Geyin, M. and Maurer, B.W. (2023). "U.S. National VS30 Models and Maps Informed by Remote Sensing and Machine Learning." *Seismological Research Letters,* 94(3): 1467-1477, doi: 10.1785/0220220181.

U.S. National V_{S30} Models and Maps Informed by Remote Sensing and Machine Learning

Geyin, M.[1] and Maurer, B.W.[2]

Abstract: The shear-wave velocity time-averaged over the upper 30 m (V_{S30}) is widely used as a proxy for site effects, forms the basis of seismic site class, and underpins site-amplification factors in empirical ground-motion models. Many earthquake simulations therefore require V_{S30}. This presents a challenge at regional scale, given the infeasibility of subsurface testing over vast areas. While various models for predicting V_{S30} have thus been proposed, the most popular U.S. national, or "background," model is a regression equation based on just one variable. Given the growth of community datasets, satellite remote sensing, and algorithmic learning, more advanced and accurate solutions may be possible. Towards that end, we develop national V_{S30} models and maps using field data from over 7,000 sites and machine learning (ML), wherein up to 17 geospatial parameters are used to predict subsurface conditions (i.e., V_{S30}). Of the two models developed, that using geologic data performs marginally better, yet such data is not always available. Both models significantly outperform existing solutions in unbiased testing and are used to create new V_{S30} maps at ~220 m resolution. These maps are updated in the vicinity of field measurements using regression kriging and cover the 50 U.S. states and Puerto Rico. Ultimately, and like any model, performance cannot be known where data is sparse. In this regard, alternative maps that use other models are proposed for steep slopes. More broadly, this study demonstrates the utility of ML for inferring below-ground conditions from geospatial data, a technique that could be applied to other data and objectives.

Introduction

Subsurface seismic-wave velocities (e.g., shear-wave velocity, V_S) affect the amplitude, duration, and frequency content of ground motions. Measurements or estimates of these velocities are thus needed to predict

[1] Norwegian Geotechnical Institute, Houston, TX USA
[2] Department of Civil and Environmental Engineering, University of Washington, Seattle, WA USA

ground motions and, by consequence, earthquake impacts. Ideally, these velocities would be obtainable: (i) quickly (i.e., by time- and cost-efficient means); (ii) at high spatial resolution (e.g., consistent with the scale at which subsurface velocities change); and (iii) over the spatial extents that experience strong motion (e.g., a metropolitan region). Problematically, state-of-practice methods for measuring V_S typically result in discrete, 1D V_S-profiles that require considerable time and cost. As a result, it is infeasible to measure V_S over vast areas, as would be required for regional earthquake simulations. Even in cases where V_S is needed for important, site-specific purposes (e.g., at seismic-recording stations, to develop empirical ground motion models, or GMMs), it is often the case that V_S is estimated, rather than measured (e.g., Ahdi et al., 2017).

Accordingly, efforts have been made to predict V_S profiles remotely (e.g., Boore and Joyner, 1997; Holzer et al., 2005; Wald and Allen, 2007; Castellaro et al., 2008; Boore et al., 2011; Thompson et al., 2014; Parker et al., 2017; Foster et al., 2019; Yu, 2021). These efforts have mostly focused on predicting the time-averaged V_S in the upper 30 m (V_{S30}), which: (i) is widely used as a proxy for site effects; (ii) forms the current basis of seismic site class; (iii) underpins site-amplification functions (e.g., Stewart et al., 2017); and (iv) is a required input to all modern empirical GMMs. V_{S30} thus serves an important role in regional earthquake simulations, post-earthquake data products (e.g., Worden et al., 2010), site-specific hazard analyses, and indirectly, the National Seismic Hazard Model (Petersen et al., 2019), given the need for V_{S30} at strong-motion stations when developing GMMs. At present, a patchwork of V_{S30} models is used in the U.S., with the national "background," model adopted by the U.S. Geological Survey (Heath et al., 2020) being a regression equation with one input – topographic slope (e.g., Wald and Allen, 2007; Allen and Wald, 2009). The underlying, seminal concept – that flat ground tends to be soft and steep ground tends to be hard – is quite useful, but also often inefficient and/or insufficient for predicting V_{S30}. Several regional models have thus aimed to improve on this approach, generally by using higher-resolution elevation models, more advanced statistical schemes, and/or by binning the data on mapped geology (e.g., Ahdi et al., 2017; Wills and Clahan, 2006; Thompson et al., 2014; Li and Rathje, 2020). Considering the growth of community geophysical datasets, satellite remote sensing, and algorithmic learning, more advanced and accurate solutions may yet be achievable, both at national and regional scales.

Toward that goal, this paper develops U.S. national V_{S30} models and maps using machine learning (ML), wherein 17 above-ground geospatial variables are used to predict below-ground V_{S30}. Examples of geospatial predictor variables, which are obtained from remote-sensing and existing mapped information, include topographic slope and various topographic indices; distance to rivers, streams, and other water bodies; and various values describing or predicting geology, hydrology, lithology, climate, etc. While such predictors lack mechanistic links to V_{S30}, they correlate in complex and interconnected ways – an ideal application for ML. Although the concept of a "geospatial" V_{S30} model is not new – all existing models could be described this way – neither algorithmic learning nor a large quantity of predictors has previously been used (whether at national or regional scale). In this regard, accurate prediction of subsurface conditions likely requires many variables, as suggested by Iwahashi et al. (2010), Yong et al. (2012), and Zhu et al. (2015; 2017) (i.e., more than topographic slope), but traditional regression requires hypotheses of what is believed to matter and how, limiting the number of variables easily modeled. Because such beliefs are unnecessary with ML, it can provide learning insights that are unlikely, if not infeasible, with traditional techniques. The adopted approach thus allows for a large body of predictive information to be utilized, with more potential for that information to be exploited. In the following, the data and methodology are first described, after which the trained ML models are compared via unbiased tests against the national "background" model of Wald and Allen (2007) and Allen and Wald (2009), as implemented by Heath et al. (2020a,b). For brevity, we refer to this slope-based model as Allen and Wald (2009), or AW09. The resulting map products, which are updated in the vicinity of field measurements using regression kriging, are then presented.

Data and Methodology

A total of 7,081 V_{S30} measurements were selected for analysis, as mapped in Figure 1 for the contiguous U.S. Not shown are 24 measurements in Hawaii, 23 in Puerto Rico, and 15 in Alaska. While these data represent a range of geographic and geologic settings, they are biased toward densely populated, high-seismicity regions where there is greater need for V_S data. As a result, some U.S. states are unrepresented in model training and testing, a limitation that is shared by all existing national models. In addition, and as will be discussed further,

the compiled data are biased with respect to both topographic slope, with <10% of measurements made on terrain with >3˚ slope, and soil sites, with <10% of measurements having V_{S30} > 760 m/s.

Significant V_{S30} data sources included the McPhillips et al. (2020), Parker et al. (2017), and Ahdi et al. (2017) compilations. The authors also computed and added V_{S30} values from profiles unrepresented in other large compilations. This augmented the available V_{S30} data by 1,021 points. Larger sources of such data included Kayen et al. (2011), Salomone et al. (2012), and Kwak et al. (2021). In computing V_{S30} from V_S profiles, the extrapolation method of Boore et al. (2011) was applied to profiles that did not reach 30 m depth. While this increases the measurement uncertainty at certain sites, it was deemed acceptable, given the incomplete coverage of V_{S30} data at national scale. Of the compiled data, 80% was randomly selected for model training and the remaining 20% was held for unbiased testing. While the definition of a truly unbiased test is debatable (e.g., test sites are occasionally located near training sites), it should be noted that the AW09 model against which comparisons will be made was originally trained using much of the data compiled herein for testing. As a result, the ensuing tests are likely biased in favor of the existing AW09 model. Finally, it must be emphasized that empirical models can be particularly unreliable when encountering unfamiliar regions or features. The limits and resolution of each predictor variable – introduced below – should thus be understood by users.

