expensive, and hence there is no benefit to skip Step 2 while decreasing the reliability of the model. This study thus develops a neural network that only replaces the step of the seismic risk analysis with high computational cost.

To begin explaining the implementation of the neural network used to predict traffic performance, some general methodological considerations for neural network calibration will be introduced. Then, the application of each consideration to the San Francisco Bay Area testbed is shown.

Methodology for neural network calibration

Several aspects need to be considered to train, validate and test a neural network. The most significant for this study are (1) definition of inputs and outputs, (2) selection of sampling protocols for training data, (3) selection of model hyperparameters, and (4) proposal of a evaluation method that is scientifically and practically insightful. The following sections elaborate on these issues.

Definition of inputs and outputs

The first step to develop a neural network is to define the inputs and outputs of the model. One aspect that should be considered when defining these variables is that the computational time required to compute the outputs of the model should be reduced significantly. In addition, to improve network accuracy, randomness in the inputs should be avoided.

Selecting sampling protocols for training data

As part of the calibration process, the sampling protocol that will be used to generate training data must be defined. The generated data should represent conditions that the neural network is expected to encounter during its utilization. Some challenges that should be accounted for during the sampling protocol definition are the selection of a training data distribution that minimizes the neural network's bias on test data. In disaster engineering, the sampling process should consider that highly disruptive events are of great interest but occur much less frequently than non-disruptive events.

Hyperparameters selection

Once the input and outputs of the model have been selected, the parameters that define the neural network, or hyperparameters, need to be determined. These include the number of layers in the neural network, the optimization algorithm, the learning rate of the optimization algorithm, and the number of neurons per layer. In addition, the connectivity between neurons across different layers of the neural network must be determined.

One common approach to determine hyperparameter values is to perform a grid search in which all possible combinations of hyperparameters are evaluated in terms of the accuracy of the neural network on test data. The hyperparameter combination that minimizes the error can then be selected. For each sampling protocol, a hyperparameter selection should be conducted to minimize accuracy error on test data.

Evaluation of the calibrated neural network

The final step of calibrating a neural network is its accuracy evaluation. Through this process, the neural network' suitability as a surrogate model is tested. The evaluation is traditionally performed by analyzing whether its accuracy is within an acceptable range for the specific study. Example metrics that can be used to evaluate the performance of regression neural networks are logarithmic loss, mean squared error, mean absolute error, and coefficient of determination R^2.

The developer of a neural network can additionally propose other evaluation metrics focused on the specific study that is being conducted. For example, instead of aggregate metrics, the developer can propose metrics that focus on cases that are critical to the model. For instance, in structural engineering, it could be essential to focus on the capabilities of a model to predict structural collapse more accurately than another kind of structural response. In this study, these other metrics are named "engineering parameters."

Neural network calibration for the San Francisco Bay Area road network

The following subsections present how the previous calibration considerations for neural networks were implemented in a traffic model for the San Francisco Bay Area road network.

Definition of inputs and outputs

The neural network proposed in this study takes damage states of bridges (input) to compute a metric of overall network performance (output). The damage state of each bridge is represented by a binary variable, which is 1 when the bridge is extensively damaged or worse, and is 0 when not. The output of the neural network is the change in traffic performance metric as presented in Equation (3).

In selecting adequate features, topological features of the network graph (such as edge degree, centrality, or graph diameters) were also analyzed as potential inputs. However, including these parameters did not significantly improve the performance of the trained network. Moreover, in the case of some centrality measures, the time involved in calculating these measures was even higher than running the traffic model, which invalidated the purpose of the neural network itself. Hence, only bridge damage states are used as inputs for the neural network.

Selecting sampling protocols for training data

This study considered three protocols to generate training data: (1) sampling hazard consistent seismic scenarios; (2) over-sampling scenarios that generate severe consequences, named here as Extreme Events; and (3) random sampling of scenarios. The first two sampling protocols align more directly with the intended risk analysis goal, whereas the random sampling protocol is used as a benchmark to compare the performance of the other two protocols. The second protocol is considered because preliminary results indicated that the first protocol did not produce enough highly disruptive events, making the prediction of extreme events less accurate. In all three protocols, simulations with no bridge damage were neglected since they do not provide helpful information for training the neural network.

An extreme event is characterized here as an event that generates an increase in traffic performance metric bigger than 50% of the most significant disruption in the hazard consistent sampling. Given that this categorization is somewhat arbitrary, different values of the extreme event thresholds were explored. The final value of 50% was selected because it provided a balance between generating enough data from hazard consistent seismic scenarios while still corresponding to highly disruptive events. With this threshold, approximately 6% of the damage maps obtained from hazard-consistent sampling are defined as extreme events. In other applications, instead of using a metric relative to the maximum observed disruption, this categorization can be defined considering the decision-maker's acceptable performance threshold and should produce comparable results. To study the effect of over-sampling Extreme Events when training the neural network, several training sets of the same size with different proportions of extreme events were considered by drawing samples of data from the extreme and non-extreme categories with varying proportions. The original pool from which these samples were drawn was larger than the final training data to allow a sufficient number of data to sample from. Specifically, a set that was one order of magnitude bigger than the biggest amount used to train the hazard consistent model was used.

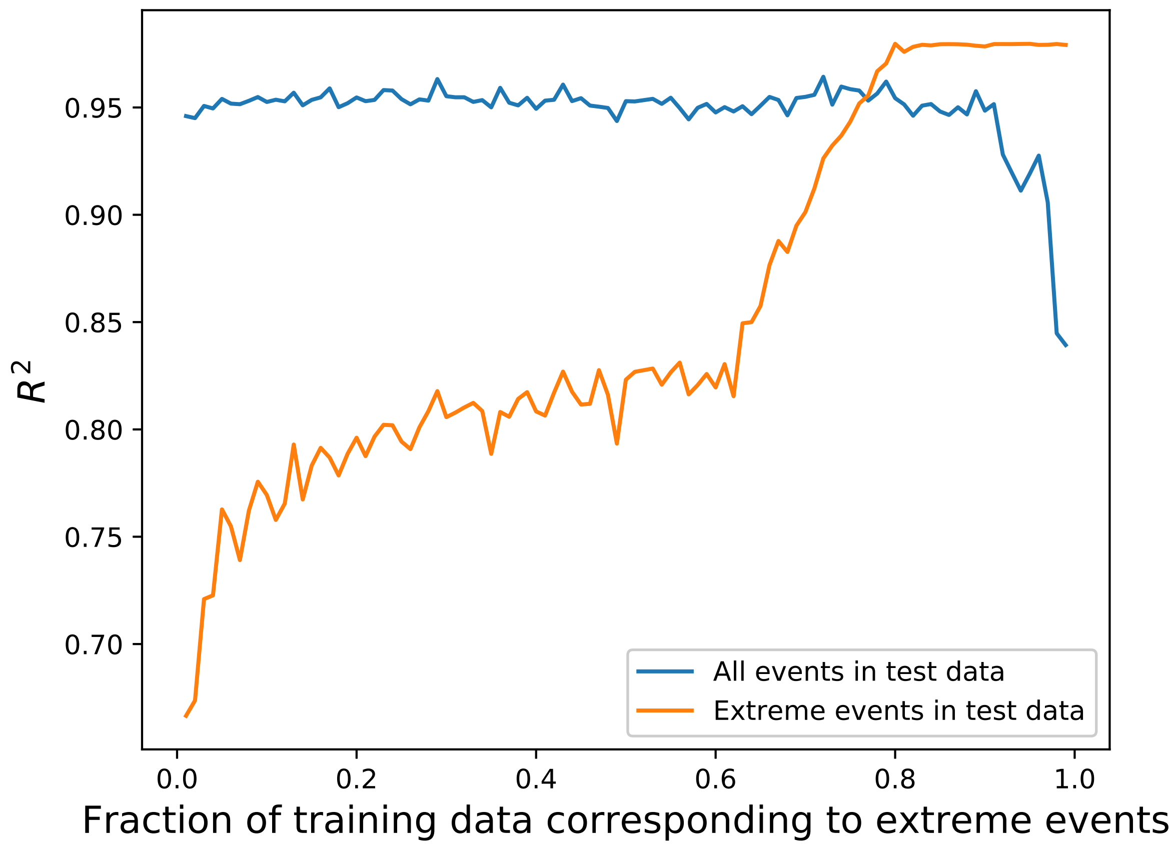

Fig. 4. Coefficients of determination computed considering all realizations in the test dataset and considering realization consistent with the definition of extreme events.

