Bulletin of the Seismological Society of America, Vol. 102, No. 1, pp. 114–128, February 2012, doi: 10.1785/0120100262

A Terrain-Based Site-Conditions Map of California with Implications for the Contiguous United States

by Alan Yong, Susan E. Hough, Junko Iwahashi, and Amy Braverman

Abstract We present an approach based on geomorphometry to predict material properties and characterize site conditions using the V_{S30} parameter (time-averaged shear-wave velocity to a depth of 30 m). Our framework consists of an automated terrain classification scheme based on taxonomic criteria (slope gradient, local convexity, and surface texture) that systematically identifies 16 terrain types from 1-km spatial resolution (30 arcsec) Shuttle Radar Topography Mission digital elevation models (SRTM DEMs). Using 853 V_{S30} values from California, we apply a simulation-based statistical method to determine the mean V_{S30} for each terrain type in California. We then compare the V_{S30} values with models based on individual proxies, such as mapped surface geology and topographic slope, and show that our systematic terrain-based approach consistently performs better than semiempirical estimates based on individual proxies. To further evaluate our model, we apply our California-based estimates to terrains of the contiguous United States. Comparisons of our estimates with 325 V_{S30} measurements outside of California, as well as estimates based on the topographic slope model, indicate our method to be statistically robust and more accurate. Our approach thus provides an objective and robust method for extending estimates of V_{S30} for regions where *in situ* measurements are sparse or not readily available.

Introduction

Spatial variability in strong ground motions has been widely observed in many past earthquakes. For example, amplification due to near-surface geology gave rise to dramatic large-scale variations in shaking and damage during the 1985 M_w 8.1 Michoacán earthquake (e.g., Singh and Ordaz, 1993), the 1989 M_w 6.9 Loma Prieta earthquake (e.g., Hough *et al.*, 1990), and the 1994 M_w 6.7 Northridge earthquake (e.g., Spudich *et al.*, 1996). Other noteworthy observations include the 1999 M_s 5.9 Athens earthquake, which caused unexpectedly heavy damage in the town of Adàmes that borders a canyon crest (e.g., Assimaki *et al.*, 2005), and the 2009 M_w 6.3 trembler in the Abruzzi region of central Italy, where reports described preferential northwest-southeast shaking intensity patterns recorded locally in the mountainous terrain of the northern Apennines that surrounds the town of L'Aquila (e.g., Ameri *et al.*, 2009; Akinci *et al.*, 2010). Clearly, it has long been consistently recognized that, while source and path effects are important factors for controlling shaking, site effects also play a key role in determining ground-motion intensities during a strong earthquake (e.g., Aki, 1988; Boore *et al.*, 1997; Field, 2000; Lee *et al.*, 2009). It has also been obvious that near-surface geology is not the only site condition that strongly controls shaking (e.g., Gao *et al.*, 1996; Magistrale *et al.*, 2000; Olsen *et al.*, 2006; Hough *et al.*, 2010).

Local site conditions are typically decoupled then analyzed on the basis of the geometric (topographic) or the material (soil) properties. In the case of geometry-related effects, the term topography has often been used either to describe the surface expression of terrain types of varying degrees of relief; or, less frequently, to characterize the irregular interface(s) between strata of differing material properties in the subsurface. To maintain consistency and avoid confusion in this paper, we restrict our references on topography to the morphometric nature of relief at the Earth's surface. As for the material properties related to site conditions, key factors include layer thickness and impedance (the product of the density of the material and the velocity of the propagating wave). Because shear-wave velocity (V_S) correlates with soil rigidity and is characterized by higher variability than density, V_S has generally been accepted as an appropriate measure of soil conditions (e.g., Borcherdt, 1970; Fumal, 1978; Aki, 1988; Borcherdt and Glassmoyer, 1992; Boore, 2006). Although variations in material properties at depths of tens to hundreds of meters below the Earth's surface are known to significantly influence ground motions, deeper variations are also important (e.g., Anderson *et al.*, 1996; Frankel *et al.*, 2002; Boore, 2004; Holzer *et al.*, 2005). Nevertheless, a number of studies (e.g., Borcherdt *et al.*, 1991; Boore *et al.*, 1993; Borcherdt, 1994; Boore *et al.*,

1994; Boore and Fumal, 1997; Wills and Silva, 1998; Dobry *et al.*, 2000) have found good correlation between observed site amplification and V_{S30}, or the time-averaged shear-wave velocity to a depth of 30 m below the surface. As a result, the V_{S30} parameter has been widely adopted as the key parameter for site characterization. Adoption of this parameter has been driven by practical as well as physical considerations. In particular, 30 m is the average depth reachable by a drill rig during one typical day's worth of work (C. Wills, personal commun., 2006). However, the simple V_{S30} parameter may not account for the complexities associated with the physical properties known to exist in the near-surface.