In the current effort, either 15 or 17 predictor variables were compiled at the sites of V_{S30} data. These consisted of: the depths to (1) bedrock and (2) groundwater, as predicted by geospatial models trained on ~1.6 million global field measurements; the mapped (3) geologic unit and (4) consolidation state; the (5) classified geomorphologic phonotype (consisting of landforms that include valley, depression, hollow, footslope, flat, and others); the measured (6) distance-to-river, (7) compound topographic index (which describes the hydrologic environment), and (8) topographic slope; the (9) profile curvature and (10) tangential curvature; the computed (11) topographic position index, (12) roughness, (13) terrain ruggedness index, and (14) vector ruggedness measure (which collectively describe the profile and heterogeneity of the surface terrain); and lastly, the geomorphologic landform's (15) Shannon diversity index, (16) uniformity, and (17) entropy (which collectively describe the diversity and spatial distribution of geomorphons in a sample area). The range,

resolution, and source of each variable are in Table 1; the reader is referred to these respective sources for methodological details and background information on all 17 variables. Supplemental Table S1 also provides additional descriptions and citations for each variable. The goal of these variables, which predominantly use above-ground data, is to predict below-ground conditions.

Except for the geologic unit and consolidation state, which were sampled from the Horton et al. (2017) U.S. national geologic map compilation, all variables are continuously available in North America, and in many cases, have global coverage. While surface geology ultimately resulted in a marginally better model, the Horton et al. (2017) compilation does not include Alaska, Hawaii, or Puerto Rico. Additionally, it will be shown that undesirable transitions occur at a few state boundaries, where differences in the state source maps result in different mapped geologies on either side of a state line, and by corollary different V_{S30}. For these reasons we ultimately present two map products – one that includes mapped surface geology ("Model 1", which performs slightly better), and one that does not ("Model 2"). The geologic unit is also a unique feature in that it was reclassified, whereas all other variables were used directly as sampled. Specifically, we: (i) grouped all sedimentary, igneous, and metamorphic rock units; and (ii) of the remaining units applicable to soils, selectively excluded those sparsely populated with V_{S30} data. In this regard, sites that do not map as either a type of rock or as alluvial, fluvial, glacial, lacustrine, peat, or terrace deposits are implicitly treated as general, unknown soil deposits; additional classification details are provided in supplemental Table S2. In addition to the predictors in Table 1, several others, including annual precipitation (Fick and Hijmans, 2017), distance to coastline (NASA, 2020), deposit age (Horton et al. (2017), and regional flags (e.g., Western US vs. Eastern US) did not improve performance and were not adopted. This apparent lack of utility could potentially change for some variables if more field data were available. The futility of geologic age and regional flags, for example, might be explained by the lack of V_{S30} measurements at rock sites, particularly from seismic site class A (e.g., <1% of the data has V_{S30} > 1225 m/s).

Having compiled V_{S30} data and predictor variables, numerical predictors were BoxCox transformed (Box and Cox, 1964) and normalized to have values between 0 and 1 to reduce spurious interactions among predictors. Several ML techniques were used to train prospective models, including support vector machines

(e.g., Vapnik, 1995), Gaussian process regression (GPR) (e.g., Rasmussen, 2003), decision trees (e.g., Rokach and Maimon, 2008), and decision tree ensembles constructed by gradient boosting, bagging, or random forests (e.g., Breiman, 1996; Piryonesi and El-Diraby, 2021). Of the resulting models, those that are easier to interpret tend to have lesser accuracy and portability (e.g., an individual decision tree), while those that tend to perform best (e.g., tree ensembles) are more convoluted. Once promising techniques were identified, the internal parameters of those techniques (i.e., "hyperparameters') were optimized to minimize the prediction error. 5-fold cross-validation was used to evaluate and mitigate overfitting, as is common. The particulars of the developed models are further described in the following.

Results and Discussion

Using the training set and all 17 predictor variables (i.e., including surface geology), many provisional models were trained. Of these, three were adopted for optimization and testing. Two were ensembles of 200 decision trees each, where relatively weak decision tree models were combined to build a stronger model. When a decision tree is trained, recursive decision forks are formed, such that a specific combination of model inputs maps to an expected output. However, because an individual tree is typically neither accurate nor portable (i.e., it is prone to overfitting), trees are generally ensembled. This modeling approach, which is found in popular ML toolkits (e.g., TensorFlow, Scikit, PyTorch), is reviewed by Friedman (2001) and practically demonstrated in detail by Elith et al. (2008). A primary distinction of tree ensembles is how the individual models are trained and combined. In this regard, "bagging" and "boosting" were respectively employed to develop the two tree ensembles. In bagging (also referred to as bootstrap aggregating), numerous versions of the training set are formed via bootstrap sampling, with each used to train a decision tree, and the predictions from the various trees are aggregated to make a final prediction. Given this resampling and averaging, bagging tends to minimize the prediction variance and reduce overfitting, relative to other ensembling methods. In boosting, a sequence of decision trees is built from weaker trees, wherein each tree attempts to learn from the prior trees by increasing the weight on observations that were poorly predicted. In this way, the most difficult cases are emphasized, such that subsequent models focus on them more. In contrast to bagging, the models

148 that perform best are weighted most. While boosting is slow, it may maximize accuracy relative to other
149 ensembling techniques, albeit at the expense of overfitting (Piryonesi and El-Diraby, 2021).

150 The last of the three adopted models was a GPR model. In contrast to other ML techniques that learn exact
151 values, both for a model's parameters and for its output, GPR infers probability distributions via the Bayesian
152 approach and is nonparametric. An important ingredient of GPR models is the prior assumption, or kernel
153 (also called the covariance function in the context of GPR), which describes how a model's predictions are
154 related, given different inputs. We ultimately adopted a squared exponential kernel function, which is the
155 default in many ML toolkits (e.g., Duvenaud, 2014), and which results in a "smooth" model, rather than one
156 in which non-differentiable behavior (e.g., multilinearity) is permitted. Benefits of GPR include the ability to
157 impart judgment via the kernel and its intrinsic use of interpolation, which makes GPR relatively less reliant
158 on a large dataset. On the random test set (i.e., the 20% of V_{S30} data held from training), the bagged ensemble,
159 boosted ensemble, and GPR models had respective mean absolute errors (MAEs) of 112 m/s, 118 m/s, and
160 110 m/s, whereas the AW09 model had an MAE of 171 m/s. This represents an average improvement of 34%.
161 The mean square errors (MSEs) suggest larger improvements, with the three respective models reducing MSE
162 by 52%, 51%, and 50%, relative to AW09.

163 Finally, while the three adopted models perform well individually, we used "meta-learning" to combine
164 them (Dzeroski and Zenko, 2004). Also known as "stacking", this approach recognizes that the base models,
165 which were each developed using different approaches, may be more (or less) effective in different situations.
166 The GPR model, for example, has the lowest MAE but the largest MSE, meaning that it prioritizes accuracy
167 at the expense of some large outliers. Stacking can result in a meta-model that performs better than any base
168 predictor and which is more stable (i.e., it avoids large swings on account of which model is chosen). While
169 stacking refers to a specific ML technique, the basic concept is ubiquitous in natural hazards modeling (e.g.,
170 ensembling of ground motion or hurricane models in a logic tree). Starting with the three base models, the
171 training set was again partitioned for 5-fold cross-validation. The out-of-fold predictions (i.e., the validation
172 data) were then used to train the meta-model using a bagging algorithm. In other words, the base models were
173 optimally coalesced through analysis of their out-of-fold predictions. The resulting meta-model, henceforth

termed "Model 1," achieved an MAE of 108 m/s on the unbiased test set and reduced the MSE by 55% relative to AW09. While these *additional* improvements are minor, the generalization that results from stacking could provide other, unrealized benefits during forward application. The overall improvement relative to AW09 is summarized in Table 2, which compiles MAE values binned on V_{S30} and topographic slope. Model 1 has lower MAE across all V_{S30} and all slopes, but especially for $V_{S30} < 180$ m/s and $537 < V_{S30} < 2000$ m/s. This may be attributable to: (i) AW09's truncation of low V_{S30} predictions at 180 m/s; and (ii) the predictors used by Model 1 (e.g., geology) that help to distinguish when relatively flat ground is rock rather than soil, where the latter is the default assumption of slope-based models.