The varying sets of training data were then used to train the neural network, and its performance was measured using the coefficient of determination (R^2) on test data sets. Separate test data sets were compiled that consisted of hazard-consistent events or that consisted of only extreme events. The latter test set allows to directly evaluate the neural network's performance for extreme events–the events of great interest for seismic risk management activities. Performance for varying sampling protocols and test data sets are shown in Figure 4. The accuracy in predicting the *All events* test data is relatively stable for training sets with $\leq$ 90% extreme events but decreases for larger percentages. As expected, the accuracy in predicting the extreme events test data increases as the proportion of extreme events increases in the training data, peaking for training data with $\geq$ 80% extreme events. Thus, the extreme event protocol used for the remainder of this work consists of a training set with a proportion of extreme events of 80%, as in this case, the resulting neural network performs well for both extreme events and all events. Note that the 80% extreme events is a much larger proportion than the 6% of extreme events produced by the hazard consistent protocol. The improved performance from this over-sampling (as seen in Figure 4) was also seen to significantly improve the performance of the neural network in later analysis stages.

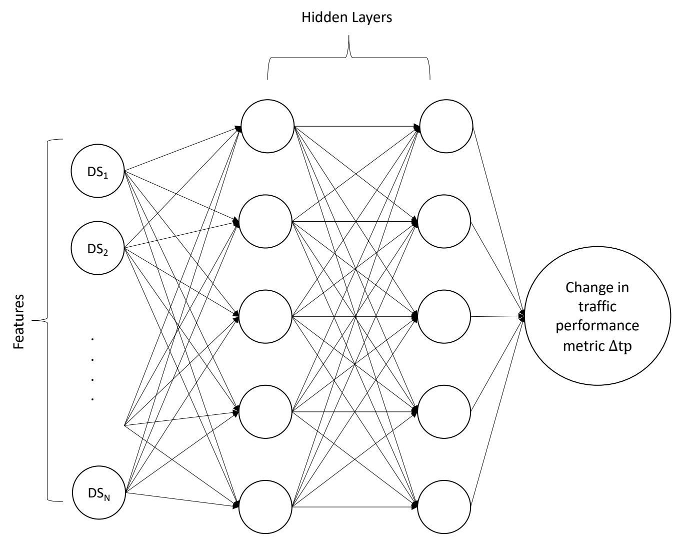

Fig. 5. Illustration of the neural network architecture. The neural network goes from a binary indicator for each bridge, through 10 hidden layers with 150 neurons each, to obtain the change in traffic performance metric.

Hyperparameters calibration

This study performed a hyperparameters calibration of the number of neurons per layer, the number of layers in the neural network, the learning rate of the algorithm, and the L2 penalty factor. These hyperparameters define the architecture of the neural network, which is illustrated in Figure 5. In addition, the calibrated neural network was considered to be fully connected. The neural network was trained using the ADAM (Adaptive Moment Estimation) (Kingma and Ba 2014). This algorithm was chosen because it converged faster than others, such as stochastic gradient descent (SGD). The rectified linear unit function (ReLU) was used for all hidden layers as the activation function. The values of hyperparameters that minimized test error are summarized in Table 1. Despite performing the hyperparameters calibration for each sampling protocol, there were no significant differences between them; therefore, the same hyperparameters were used throughout different sampling protocols.

Evaluation of the calibrated neural network

Using the hyperparameters shown in Table 1, the neural network for each protocol is evaluated using statistical evaluation for neural networks. The first analysis conducted is a measure of model accuracy through the use of R^2. The results of the different regression models are shown in Table 2. Extreme events sampling and hazard-consistent event sampling have a value of R^2 of 0.967 and 0.958, respectively, on the test data, which has not been used to train nor validate the neural network. To verify how good this value is, a linear regression and a ridge regression is performed using the same training data, observing R^2 values of 0.842 and 0.864, respectively. To conduct this regression, we use all the damage states of the bridges as variables in the same way as it is done for the neural network. Thus, neural networks produce more accurate results than other more

TABLE 1. Hyperparameters used to calibrate the neural network model

Hyperparameter	Values
Optimization algorithm	ADAM
Activation function	ReLU
Number of layers	10
Number of neurons per layer	150
Learning rate α	0.001
Penalty Factor L2	0.001

TABLE 2. Coefficient of determination, as a metric of accuracy, for different regression models and sampling protocols.

Regression Model	Coefficient of determination R^2
Neural Network - Extreme Events	0.967
Neural Network - Hazard Consistent	0.958
Neural Network - Random Sampling	0.888
Linear Regression	0.842
Ridge Regression	0.864

straightforward regression techniques.

As mentioned before, the motivation to include a sampling protocol focused on extreme events was to accurately represent events that can be particularly disruptive and of interest to decision-makers. Taking this into account, to show the difference between the predicted values of the neural network that uses hazard-consistent sampling and the one that uses extreme events sampling, Figure 6 compares the Δtp values predicted by the neural network and by the traffic model for both sampling protocols on the test set, showing all test data in grey and events that are extreme in blue. As expected, for extreme events, the prediction performed by extreme events sampling is more accurate than the one performed by hazard-consistent sampling. On the other hand, the prediction of extreme events sampling for non-extreme events is not as accurate as the prediction given by the hazard consistent sampling protocol. Considering that the events that are not that extreme have a smaller role in the risk of road systems, the observed difference is not significant. Possible overfitting of extreme events in Figure 6 was checked by taking several sets of extreme events not used on the training data, and the value of R^2 remained stable and lower than the one observed on the training data. Whether or not the increase in accuracy over extreme events was a result of just including more extreme events into the training set was also explored, concluding that it was not the result of an increase of training data but rather an increase in the proportion of extreme events with respect to the total. This dependence on the proportion is consistent with assigning bigger weights to extreme events while minimizing the training error in the calibration

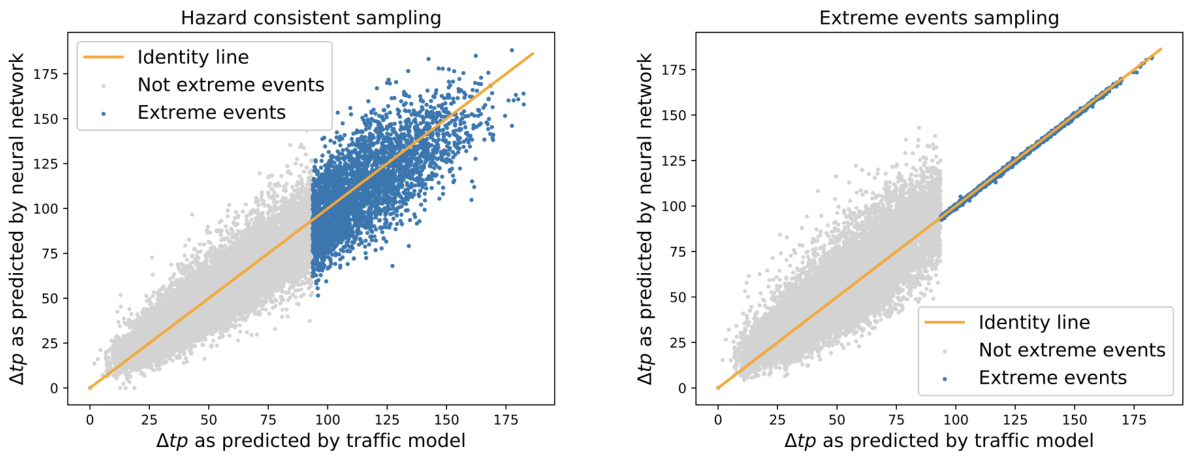

Fig. 6. Scatter plots showing a comparison between predicted values by the neural network, and data obtained using the traffic model. Results for hazard consistent protocol are shown on the left and for extreme events on the right.

of the neural network. The sudden drop observed in Figure 6 is a result of the definition of extreme events and their consequences on the sampling protocol. If other thresholds to characterize extreme events were defined, the sudden drop would be observed in other values of network performance; on the other hand, if a probability distribution was assigned to the sample of extreme events that were not binary and defined by a threshold, then the drop could be minimized. For the effects of this study, the sudden drop does not affect the purpose of the neural network calibration, which is to allow seismic risk assessment. The previous statements regarding how to control the location and magnitude of the drop observed on Figure 6 depend on specific implementations of the reader.

To further illustrate the benefit of extreme events sampling, Figure 7 shows the residuals as a function of Δtp using each model. It is observed that the hazard-consistent protocol has a more considerable bias and that the difference in bias is more significant when predicting extreme events. Note that random sampling is the protocol that leads to the smallest bias for results greater than 75 [%UD]; however, its accuracy is significantly worse than the other models when observing Figure 8. Given the test data structure and how R^2 is computed, there is a trade-off between bias and accuracy. Test data, which represents how modelers will evaluate the network, has seismic events that are hazard consistent, which means an underlying distribution of the data focused on minor events. This distribution is different from the implicit uniform distribution used to compute R^2 since hazard consistency adds more weight to events with less disruption, which happen more frequently. The trade-off between R^2 and the bias comes from this underlying distribution since higher R^2 values will minimize the error for smaller events and at the cost of increasing the error for disruptive events. Different training data distributions were considered, but none could minimize the bias while obtaining a high value of R^2.

Another experiment conducted to verify the network's performance was to explore the accuracy of the neural network as a function of the number of realizations included in the training data. Both the neural network trained with extreme events sampling and hazard-consistent sampling are

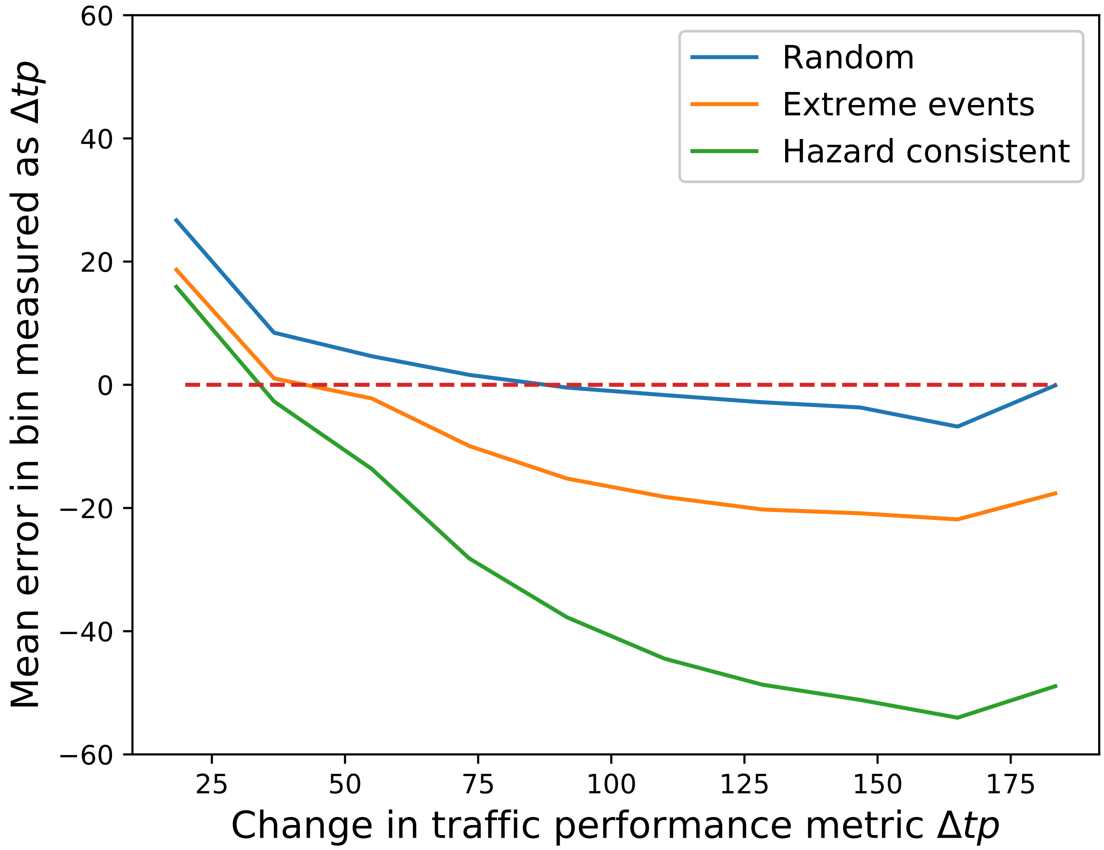

Fig. 7. Mean residuals for extreme events computed using a neural network trained with each of the three sampling protocols. Optimal zero bias shown with red dashed line.

significantly more accurate than the neural network trained using random sampling. Moreover, Figure 8 shows that after approximately 10000 realizations, the neural network's performance remains stable. This information suggests that although it may be computationally expensive to train the network, it is feasible, and the benefits of using the neural network surpass this cost. For reference, for a hazard consistent seismic assessment of the San Francisco Bay Area, 1992 scenarios are used as provided by implementing UCERF2 according to Field et al. (2009). Therefore, training this neural network is equivalent to performing five analyses, which is not a considerable number taking into account that every time a single retrofitting strategy is evaluated, all those 1992 scenarios need to be considered.

As a conclusion of the comparison between the neural networks trained with the three sampling protocols, the neural network that uses extreme events can predict Δtp as well as the neural network that uses hazard-consistent sampling, without inducing a bias on the extreme events, which are particularly important for decision-making in risk management. Given this, the following sections only present results of the neural network trained using extreme events sampling.