Plotted in Figure S1 of the electronic supplement are measured vs. predicted V_{S30} values for the compiled dataset, both for Model 1 and AW09. The corresponding prediction residuals, defined as r = ln (observed/predicted), are in Figure 2. Also shown via green lines are the residual standard deviations, computed as 0.218 and 0.555 for Model 1 and AW09, respectively. Model 1 residuals are thus less dispersed (e.g., $R^2 = 0.72$ vs. 0.02) and minimally biased, whereas AW09 tends to overpredict lower V_{S30} values and underpredict higher V_{S30} values. It can similarly be shown that the Model 1 residuals are unbiased with respect to each input variable. In this regard, residuals are plotted vs. each numerical input in Figure S2. Collectively, the results suggest that Model 1 warrants adoption and further evaluation as a national background model.

While simplified interpretations of model structure are often infeasible with ML (i.e., relative to traditional regression), insights can be gained via the computed predictor importance (e.g., Auret and Aldrich, 2011), which may be interpreted as each variable's relative contribution to the accuracy of a model. Accordingly, the relative importance of each variable was computed and is plotted in Figure 3, where variables are sorted from most to least important. This approach to model interpretation was also used by Durante and Rathje (2021) and Geyin et al. (2022), who developed ML models for liquefaction-induced ground failure. The most influential variables in Model 1 include the predicted depth to bedrock, measured topographic slope, three different indices of surface roughness, and the mapped geologic unit and geomorphologic phonotype. These predictors are ~3-5 times more influential than the least important variable – distance to river. These results also reflect both the utility and insufficiency of topographic slope, which is useful, but which alone cannot

predict when flat ground is relatively hard or when sloping ground is relatively soft. Lastly, these results have important implications for forward mapping, given that spatial biases or discontinuities in important variables (e.g., a mispredicted depth to bedrock or surface geology) can be expected to cause similar problems in the predicted V_{S30}.

Using Model 1, V_{S30} predictions were next mapped throughout the contiguous US, wherein regression kriging (Hengl et al., 2007) of model residuals was used to update predictions in the vicinity of measurements (i.e., to bring them into agreement). With this approach, which was used by Thompson et al. (2014) to map V_{S30} in California, a model trained on various predictors (i.e., "regression") is combined with spatial interpolation of that model's residuals (i.e., "kriging"). Thus, the residuals are predicted at unsampled locations using nearby measurements (where residuals are known) and are used to update the model's predictions in the vicinity. Defining the residuals as r = ln (observed/predicted), which pass the Lilliefors (1967) test for normality, an exponential semivariogram model was selected for its best fit of the data:

$$Semivariance\ (h) = c_0 \left(1 - e^{-h/a}\right) \quad (1)$$

Where c_0 and a are respectively the semivariogram sill and range, defined as c_0 = 1.1576 and a = 4.7667, and h is the separation distance between locations. This semivariogram and its fit of the empirical data are shown in Figure S3. Using this information, which describes spatial correlation, residuals were predicted nationally. As a representative example, the krigged residuals are shown in Figure 4 for the Puget Sound region of Washington State. Predicted residuals approach the computed residual at sites of V_{S30} measurement and attenuate with distance toward zero (i.e., the model's mean residual). The rate of this attenuation is controlled by the semivariogram in Eq. (1). Similarly, the standard deviation of the krigged residual approaches zero at measurement locations (reflecting the "known" error) and increases to σ = 0.218 (i.e., the overall model uncertainty) at locations far away. It should be noted, however, that the semivariance at a separation distance of zero (i.e., the nugget) is zero in Eq. (1), meaning that V_{S30} measurement uncertainties are not considered. Additionally, the nugget could contain small-scale spatial uncertainty (i.e., variability in V_{S30} over a distance

that is greater than zero, but smaller than the smallest separation distance in the dataset). In other words, V_{S30} may not be uniform over an individual cell/pixel in the resulting V_{S30} map. However, because the nugget is poorly constrained by the data (e.g., see Figure S3) and would require additional measurements and/or subjectivity to define, it is neglected in the present effort. Had the nugget been non-zero, the mapped V_{S30} values at sites of V_{S30} measurement would not equal the measured values (i.e., the ML model prediction would be given more weight) and the uncertainties would not be zero. This issue can and should be revisited in the future. Measurement uncertainties, for example, could be assigned via regression kriging or the multivariate normal method (Worden et al., 2018; Foster et al., 2019). The primary benefit of the latter is that it allows for site-specific uncertainty assignments, although this would require a rigorous, judgment-based analysis of the more than 7,000 V_{S30} measurements.

A national V_{S30} map was next created by computing the product V_{S30}*exp(r), where V_{S30} is the prediction from Model 1 and r is the krigged residual. This process scales the prediction up or down in the vicinity of measurements, thereby correcting for local or sub-regional prediction bias (e.g., where V_{S30} is mispredicted at a site or across a city). It should be noted, however, that biases at larger scales (e.g., state-scale) were not observed. As a representative example, the resulting krigged V_{S30} map is shown in Figure 5 for the Puget Sound and is compared to AW09. Aside from local V_{S30} discrepancies, the most notable difference is the shift in predicted V_{S30} across mountainous terrain, with AW09 consistently predicting higher V_{S30} on steeper slopes.

As previously mentioned, the compiled dataset is biased toward sites that are flatter and softer, with very few measurements in mountainous terrain. Plotted in Figure S4 of the electronic supplement, for example, is the cumulative distribution of the compiled data with respect to slope, which indicates that ~5% of measurements are from sites >5˚ slope. This is because V_{S30} is of greatest interest where infrastructure exists (flatter ground), and where subsurface conditions have the potential to alter ground motions (soil sites). Although the test data from steeper terrain and harder sites indicate that Model 1 outperforms AW09 (see Table 2), these predictions should nonetheless be viewed skeptically, given the paucity of data. Our interpretation is that the AW09 model was based on judgement for steeper slopes and was not strictly a regression of the V_{S30} data then available. Inspection of the AW09 data, for example, reveals that the AW09

model tended to severely overpredict the measured V_{S30} values for slopes exceeding ~6˚. This departure from the data – seemingly driven by a belief that V_{S30} values must be higher than suggested by the limited data – likely explains the difference across mountainous terrain illustrated in Figure 5. Whereas AW09 purposefully mispredicted the measured values on steeper slopes, the models developed herein do not. Mirroring AW09, however, it is also our judgement that Model 1 generally underpredicts V_{S30} on steep slopes. While predictions on steeper slopes are generally less consequential for engineering purposes, we created an alternative map termed "Model 1alt." Here, Model 1 is heuristically blended with AW09 using a weighting scheme in which Model 1 predictions are adopted for slopes ≤ 5˚, AW09 predictions are adopted for slopes ≥ 10˚, and otherwise:

$$V_{S30} = Model\ 1 * \left(-\frac{1}{5} * (slope) + 2\right) + AW09 * \left(\frac{1}{5} * (slope) - 1\right) \quad (2)$$

where V_{S30} is the Model 1alt prediction and slope is measured in degrees. This scheme is based on the data available for analysis (see Figure S4) but is ultimately subjective. While the performance of Model 1alt is slightly less than that of Model 1, it is our judgement that the blended predictions are more reasonable across the full domain of topographic slope, given that steep slopes are almost necessarily comprised of rock. Ultimately, additional measurements are needed from steeper slopes, or else judgement must continue to be relied upon. An example of the Model 1alt map is shown in Figure 6 for the Puget Sound. Both the original and alternative maps can be downloaded as ~220 m resolution geotiff files from Geyin and Maurer (2022) (see Data and Resources) and provide continuous coverage of the contiguous U.S.

Although the developed model, with 17 predictors, performs better than any other on the training and unbiased test data, it: (i) covers only the contiguous U.S., given the extents of the Horton et al. (2017) geology compilation; and (ii) results in discontinuities at a few state borders, an example of which at the Nebraska-Kansas border is shown in Figure S5. While the first of these problems could be rectified by augmenting the Horton et al. (2017) national compilation with additional maps, the latter problem, which results in minor but unreasonable shifts in the predicted V_{S30}, would be resolved only through a rigorous reinterpretation of the state source maps. Given these problems, the preceding effort for Model 1 was repeated without surface

geology (i.e., the mapped geologic unit and consolidation state). The resulting model, henceforth termed "Model 2," achieved an MAE of 115 m/s on the unbiased test set (vs. 108 m/s by Model 1 and 171 m/s by AW09) and reduced the MSE by 49% relative to AW09 (vs. a 55% reduction by Model 1). Thus, while surface geology is useful, Model 2 provides a serviceable alternative given the stated limitations. Following the same methodology, the relative predictor importance was computed for Model 2 and is shown in Figure S6. The ranking of variables is very similar to Model 1 (see Figure 3), except for: (i) the absence of surface geology; and (ii) a slight upward shift in the importance of groundwater depth, which suggests that it provides additional utility in the absence of geologic mapping. This is unsurprising, given that surface geology and groundwater depth are correlated under certain conditions.

Analogous to Figure 2, the Model 2 prediction residuals, defined as $r = \ln(\text{observed/predicted})$, are plotted in Figure S7 and have a standard deviation of 0.264 (vs. 0.218 for Model 1 and 0.555 for AW09). Again, use of surface geology is beneficial, but a large improvement over slope-based methods is still achieved in its absence. Finally, following the prior methodology, two maps were created using regression kriging to update Model 2 in the vicinity of field measurements. The semivariogram defining the spatial correlation of Model 2 residuals is shown and defined in Figure S8. Using the weighting scheme given in Eq. (2) the Model 2 map was blended with AW09, such that predictions shift toward AW09 predictions at larger topographic slope, creating "Model 2alt." Both the original and alternative maps (Model 2 and Model 2alt) can be downloaded from Geyin and Maurer (2022) and provide continuous coverage of the 50 U.S. states and Puerto Rico at ~220 m resolution. While the Model 1 and Model 2 maps are provided for transparency into the modeling process, we recommend adoption and further testing of the Model 1alt and Model 2alt products, the first of which is mapped in Figure 7 for the contiguous U.S.

Limitations, Uncertainties, and Future Work

The developed models are inherently tied to the data compiled for analysis. While this is true of any model, it is especially true of empirical models, given the lack of mechanistic links between the prediction variables and target. ML models are unfortunately no exception. As with any empirical model trained by finite data,

299 correlations do not necessarily indicate causality and may not necessarily transfer to larger datasets. While
300 several techniques were used to mitigate overfitting, including the use of an unbiased test set, k-fold cross
301 validation, and model stacking, additional data is inevitably needed to confirm and/or improve model
302 portability. Thus, while Models 1 and 2 improve upon a national slope-based model in unbiased tests, their
303 performance in data-poor regions cannot be known. The models should be used cautiously in these locations
304 (e.g., Colorado, Florida, etc.), where the model uncertainty may exceed that suggested by the presented test
305 statistics, given that neither the training nor test data represent those locales. Nonetheless, the merits of the
306 presented approach and models, which warrant adoption and further testing alongside other solutions, are
307 arguably compelling. In the future, this approach could be improved in several ways.

308 First, it is obvious that more V_{S30} data is desirable, both for training and testing models, and for anchoring
309 model predictions at sites of measurement. There is also currently a strong imbalance towards sites that are
310 flatter and softer, where there is more interest in measuring V_{S30}. There is almost a complete lack of
311 measurements from seismic site class A, for example, which is problematic for the Central and Eastern U.S.
312 This imbalance results in models with questionable performance at very hard sites and on steep slopes. While
313 oversampling of underrepresented data was attempted during model development, there is so little data at high
314 V_{S30} and on steep slopes that resampling a very small amount of data a very large number of times produces
315 an overfit model. To circumvent this problem, we prefer judgment-based estimates on steep slopes, following
316 from AW09. Neither oversampling nor judgment are ideal, of course, and the problem ultimately awaits
317 additional data. Second, it is well known that ML (like any algorithmic, or "AI," learning technique) can make
318 strong models, but is generally weak in explaining the "why." It can be difficult, for example, to explain the
319 influences and interactions of variables, or the physical structure of the resulting model. This is particularly
320 true when multiple models are "stacked" to produce an ensemble that is more effective, but also more
321 convoluted. Thus, focused efforts to identify new geospatial variables that more efficiently and sufficiently
322 correlate to V_{S30} are warranted and could produce additional gains.

323 Third, the models rely on the accuracy and spatial resolution of inputs, some of which are themselves
324 predictions (e.g., depth to bedrock, surface geology). Mispredictions may therefore occur in the vicinity of

geomorphic transitions (e.g, at the base of a mountain, as in Salt Lake City, UT), where the resolution of input variables may not capture local conditions, or where one or more variables is inaccurate (e.g., among other examples, the unmapped presence of artificial fill, as in Seattle, WA). The adopted approach should therefore improve as the accuracy and resolution of the geospatial predictors improves. Fourth, the uncertainties of V_{S30} measurements, which are especially non-trivial for surface-wave inversion methods, were not included in the present effort but could be in the future, as could uncertainty more broadly (also neglected, for example, is the small-scale spatial variability than may occur across a map pixel). Fifth, regression kriging is one of several approaches for updating predictions with field data. Other methods (e.g., Worden et al., 2018; Foster et al., 2019) may provide advantages in certain situations, such as when site-specific measurement uncertainties are available. Moreover, the geostatistical updating was not bound by predictor variables, but potentially should be. As one example, an underpredicted V_{S30} in a unit mapped as igneous rock shouldn't necessarily suggest that V_{S30} is also underpredicted 1 km away in a unit mapped as alluvium, contrary to what a univariate semivariogram suggests. This possibility could be evaluated in the future. While improvements are inevitably warranted, this study demonstrated the utility of ML for inferring V_{S30} from geospatial information. Ultimately, more data and future research will confirm or update the findings presented herein and succinctly summarized below.

Conclusions

While not a panacea for describing seismic site conditions and response, V_{S30} is an important input parameter for many earthquake applications. This paper developed U.S. national V_{S30} models using ML and geospatial information. Using these models, predictions were mapped at national scale and updated in the vicinity of field measurements. Of the resulting maps, Model 1alt and Model 2alt, which each defer to existing models on steeper slopes, are recommended. Of these, Model 1alt performed slightly better, but requires geologic information that may be unavailable or otherwise problematic. Based on the presented tests, these maps warrant adoption and further evaluation alongside existing solutions. More broadly, the approach employed

herein can be applied to other subsurface data and objectives (e.g., predicting liquefaction, as demonstrated by Geyin et al., 2022).