MODIFIED LIME ALGORITHM AS A RETROFITTING STRATEGY

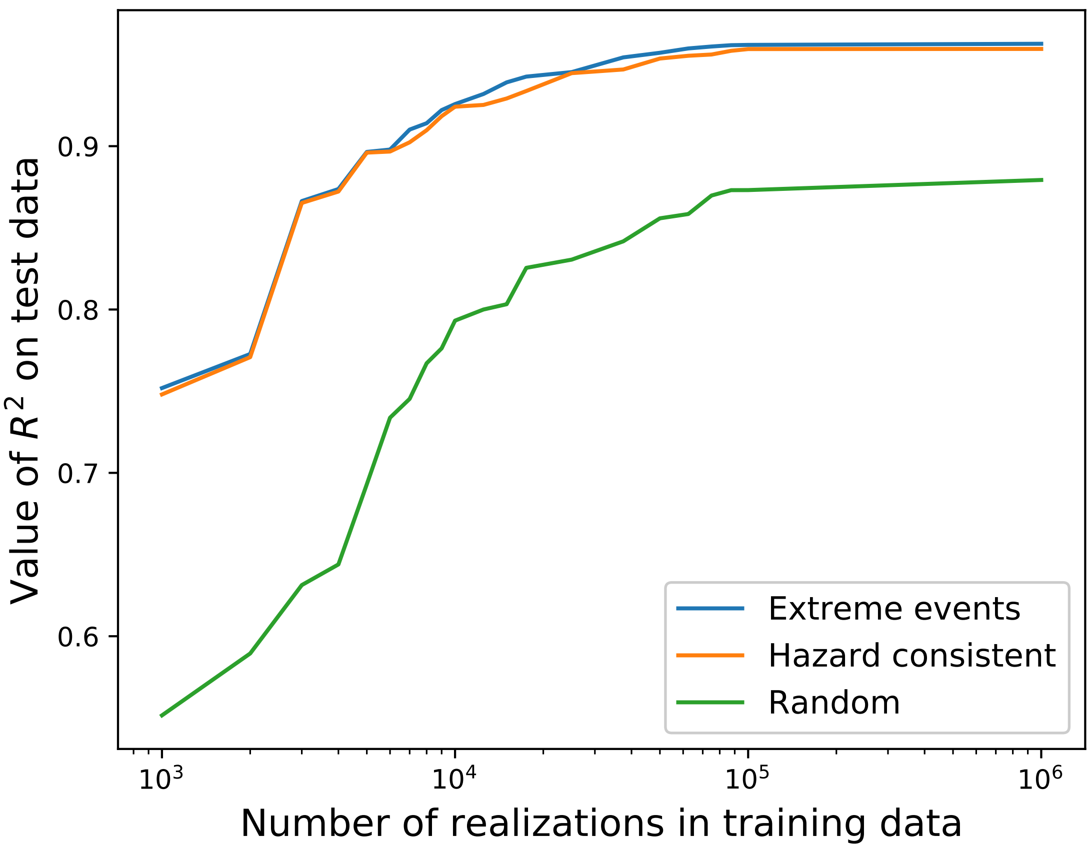

Fig. 8. Evolution of neural network accuracy as a function of the number of realizations included in the training data. Results are shown for random, extreme-events, and hazard-consistent sampling.

In addition to predicting the performance of the damaged road network, the second goal of this study is to use the neural network to identify the bridges whose damage contributes most to network disruption. Given that the input variables are the bridge damage states, the influential bridges can potentially be identified by evaluating the most important input variables in predicting traffic performance. By ranking the importance of the bridges in the model, a new retrofitting strategy can be proposed. This calculation is the third step of the Figure 1 framework.

LIME as a variable importance algorithm

To establish which variables within a model have a more prominent role in predicting its output, it is necessary to use a variable importance algorithm. In that regard, this study uses a Local Interpretative Model-Agnostic Explanation (LIME) to efficiently determine which variables had a more significant role in predicting the output variable (Ribeiro et al. 2016). The LIME algorithm develops local linear regressions around a singular realization of interest using the input variables of the neural network as independent variables and the values predicted by the neural network around the realization as dependent variables. This LIME algorithm was initially meant to explain individual predictions by the neural network; however, it can be used in several realizations to extrapolate which variables contribute more frequently to the neural network model as a whole. This process, named "submodularity pick" in Ribeiro et al. (2016), consists of three steps: (1) select a set of individuals realizations to be explained by the LIME algorithm, (2) compute coefficients W_{jk} for each feature on each realization, and (3) aggregate for each feature the coefficients obtained for every single realization. As a result of this algorithm, each independent variable, which corresponds to a bridge, has a weight I_j that describes how much it contributes to the whole neural network. In the original formulation of LIME, the coefficients of the features k for each realization j are the coefficients of a local linear regression W_{jk}. Then, following step 3, the global coefficient of a feature I_j is defined by

$$I_j = \sqrt{\sum_{k=1}^{d} |W_{jk}|} \tag{7}$$

Where
I_j = Final weight of variable j.
W_{jk} = Weight of variable j in realization k.
d = Number of realizations considered in the subsampling.

Bridges are ranked according to their I_j values, which represent their importance in estimating traffic disruption.

LIME accounting for Traffic Impacts (LIME-TI)

The original submodularity pick calculation above was revised to be more suitable for this application. The modified approach, termed LIME with Traffic Impacts, incorporates the bridge's contribution to traffic performance and the probabilistic nature of seismic events. To account for these two factors, Steps 2 and 3 of LIME were changed. To incorporate bridge importance in traffic performance, Step 2 is modified by computing the coefficients C_{ij} of each bridge j on each realization i as the marginal increase in traffic performance once the bridge has been repaired according to the order suggested by the local explanation of LIME for the realization i, as shown in Equation (8). For instance, if a realization has two bridges A and B, and the explanation of the realization from LIME establishes that bridge A has higher importance than B, then the Traffic Impacts (TI) coefficient for bridge A, C_{iA}, will be the improvement in network performance metric Δtp when only bridge A is repaired. Regarding bridge B, its TI-coefficient C_{iB} will be the improvement on the traffic performance metric Δtp after repairing B, given that A was already repaired. This conditioning on having repaired more important bridges before is why the TI-coefficient is defined as the marginal improvement on the traffic performance metric.

$$C_{jk} = tp_k - tp_{b \geq j,k} \tag{8}$$

Where
C_{jk} = TI-coefficient of bridge j in realization/damage map k.
tp_k = Traffic performance metric computed for damage map k.
$tp_{b \geq j,k}$ = Traffic performance metric computed when all bridges with a higher LIME ranking than bridge j, and bridge j itself have been repaired for damage map k.

In addition to redefining LIME coefficients, this study also modifies how these coefficients are aggregated in Step 3. In place of using Equation (7), which adds all realizations equally, the realizations are now combined using their rates of occurrence w_k as shown in Equation (9). By aggregating all realizations using w_k, each bridge contribution is normalized by their vulnerability and the occurrence rate of seismic scenarios.

$$I_j^* = \sum_{k=1}^{n} w_k |C_{jk}| \tag{9}$$

Where
I_j^* = LIME-TI coefficient of bridge j.
C_{jk} = TI-coefficient of bridge j in realization/damage map k.
w_k = Rate of occurrence of damage map k as used in (5) and (6).

The result of this LIME-TI algorithm is a ranking of bridges that are important to predict the traffic disruption, but it also accounts for their vulnerability and the contribution of other damaged bridges.

Evaluation of retrofitting strategies on the San Francisco Bay Area road network

Once the ranking from LIME-TI is obtained, as shown in Figure 1, its effectiveness to reduce road network disruption is compared to those of rankings obtained using three other potential metrics of importance: structural vulnerability, one-at-a-time (OAT) sensitivity analysis, and traffic capacity of the bridges. The vulnerability ranking is obtained by sorting the number of times the bridge was damaged in all training data scenarios. The more times a bridge experiences damage, the higher its vulnerability ranking. The OAT analysis uses the value of Δtp generated by damaging each bridge individually; then, bridges are ranked in descending order according to the Δtp that their damage causes. Finally, the traffic capacity ranking is based on the amount of traffic that the bridge allows in free-flow conditions. Rankings from the original LIME formulation were also checked, but results are not reported since they were less effective than LIME-TI.

The retrofitting strategies were evaluated by computing the expected annual change in road performance metric $E[\Delta tp]$ for different numbers of bridges retrofitted. For a case of n retrofits, the n highest-ranking bridges according to each metric were assumed to be retrofit. A retrofit bridge has its fragility function in Equation 1 modified, and the entire risk assessment procedure is repeated (with the retrofit bridge being damaged less often). $E[\Delta tp]$ is computed using Equation (6). The performance evaluation considers many scenarios; the traffic performance metric is computed using the neural network surrogate model to reduce computing time. The results of this evaluation are presented in Figure 9, which shows that the ranking based on LIME-TI submodularity yields a bigger improvement in traffic performance than ranking bridges by structural vulnerability, OAT analysis, or traffic capacity.

Explanation on bridges selected by LIME-TI

The superiority of the proposed LIME-TI ranking is reasonable as it is the only ranking that combines the effect of vulnerability and traffic importance. OAT analysis and traffic capacity only quantify the role of the bridge in traffic performance, whereas structural vulnerability only accounts for bridge damage probability without including the importance of the bridge to the system's performance.