Data and Resources

All data analyzed in this study is publicly available, as described and referenced in the text. The resulting V_{S30} maps are downloadable from Geyin and Maurer (2022) (https://doi.org/10.17603/ds2-80d8-9m83). Supplemental Material for this article includes additional figures and tables, as described in the main text.

Declaration of Competing Interests

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

Acknowledgments

This study is based on work supported by the National Science Foundation (NSF), US Geological Survey (USGS), and Washington State Dept. of Transportation (WSDOT) under Grant Nos. CMMI-1751216, G18AP-00006, and T1461-74, respectively. However, any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and may not reflect the views of NSF, USGS, or WSDOT.

References

Ahdi, SK, Stewart, JP, Ancheta TD, Kwak DY, and Mitra D. (2017). Development of VS Profile Database and Proxy-Based Models for V_{S30} Prediction in the Pacific Northwest Region of North America. *BSSA* **107**, no. 4, 1781-1801.

Allen, T.I. and Wald, D.J. (2009). On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}). *BSSA* **99**, no. 2A, 935-943.

Amatulli, G., Domisch, S., Tuanmu, M.N., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W. (2018). A suite of global, cross-scale topographic variables for environmental and biodiversity modeling. *Scientific data* **5**, no. 1, 1-15.

Boore, D.M. (2016). Determining Generic Velocity and Density Models for Crustal Amplification Calculations, with an Update of the Boore and Joyner (1997) Generic Site Amplification for $V_S(Z)$ = 760 m/s. *BSSA* **106** no. 1, 316–320.

Boore, D.M. and Joyner, W.B. (1997). Site amplifications for generic rock sites. *BSSA* **87**, no. 2, 327-341.

Boore, D.M., Thompson, E.M., Cadet, H., (2011). Regional Correlations of V_{S30} and Velocities Averaged Over Depths Less Than and Greater Than 30 Meters. *Bulletin of the Seismological Society of America* **101**, no. 6, 3046-3059.

Box, G.E.P and Cox, D.R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*. **26**, no. 2, 211–252.

Breiman, L. (1996). Bagging predictors. *Machine Learning* 24, no. 2, 123–140.

Castellaro, S., Mulargia, F., and Rossi, P.L. (2008). V_{S30}: Proxy for seismic amplification? *SRL* **79**, no. 4, 540-543.

Durante, M.G., and Rathje, E.M. (2021). An exploration of the use of machine learning to predict lateral spreading. *Earthquake Spectra*, **37**, no. 4, 2288-2314.

Duvenaud, D. (2014). Automatic model construction with Gaussian processes, Ph.D. thesis, University of Cambridge.

Dzeroski, S., and Zenko, B. (2004). Is combining classifiers with stacking better than selecting the best one? *Machine learning* **54**, no. 3, 255-273.

Elith, J., Leathwick, J.R., and Hastie, T. (2008). A working guide to boosted regression trees. *Journal of animal ecology* **77**, no. 4, 802-813.

Fan, Y., Li, H., and Miguez-Macho, G. (2020). RE: updated dataset for Global Patterns of Groundwater Table Depth, Science, 339 (6122): 940-943, doi: 10.1126/science.1229881, by the authors. http://thredds-gfnl.usc.es/thredds/catalog/GLOBALWTDFTP/catalog.html

Fick, S.E. and Hijmans, R.J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, no. 12, 4302-4315.

Foster, K.M., Bradley, B.A., McGann, C.R., and Wotherspoon, L.M. (2019). A VS_{30} map for New Zealand based on geologic and terrain proxy variables and field measurements. *Earthquake Spectra* **35**, no. 4, 1865-1897.

Friedman, J.H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics* **29**, no. 5, 1189-1232.

Geyin, M., Maurer, B.W., and Christofferson, K. (2022). An AI driven, mechanistically-grounded geospatial liquefaction model for rapid response and scenario planning. *SDEE* **159**, 107348.

Geyin, M. and Maurer, B.W. (2022). "U.S. national Vs30 maps informed by remote sensing and machine learning." DesignSafe-CI. https://doi.org/10.17603/ds2-80d8-9m83.

Heath, D.C., Wald, D.J., Worden, C.B., Thompson, E.M., and Smoczyk, G.M. (2020a). A global hybrid V_{S30} map with a topographic slope–based default and regional map insets. *Earthquake Spectra* **36**, no. 3, 1570-1584.

Heath, D.C., Wald, D.J., Worden, C.B., Thompson, E.M., and Smoczyk, G.M., (2020b). A Global Hybrid V_{S30} Map with a Topographic-Slope-Based Default and Regional Map Insets: U.S. Geological Survey data release, https://doi.org/10.5066/P96HFVXM.

Hengl, T., Heuvelink, G.B., & Rossiter, D.G. (2007). About regression-kriging: From equations to case studies. *Computers and Geosciences*, **33**, no. 10, 1301-1315.

Holzer, T.L., Padovani, A.C., Bennett, M.J., Noce, T.E., and Tinsley, J.C. (2005). Mapping NEHRP V_{S30} site classes. *Earthquake Spectra*, **21**, no. 2, 353-370.

Horton, J.D., San Juan, C.A., and Stoeser, D.B. (2017). The State Geologic Map Compilation (SGMC) geodatabase of the conterminous United States (ver. 1.1, August 2017): U.S. Geological Survey data release, https://doi.org/10.5066/F7WH2N65.

Iwahashi, J., Kamiya, I., and Matsuoka, M. (2010). Regression analysis of Vs30 using topographic attributes from a 50-m DEM. *Geomorphology* **117**, nos. 1-2, 202-205.

Kayen, R.E., Carkin, B.A., Corbett, S.C., Zangwill, A., Estevez, I., and Lai, L. (2011). Shear wave velocity and site amplification factors for 25 strong-motion instrument stations affected by the M5.8 Mineral, Virginia, earthquake of August 23, 2011. *USGS Report* No. 2015-1099

Kuhn, M. and Johnson, K. (2013). Measuring predictor importance. In *Applied predictive modeling*, 463-485, Springer, New York.

Kwak, D.Y., Ahdi, S.K., Wang, P., Zimmaro, P., Brandenberg, S.J., Stewart, J.P. (2021). Web portal for shear wave velocity and HVSR databases in support of site response research and applications. UCLA Geotechnical Engineering Group, DOI:10.21222/C27H0V.

Lehner, B. and Grill, G. (2013). Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrological Processes* **27**, no. 15, 2171–2186.

Li, M. and Rathje, E. (2020). Comprehensive Map of Vs30 for Texas. DesignSafe-CI. Available at: https://www.designsafe-ci.org/data/browser/public/designsafe.storage.published/PRJ-2651).

Lilliefors, H.W. (1967). On the Kolmogorov-Smirnov test for normality with mean and variance unknown. *Journal of the American Statistical Association* **62**, no. 318, 399–402.

McPhillips, D.F., Herrick, J.A., Ahdi, S., Yong, A.K., and Haefner, S. (2020). Updated Compilation of V_{S30} Data for the United States. *USGS data release*, https://doi.org/10.5066/P9H5QEAC.

NASA (2020). Distance to nearest coastline. *NASA Ocean Biology Processing Group (OBPG).* < *https://oceancolor.gsfc.nasa.gov/docs/distfromcoast/> last accessed July 2021.*

Oliver, M.A., and Webster, R. (2015). *Basic steps in geostatistics: the variogram and kriging*, 15-42. Springer International Publishing.

Parker, G.A., Harmon, J.A., Stewart, J.P., Hashash, Y.M., Kottke, A.R., Rathje, E.M., ... & Campbell, K.W. (2017). Proxy-based V_{S30} estimation in central and eastern North America. *BSSA* **107**, no. (1), 117-131.

Petersen, M.D. et al. The 2018 update of the US National Seismic Hazard Model: Overview of model and implications." *Earthquake Spectra* **36**, 5–41.

444 Piryonesi, S.M. and El-Diraby, T.E. (2021). Using machine learning to examine impact of type of performance
445 indicator on flexible pavement deterioration modeling. *Journal of Infrastructure Systems* **27**, no. 4,
446 04021005.
447 Rasmussen, C.E. (2003). Gaussian processes in machine learning. In *Summer school on machine learning*, 63-
448 71. Springer, Berlin, Heidelberg.
449 Rokach L and Maimon O (2008) *Data mining with decision trees: theory and applications*. World Scientific
450 Pub Co Inc. ISBN 978-9812771711.
451 Salomone, L.A., Hamel, J.F., and Kassawara, R.P. (2012). EPRI (2004, 2006) ground-motion model (GMM)
452 review project: Shear wave velocity measurements at seismic recording stations. Electric Power Research
453 Institute, Report EP-P43952/C19088.
454 Shangguan, W., Hengl, T., de Jesus, J.M., Yuan, H., and Dai, Y. (2017). Mapping the global depth to bedrock
455 for land surface modeling. *Journal of Advances in Modeling Earth Systems* **9**, no. 1, 65-88.
456 Stewart, J.P., Parker, G.A., Harmon, J.A., Atkinson, G.M., Boore, D.M., Darragh, R.B., Silva, W.J. and
457 Hashash, Y.M. (2017). Expert panel recommendations for ergodic site amplification in central and eastern
458 North America. *PEER Report* 2017-04. Berkeley, CA.
459 Thompson, E. M., Wald, D. J., & Worden, C. B. (2014). A V_{S30} map for California with geologic and
460 topographic constraints. *Bulletin of the Seismological Society of America* **104**, no. 5, 2313-2321.
461 Vapnik, V. (1995). *The Nature of Statistical Learning Theory.* Springer, New York.
462 Verdin, K.L., (2017). Hydrologic Derivatives for Modeling and Applications (HDMA) database: U.S.
463 Geological Survey data release, https://doi.org/10.5066/F7S180ZP.
464 Wald, D.J. and Allen, T.I. (2007). Topographic slope as a proxy for seismic site conditions and amplification.
465 *BSSA* **97**, no. 5, 1379-1395.
466 Wills, C. and Clahan, K. (2006). Developing a map of geologically defined site-condition categories for CA.
467 *BSSA* **96**, 1483–1501.
468 Worden, C., Wald, D., Allen, T., Lin, K., Garcia, D. and Cua, G. (2010). A revised ground-motion and intensity
469 interpolation scheme for ShakeMap. *BSSA* **100**, no. 6, 3083–3096.

Worden, C.B., Thompson, E.M., Baker, J.W., Bradley, B.A., Luco, N., and Wald, D.J., (2018). Spatial and spectral interpolation of ground-motion intensity measure observations. *BSSA* 108, no. 2, 866–875.

Yong, A., Hough, S. E., Iwahashi, J., & Braverman, A. (2012). A terrain-based site-conditions map of California with implications for the contiguous United States. *Bulletin of the Seismological Society of America* **102**, no. 1, 114-128.

Yu, Q. (2021). Soil velocity models informed by remote sensing and artificial intelligence. MS Thesis, Department of Civil and Environmental Engineering, University of Washington, Seattle.

Zhu, J., Daley, D., Baise, L. G., Thompson, E. M., Wald, D. J., & Knudsen, K. L. (2015). A geospatial liquefaction model for rapid response and loss estimation. *Earthquake Spectra*, **31**, no. 3, 1813-1837.

Zhu, J., Baise, L.G., and Thompson, E.M. (2017). An updated geospatial liquefaction model for global application. *Bulletin of the Seismological Society of America **107***, no. 3, 1365-1385.

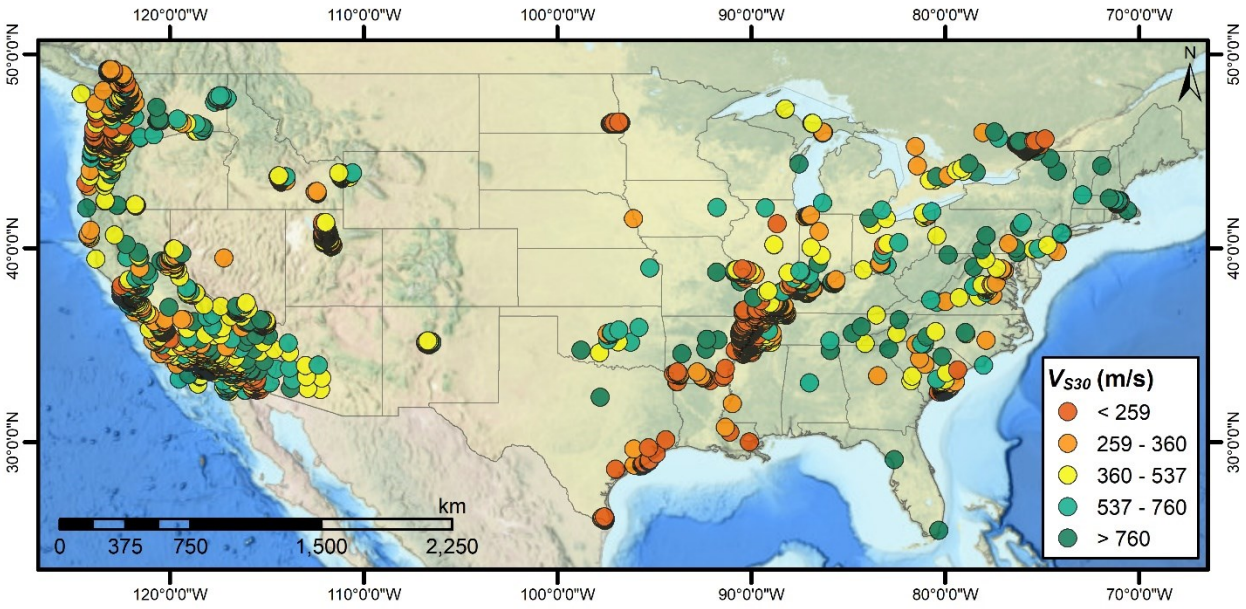

Figure 1. Spatial distribution of V_{S30} measurements in the contiguous U.S.

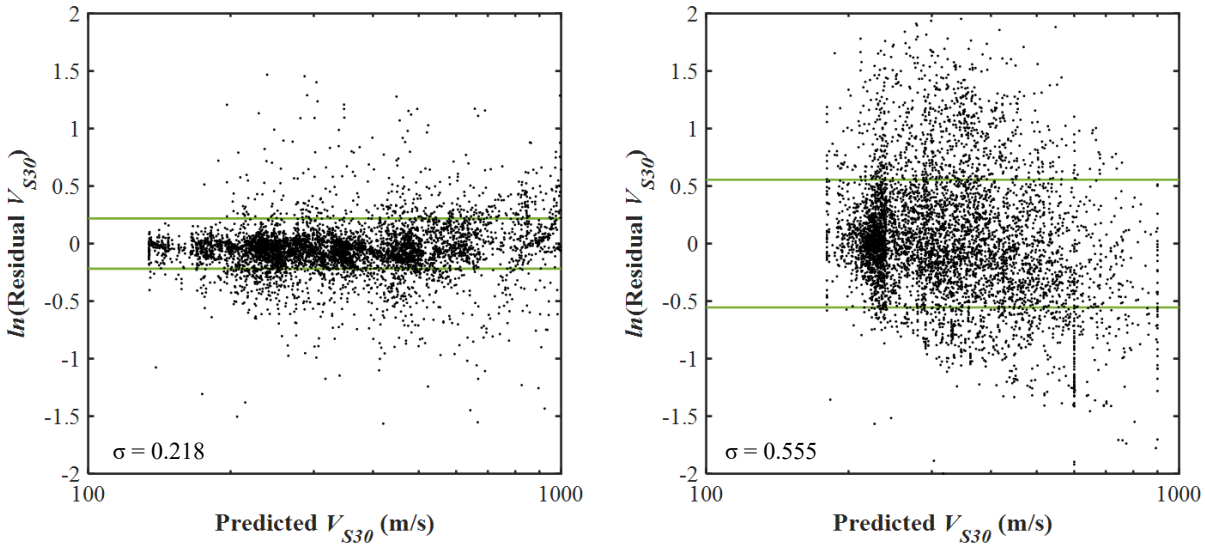

Figure 2. Prediction residuals [ln (observed/predicted)] computed for (a): Model 1; and (b) AW09. The green bands depict the standard deviations of the residuals for each model.