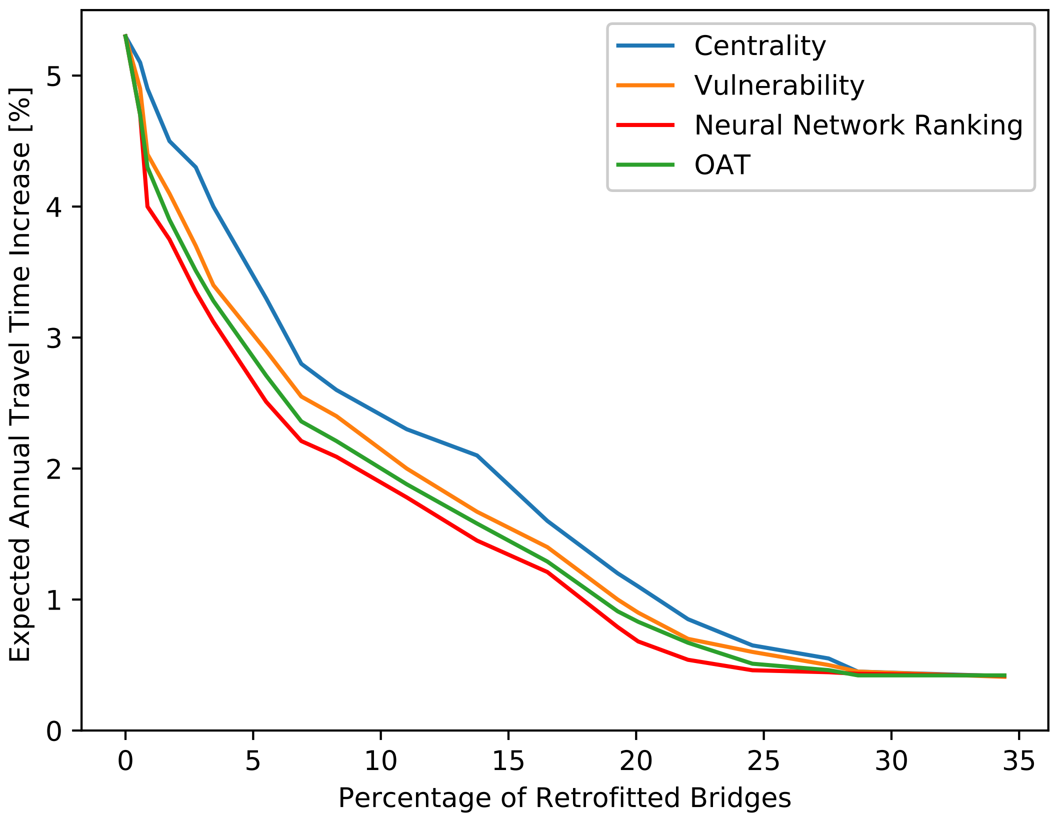

Fig. 9. Expected annual increase in traffic performance metric for different retrofitting strategies as a function of the number of retrofitted bridges.

This study next explores the characteristics of the bridges that the LIME-TI algorithm ranks highly. There was no strong correlation between the LIME-TI ranking and the traffic capacity of the bridges or their structural vulnerability, separately. However, some correlations can be observed with the bridges selected by LIME-TI when using combined traffic capacity and structural vulnerability measures. This correlation exists because some bridges may have an important role in the network, but they may not be very vulnerable and contribute significantly to the seismic risk. On the other hand, the most vulnerable bridges in the network may not have an essential role in traffic performance; therefore, retrofitting those bridges is not effective in minimizing the impacts of earthquakes. The LIME-TI selection of important bridges comes from the magnitude of coefficients C_{jk}, which detects the vulnerability of bridges by the occurrence of bridge damage on the individual realizations, combined with the occurrence rates used to aggregate different realizations in Equation (9). Motivated by this correlation, rankings based on combinations of OAT and vulnerability were also evaluated, but even combinations of these factors did not outperform the LIME-TI ranking.

Besides correlations, the features that make the bridges proposed by LIME-TI improve the network's performance better than other strategies were also explored. The first bridges selected by LIME-TI have a more significant individual contribution to the expected annual increase of traffic

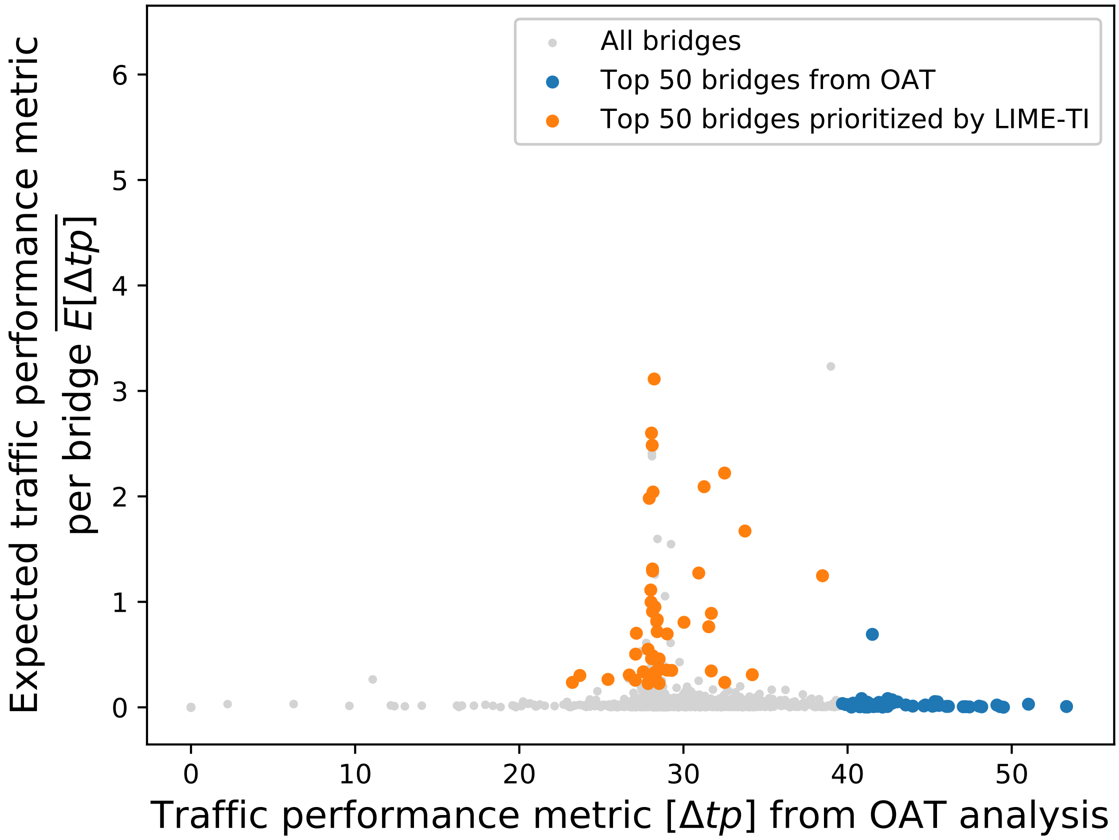

Fig. 10. Scatter plot of traffic performance from OAT analysis versus expected traffic performance metric, for all bridges in the network. The top 50 bridges ranked by OAT are shown in blue and the top 50 bridges ranked by LIME-TI are shown in orange.

performance metric of the traffic network. Figure 10 shows on the X-axis the change in traffic performance when each bridge is damaged individually, which is defined in previous sections as the OAT analysis. The Y-axis shows the expected change in traffic performance metric per bridge as defined in Equation (10).

$$\overline{E[\Delta tp]}_j = \sum_k w_k x_{jk} \overline{TP_{jk}} \tag{10}$$

Where:
$\overline{E[\Delta tp]}$ = Expected change in traffic performance metric associated to bridge j .
w_k = Rate of occurrence of damage map k.
x_{jk} = Indicator of damage to bridge j, in damage map k. = 1 if the bridge is damaged, 0 otherwise.
$\overline{TP_{jk}}$ = Contribution of bridge j to traffic performance metric change in scenario k, defined in Equation (11).

$$\overline{TP_{jk}} = \frac{\Delta tp_k}{nb_k} \tag{11}$$

Where:
$\overline{TP_{jk}}$ = Contribution of bridge j to traffic performance metric change in scenario k.
$\Delta tpki$ = Change in traffic performance metric for scenario k
nb_k = Number of bridges damaged in scenario k

Figure 10 shows that the bridges selected by LIME-TI may not individually cause substantial disruption to the network when they experience damage. However, they generate significant traffic disruption in the seismic scenarios considered on the training data, resulting from their vulnerability and importance to the network.