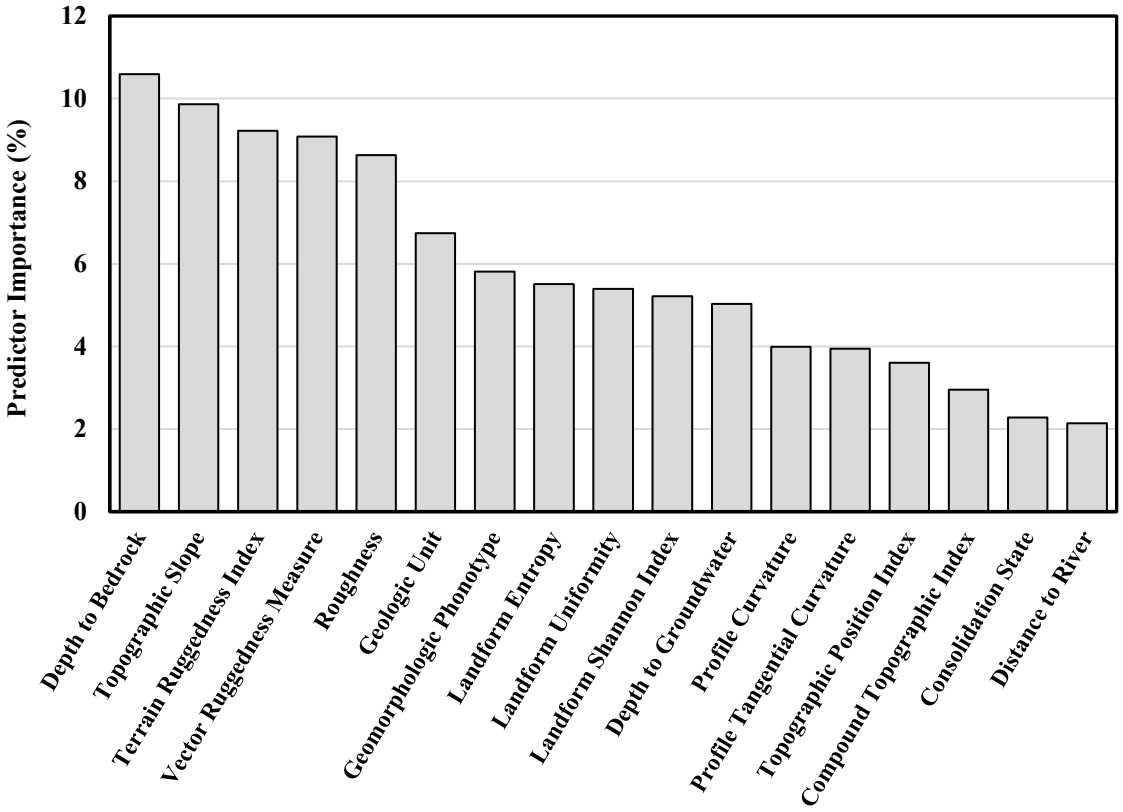

Figure 3. Relative predictor importance ranking for Model 1.

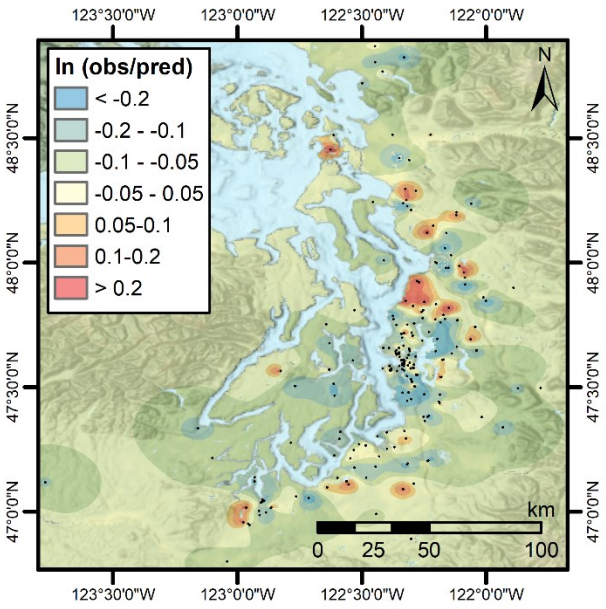

Figure 4. Krigged residuals in the Puget Sound region of Washington State.

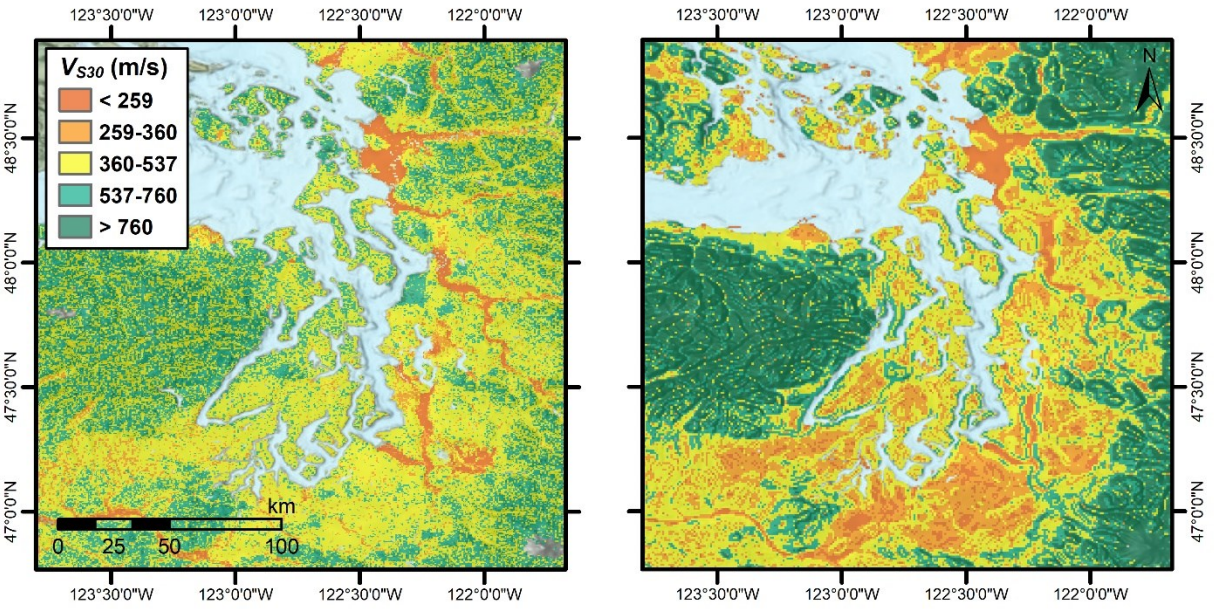

Figure 5. V_{S30} predicted by: (a) Model 1 with residual kriging; and (b) AW09 in the Puget Sound.

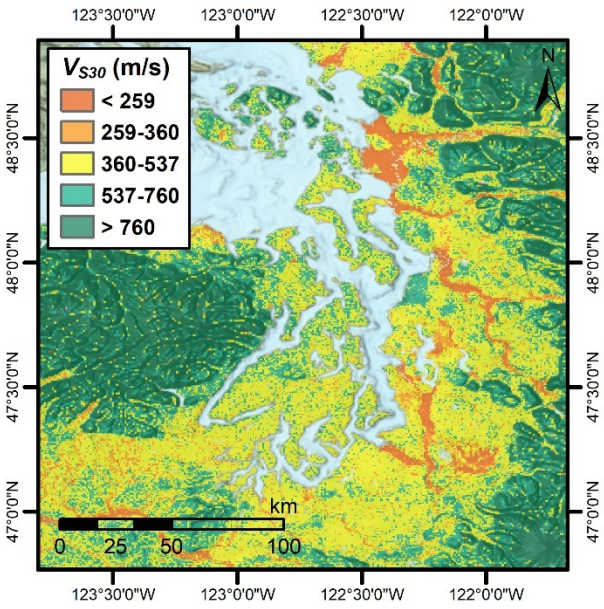

Figure 6. V_{S30} predicted by Model 1alt with residual kriging in the Puget Sound.

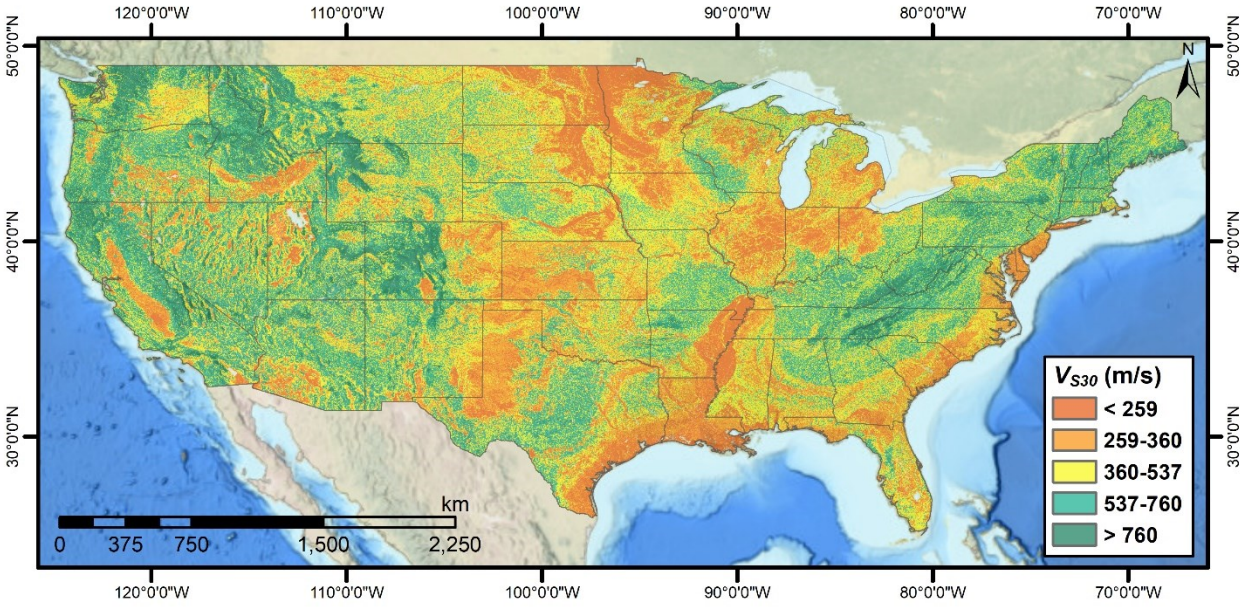

Figure 7. V_{S30} predicted by Model 1alt with residual kriging in the contiguous United States.

Table 1. Range, spatial resolution, and sources of predictor variables in the dataset.

Variable (Units)	Source	Range in Dataset	Spatial Resolution
Depth to bedrock (cm)	Shangguan et al. (2017)	0 to 43,437	250 m
Depth to groundwater (m)	Fan & Miguez-Macho (2020)	0 to 216	~1000 m (30 arc-sec)
Geologic unit	Horton et al. (2017)	Categorical	25 m to 500 m (varies)
Consolidation state	Horton et al. (2017)	0 or 1	25 m to 500 m (varies)
Distance to river (m)	Lehner and Grill (2013)	0 to 8.4×10^4	~90 m (3 arc-sec)
Compound topographic index	Verdin et al. (2017)	484 to 2858	~90 m (3 arc-sec)
Geomorphologic phonotype	Amatulli et al. (2018)	Categorical	~1000 m (30 arc-sec)
Topographic slope (%)		0 to 26.7	~1000 m (30 arc-sec)
Topographic position index		-37.38 to 22.94	~1000 m (30 arc-sec)
Profile curvature		-0.0012 to 0.0013	~1000 m (30 arc-sec)
Tangential curvature		-9.0577×10^{-4} to 9.35069×10^{-4}	~1000 m (30 arc-sec)
Roughness		0 to 284	~1000 m (30 arc-sec)
Terrain ruggedness index		0 to 90.88	~1000 m (30 arc-sec)
Vector ruggedness measure		0 to 0.0457	~1000 m (30 arc-sec)
Landform entropy		0 to 2.9572	~1000 m (30 arc-sec)
Landform uniformity		0.0536 to 1	~1000 m (30 arc-sec)
Landform Shannon index		0 to 2.0467	~1000 m (30 arc-sec)

Table 2. Mean absolute errors (MAE), binned on V_{S30} and topographic slope, for the unbiased test set.

Bin Variable	Bin Range	Model 1 MAE (m/s)	AW09 MAE (m/s)	Improvement (%)
V_{S30} (m/s)	0-180	55.64	164.76	66.23
	180-259	55.84	57.38	2.70
	259-360	77.65	84.61	8.23
	360-537	98.43	126.09	21.93
	537-760	148.32	239.39	38.04
	760-1150	296.73	520.70	43.01
	1150-2000	531.23	1055.53	49.67
	>2000	1484.25	1700.35	12.71
Slope (deg)	0.00-0.13	39.26	46.03	14.70
	0.13-0.21	44.08	76.05	42.04
	0.21-0.30	88.51	140.10	36.82
	0.30-0.40	101.93	200.77	49.23
	0.40-0.55	101.94	159.17	35.96
	0.55-0.78	119.46	198.80	39.91
	0.78-1.24	155.38	280.24	44.56
	>1.24	168.70	200.95	16.05

Figure Captions

Figure 1. Spatial distribution of V_{S30} measurements in the contiguous U.S.

Figure 2. Prediction residuals [ln (observed/predicted)] computed for (a): Model 1; and (b) AW09. The green bands depict the standard deviations of the residuals for each model.

Figure 3. Relative predictor importance ranking for Model 1.

Figure 4. Krigged residuals in the Puget Sound region of Washington State.

Figure 5. V_{S30} predicted by: (a) Model 1 with residual kriging; and (b) AW09 in the Puget Sound.

Figure 6. V_{S30} predicted by Model 1alt with residual kriging in the Puget Sound.

Figure 7. V_{S30} predicted by Model 1alt with residual kriging in the contiguous United States.

Table Captions

Table 1. Range, spatial resolution, and sources of predictor variables in the dataset.

Table 2. Mean absolute errors (MAE), binned on V_{S30} and topographic slope, for the unbiased test set.

Mailing Addresses

Brett W. Maurer
132 More Hall
University of Washington
Seattle, WA 98195

Mertcan Geyin
Norwegian Geotechnical Institute
10615 Shadow Wood Dr Suite 100
Houston, TX 77043