Additionally, the locations of the bridges selected by LIME-TI within the road network are examined. It is observed that LIME-TI bridges often correspond to "leading bridges." A leading bridge is a bridge that does not generate a significant change in tp when it experiences damage, but that leads or connects to a bridge that, when damaged, does generate a considerable change in tp. An OAT analysis did not identify these bridges since they are not the most important ones, and their ranking in OAT is lower than the one assigned by LIME-TI, which accounts for their vulnerability. Examples of leading bridges are shown in Figure 11, in which some of the top bridges selected by LIME-TI are shown with blue circles. These bridges are on highways that lead to two critical bridges in the San Francisco Bay Area, the Golden Gate Bridge and the Bay Bridge, which are not particularly vulnerable in our model.

NEURAL NETWORKS TO PREDICT SEISMIC RISK

The neural network of this study is a tool to inform engineering decisions. Thus, besides the evaluation of a neural network based on prediction errors, the model should predict seismic risk results used by policymakers. In particular, this work studies the model's accuracy at predicting loss curves and expected annual loss for the system.

The first seismic risk metric that the neural network should estimate accurately is a loss curve, as shown by Equation (5). Thus, the loss curves obtained using the traffic model described in the previous section are compared to the curves computed using the neural network. Figure 12 shows the loss curves for the case without retrofitting and for the case where 300 bridges are retrofitted based on their LIME-TI ranking. In both cases, the neural network can reproduce the curve generated by running the traffic model.

Figure 13 presents the expected annual increase of traffic performance computed by using the traffic model and by using the neural network model for different numbers of retrofitted bridges. The sets of bridges to retrofit correspond to those from the LIME-TI ranking, which coincide with those used in Figure 9, and the associated expected annual loss is computed using Equation (6). The prediction obtained from the neural network has almost no bias and has low error relative to the real value for all numbers of retrofitted bridges.

These results are significant because these risk metrics incur a high computational cost if obtained using the traffic model. Thus, the fact that the neural network is highly accurate allows reliable estimates of the effect of retrofitting bridges in a matter of seconds.

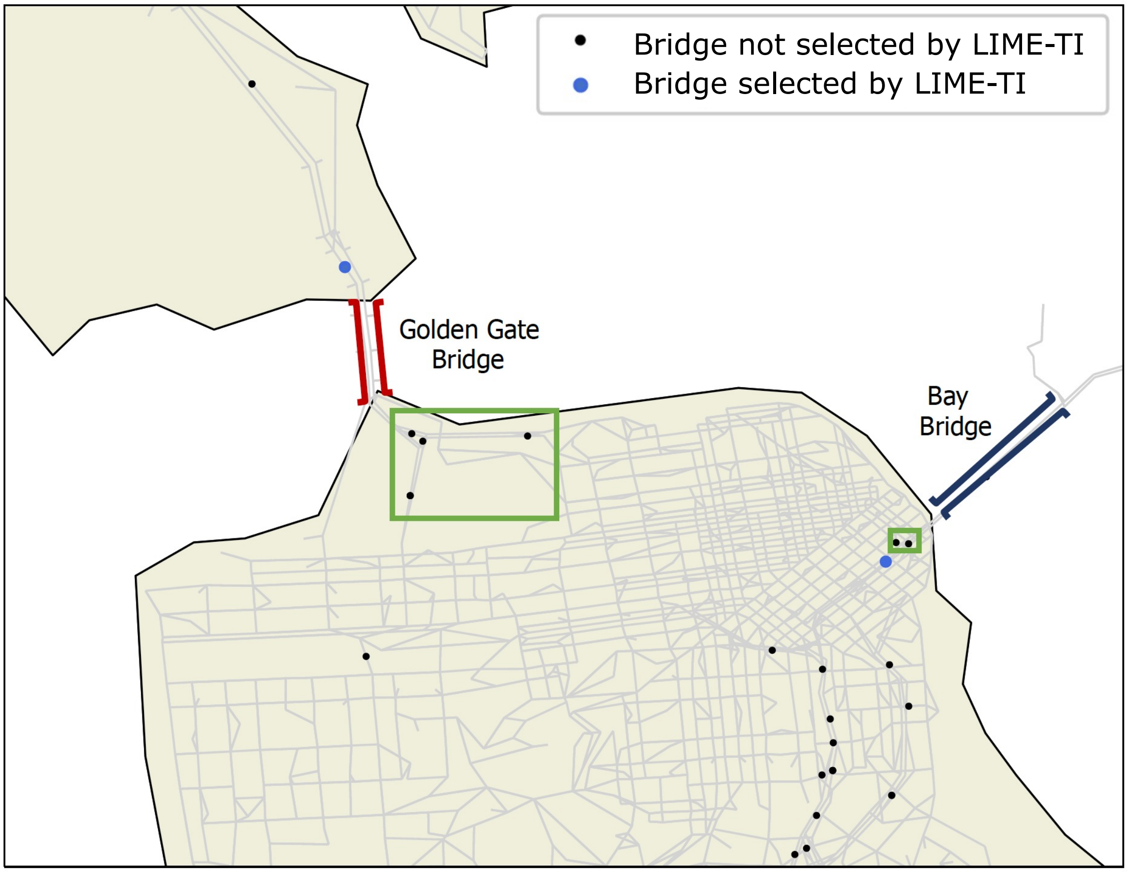

Fig. 11. A map of a portion of the testbed, illustrating "leading bridges." Bridges among the top 50 selected by LIME-TI are shown with blue dots. Other bridges are shown with black dots. Bridges that could be leading bridges but that are not vulnerable enough to be highly ranked are marked in a green box.

DISCUSSION

The results obtained in this study prove that neural networks can be extended to perform a seismic risk assessment of distributed systems in a rapid and accurate way. In terms of time savings, the neural network can perform analysis approximately 100,000 times faster than the original traffic model. This significant decrease of computational time did not impact the model's accuracy significantly, with the coefficient of determination of the model being $R^2 = 0.97$. Moreover, besides this statistical evaluation of the neural network, it was also observed that the surrogate model could predict practical parameters for decision-making. In particular, the use of the model allowed the fast computation of loss curves and expected annual losses.

Training neural networks to predict the performance of distributed systems introduces an alternate way to efficiently assess risk of critical infrastructure. Prior to this study, scenario reduction techniques, such as those proposed by Miller (2014) and (Han and Davidson 2012), have been used as methods to optimize computational resources without compromising the accuracy of the risk assessment. With the introduction of surrogate models such as the ones proposed in this work, the

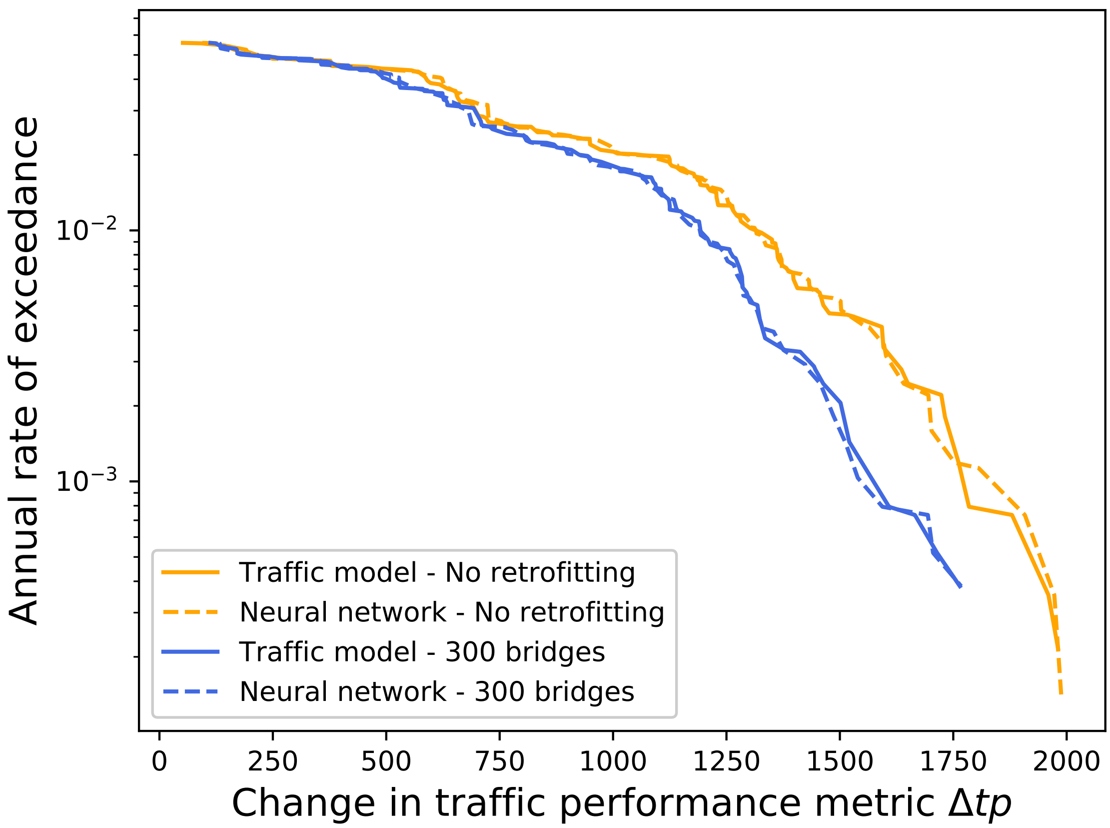

Fig. 12. Comparison of loss curves obtained using the neural network and the traffic model.

time required to analyze the consequences of each scenario is minimal, therefore limiting the need to reduce the number of scenarios included in the risk assessment.

From a practical point of view, these rapid and accurate surrogate models can be instrumental for decision-makers managing complex systems. The ability of the model to predict metrics such as expected annual loss or loss curves in a short time can facilitate the exploration of management strategies without requiring access to high-performance computing resources.

CONCLUSIONS

This study proposes a neural network surrogate model that accurately and rapidly estimates traffic disruption due to earthquakes. The neural network uses damage states of all bridges to predict a traffic performance metric and achieves a value of $R^2 = 0.97$ when compared with the traffic model in the testbed study, but with greatly reduced computational time. In addition, the neural network performs well based on risk measures such as loss curves and expected annual loss, which is relevant for evaluating the impact of retrofitting strategies.

In addition, the study proposes a new retrofitting strategy based on ranking bridges according to their role in the disruption to the road network. To quantify the role of the bridges, this work uses a modified version of a Local Interpretable Model-Agnostic Explanation (LIME), which is a

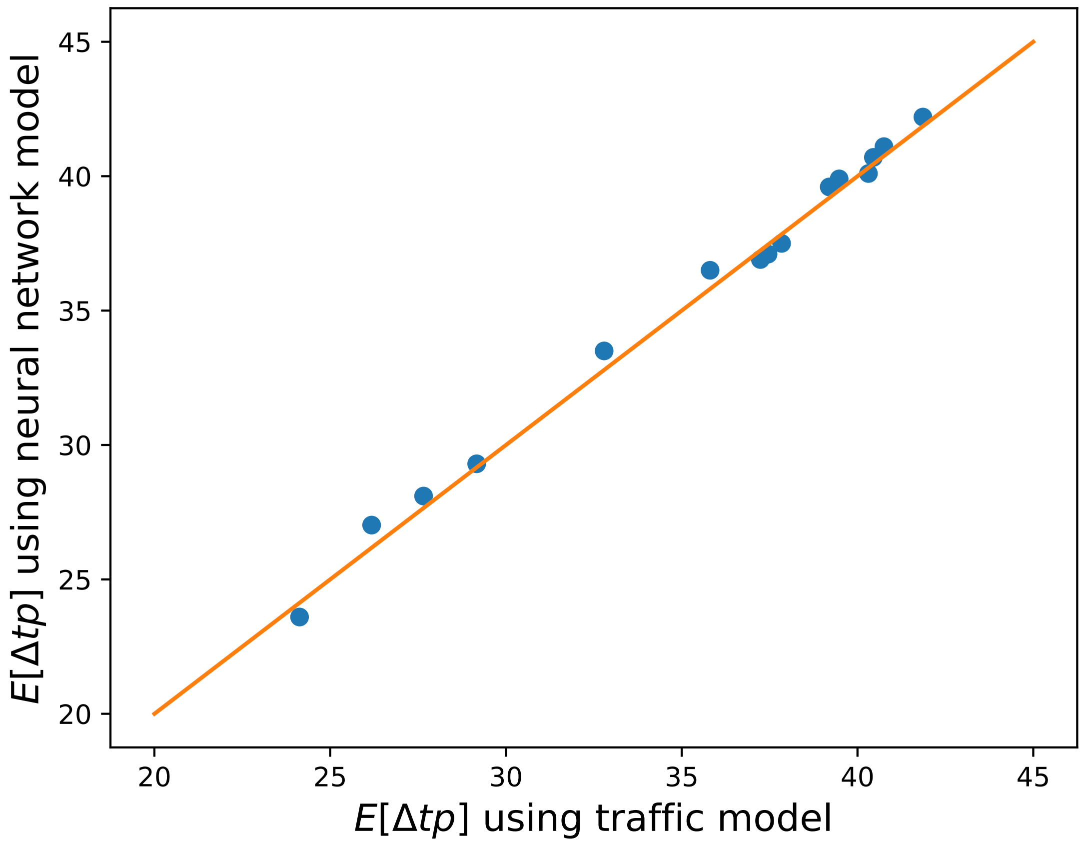

Fig. 13. Comparison of expected annual loss computed using the neural network and using the traffic model. Each point represents a different number of retrofitted bridges consistent with the sets of bridges used in Figure 9. Points with higher $E[\Delta tp]$ values are associated with fewer retrofitted bridges. The diagonal line indicates cases where the neural network results perfectly predict the traffic model result.

variable importance algorithm applicable to neural networks. In the modified version (LIME-TI), a new submodularity pick that accounts for the vulnerability of bridges and their importance in the performance of the road network is proposed. The proposed retrofitting strategy yields smaller increases in expected traffic disruption than other intuitive retrofitting strategies, such as ranking bridges by vulnerability or the volume of traffic they carry.

It is observed that some of the bridges selected by LIME-TI have a significant individual contribution to the seismic risk of the network. Moreover, some of these bridges are termed a leading bridge, a bridge whose sole impairment does not generate a significant disruption to the network but leads to a bridge with a significant role in the road network.

To construct the neural network, three different sampling protocols to generate training data are explored, and this study concludes that it is essential to incorporate enough highly disruptive events in the training data to predict these infrequent events accurately. This predictive capability is also important for the seismic management of road systems because decision-makers may rely on scenario-based approaches to explore the impact of different retrofitting strategies, which are usually selected to be highly disruptive. The authors also conclude that random sampling is less accurate than using hazard consistent or extreme events sampling.

The surrogate model requires a significant number of network performance simulations for training. However, the simulations are independent, and hence they can be performed in parallel. In addition, the objective of the neural network is to serve as a tool to evaluate retrofitting policies. As such, it still massively reduces the overall computational cost of the results shown above because risk metrics were computed for several retrofit permutations. Each retrofit case would take several hours or days of computational time to evaluate, whereas the trained neural network only takes fractions of a second. In terms of policy, the network could be trained by a research institution or consultant with access to high-performance computing and then transferred to decision-makers for use. The number of damage realizations required to train the network above was equivalent to that from evaluating the effectiveness of five retrofitting strategies. Thus, for any substantial number of retrofit strategy evaluations, the surrogate model will greatly reduce the overall computational cost. For instance, to obtain Figure 13, it takes 112.68 hours to run the original traffic model on a 2016 MacBook with a 2.6 GHz Quad-Core Intel Core i7 processor. The neural network takes 10.89 seconds to generate the same data on the same computer.

This study indicates several considerations for future applications on other systems that should allow the method to be generally valuable. First, properly trained neural networks can predict flows in infrastructure networks and reflect complex relationships between component damage and performance. Second, it is likely necessary to over-sample high-damage realizations of the network during training to avoid over-fitting small damage cases and failing to effectively predict the high-disruption cases of interest in risk management problems. Finally, a trained surrogate model offers tremendous utility in allowing decision-makers to explore mitigation strategies such as component retrofit by capturing complexities of the network performance at a feasible computational cost.

DATA AVAILABILITY STATEMENT

Some data, models, or code generated or used during the study are available in a repository online in accordance with funder data retention policies, which can be found in: DOI `10.5281/zenodo.5161259`.

ACKNOWLEDGMENTS

The State of California supported this work through the Transportation System Research Program of the Pacific Earthquake Engineering Research Center (PEER) and by the Shah Family Fund Fellowship. Some of the computing for this project was performed on the Sherlock cluster. The authors would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect those of the funding agency.

REFERENCES

Alemzadeh, S., Talebiyan, H., Talebi, S., Duenas-Osorio, L., and Mesbahi, M. (2020). "Deep learning-based resource allocation for infrastructure resilience." *arXiv preprint arXiv:2007.05880*.

Basoz, N. and Kiremidjian, A. S. (1995). *Prioritization of bridges for seismic retrofitting*. National Center for Earthquake Engineering Research.

Beckmann, M., McGuire, C. B., and Winsten, C. B. (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.

Bommer, J., Scott, S., and Sarma, S. (2000). "Hazard-consistent earthquake scenarios." *Soil Dynamics and Earthquake Engineering*, 19(4), 219–231.

Bommer, J., Spence, R., Erdik, M., Tabuchi, S., Aydinoglu, N., Booth, E., Del Re, D., and Peterken, O. (2002). "Development of an earthquake loss model for turkish catastrophe insurance." *Journal of seismology*, 6(3), 431–446.

Boore, D. M. and Atkinson, G. M. (2008). "Ground-motion prediction equations for the average horizontal component of pga, pgv, and 5%-damped psa at spectral periods between 0.01 s and 10.0 s." *Earthquake Spectra*, 24(1), 99–138.

Caltrans, S. (2019). "Caltrans seismic design criteria." *California Department of Transportation, Sacramento, CA*.

Chen, M. and Alfa, A. S. (1991). "A network design algorithm using a stochastic incremental traffic assignment approach." *Transportation Science*, 25(3), 215–224.

Elhag, T. M. and Wang, Y.-M. (2007). "Risk assessment for bridge maintenance projects: neural networks versus regression techniques." *Journal of computing in civil engineering*, 21(6), 402–409.

Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., Weldon, R., et al. (2009). "Uniform california earthquake rupture forecast, version 2 (ucerf 2)." *Bulletin of the Seismological Society of America*, 99(4), 2053–2107.

Field, E. H., Jordan, T. H., and Cornell, C. A. (2003). "Opensha: A developing community-modeling environment for seismic hazard analysis." *Seismological Research Letters*, 74(4), 406–419.

Gomez, C. and Baker, J. W. (2019). "An optimization-based decision support framework for coupled pre-and post-earthquake infrastructure risk management." *Structural Safety*, 77, 1–9.

Gomez, C., Sanchez-Silva, M., Dueñas-Osorio, L., and Rosowsky, D. (2013). "Hierarchical infrastructure network representation methods for risk-based decision-making." *Structure and Infrastructure Engineering*, 9(3), 260–274.

Han, Y. and Davidson, R. A. (2012). "Probabilistic seismic hazard analysis for spatially distributed infrastructure." *Earthquake Engineering & Structural Dynamics*, 41(15), 2141–2158.

Jayaram, N. and Baker, J. W. (2009). "Correlation model for spatially distributed ground-motion intensities." *Earthquake Engineering & Structural Dynamics*, 38(15), 1687–1708.

Kingma, D. P. and Ba, J. (2014). "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*.

Lim, H.-W., Song, J., and Kurtz, N. (2015). "Seismic reliability assessment of lifeline networks using clustering-based multi-scale approach." *Earthquake Engineering & Structural Dynamics*, 44(3), 355–369.

Liu, C., Fan, Y., and Ordóñez, F. (2009). "A two-stage stochastic programming model for transportation network protection." *Computers & Operations Research*, 36(5), 1582–1590.

Lundberg, S. and Lee, S.-I. (2017). "A unified approach to interpreting model predictions." *arXiv preprint arXiv:1705.07874*.

Miller, M. (2014). "Seismic risk assessment of complex transportation networks." Ph.D. thesis, Stanford University, Stanford, CA.

Miller, M. and Baker, J. (2015). "Ground-motion intensity and damage map selection for probabilistic infrastructure network risk assessment using optimization." *Earthquake Engineering & Structural Dynamics*, 44(7), 1139–1156.

Miller-Hooks, E., Zhang, X., and Faturechi, R. (2012). "Measuring and maximizing resilience of freight transportation networks." *Computers & Operations Research*, 39(7), 1633–1643.

MR, H.-M. (2003). "Multi-hazard loss estimation methodology: Earthquake model." *Department of Homeland Security, FEMA, Washington, DC*, 235–260.

Nabian, M. A. and Meidani, H. (2018a). "Accelerating stochastic assessment of post-earthquake transportation network connectivity via machine-learning-based surrogates." *Transportation Research Board 97th Annual Meeting*, Washington, DC.

Nabian, M. A. and Meidani, H. (2018b). "Deep learning for accelerated seismic reliability analysis of transportation networks." *Computer-Aided Civil and Infrastructure Engineering*, 33(6), 443–458.

Olden, J. D., Joy, M. K., and Death, R. G. (2004). "An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data." *Ecological modelling*, 178(3-4), 389–397.

Özdamar, L. and Demir, O. (2012). "A hierarchical clustering and routing procedure for large scale disaster relief logistics planning." *Transportation Research Part E: Logistics and Transportation Review*, 48(3), 591–602.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). ""why should I trust you?": Explaining the predictions of any classifier." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, 1135–1144.

Rokneddin, K., Ghosh, J., Dueñas-Osorio, L., and Padgett, J. E. (2013). "Bridge retrofit prioritisation for ageing transportation networks subject to seismic hazards." *Structure and Infrastructure Engineering*, 9(10), 1050–1066.

Tomar, A. and Burton, H. V. (2021). "Active learning method for risk assessment of distributed infrastructure systems." *Computer-Aided Civil and Infrastructure Engineering*, 36(4), 438–452.

Wald, D. J. and Allen, T. I. (2007). "Topographic slope as a proxy for seismic site conditions and amplification." *Bulletin of the Seismological Society of America*, 97(5), 1379–1395.

Werner, S. D., Taylor, C. E., Cho, S., Lavoie, J.-P., Huyck, C. K., Eitzel, C., Chung, H., and Eguchi, R. T. (2006). "Redars 2 methodology and software for seismic risk analysis of highway systems." *Report no.*

Xie, Y., Ebad Sichani, M., Padgett, J. E., and DesRoches, R. (2020). "The promise of implementing machine learning in earthquake engineering: A state-of-the-art review." *Earthquake Spectra*, 36(4), 1769–1801.

Yu, H., Wu, Z., Wang, S., Wang, Y., and Ma, X. (2017). "Spatiotemporal recurrent convolutional networks for traffic prediction in transportation networks." *Sensors*, 17(7), 1501.