The use of proxies, that is, geologic properties and/or units, topography, terrain types, and so forth, to infer V_S or V_{S30} has been previously applied when direct measurements are prohibitive, sparse, or not readily available. Because these conditions are common everywhere, it is likely that other proxy-based V_{S30} maps were developed elsewhere besides California and Japan. With no intention to purposefully exclude other regional investigations, we herein describe selected studies that directly relate to the development of our V_{S30} estimation model.

A number of proxy-based maps have been developed for all or parts of California. Joyner *et al.* (1981), for example, proposed using velocity to a depth equal to one-quarter of the wavelength corresponding to the period of interest to account for site conditions when developing ground-motion prediction equations. Following this idea, Tinsley and Fumal (1985) mapped Quaternary sedimentary units in the Los Angeles basin on the basis of the variation of V_S with age, grain size, and depth, and then, with consideration given to both the characteristic V_S and the thickness of the unit, calculated mean V_S to one-quarter wavelength for a 1-s shear-wave as described in Fumal and Tinsley (1985). Based on the mean V_S assignments from Fumal and Tinsley (1985), and, V_S profiles from the U.S. Geological Survey (Fumal *et al.*, 1981; 1982; 1984; Gibbs *et al.*, 1980) and VIC, Inc. (VIC, 1993), Park and Elrick (1998) developed a Quaternary Tertiary Mesozoic (QTM)-based V_{S30} map grouped into nine units of similar V_{S30} values.

Wills *et al.* (2000) presented the first cut of a geologic unit-based V_{S30} map for California that has found widespread usage in recent years. Further subdividing the original eight site classes used in the 2000 study, Wills and Clahan (2006) refined the urban areas (C. Wills, personal commun., 2006) of the earlier map by introducing 19 site classes based on a number of factors in addition to surficial geology, such as grain size, thickness of units, deep alluvium assumed for major basins, shallow depth for small and narrow basins, and degree of coarseness of sediments as a function of distance from mountain fronts. Both the Wills *et al.* (2000) and Wills and Clahan (2006) maps were derived from existing geologic maps of California. To develop the maps, geologic units were first correlated with V_{S30} values on the basis of shear-wave velocity measurements mostly located in the major metropolitan areas (Wills and Silva, 1998). To map V_{S30} values throughout the state, Wills *et al.* (2000) considered several geologic map series. After evaluating many readily available map series for appropriate scale and resolution, completeness of coverage, uniformity in descriptive nomenclature, continuity of data, and other relevant factors, Wills *et al.* (2000) decided on a compromise between readily available digital maps with insufficient details and an unmanageable number of detailed maps in paper form. Hence, despite the imperfect nature of the series, the authors used two sets of 1:250,000 scale maps: the Geologic Atlas of California and the Regional Geologic Map series. For example, in addition to lacking uniformity in unit nomenclature, the regional geologic series (1981 to 1992) described more recent mapping details of Quaternary geologic features than the older (1958 to 1974) atlas series (Wills *et al.*, 2000). As acknowledged by Wills *et al.* (2000), the ideal approach to solving these fundamental issues was to first systematically (digitally) group all related geologic units, assign V_{S30} on the basis of their shear-wave velocity properties as defined by Wills and Silva (1998), and then combine all units with the same V_{S30} classifications. Due to practical limitations, Wills *et al.* (2000) first grouped the geologic units by V_{S30} classification, and then prepared a tracing of the geologic map along the contacts that separated the units of differently inferred V_{S30} classifications.

In Japan, Matsuoka and Midorikawa (1995) developed a geographic information systems (GIS)-based amplification map using peak ground velocity (PGV) and the correlation between PGV and V_{S30} from 459 sites to demonstrate that terrain-based classification is appropriate for seismic hazards mapping in large metropolitan areas. Later, using map information from the Digital National Land Information (DNLI) database, Yamazaki *et al.* (2000) combined the use of both geomorphological and geological (subsurface) classifications and found that they yield the best estimates of site amplification ratios for 77 Japan Meteorological Agency strong-motion stations. On the basis of the relationship between amplification ratio and the GIS classes, Yamazaki *et al.* (2000) suggested that the DNLI-derived data are appropriate for the estimation of strong ground motion over large areas in Japan. In another study, Matsuoka *et al.* (2006) developed a geomorphology-based V_{S30} proxy model derived from the Japan Engineering Geomorphology Classification Map (JEGM). The JEGM is a collection of 1:50,000 scale terrain classification maps produced by a consortium consisting of local governments (Wakamatsu *et al.*, 2004). Analyzing data from approximately 2000 sites, Matsuoka *et al.* (2006) concluded that there is good correlation between V_{S30} and the set of metrics that include the manually classified terrain units, elevations, slope angles, and distances from mountains or hills. More recently, Matsuoka joined Iwahashi *et al.* (2010) to examine results from multiple regression analyses of V_{S30} measurements at 1646 locations with topographic attributes (slope gradient, surface texture, and the logarithm of elevation) based on 50-m digital elevation models (DEMs) and found that estimations of V_{S30} from a DEM are useful for earthquake vulnerability assessment.

While V_{S30} maps have traditionally been derived directly from geological maps, in recent years several studies have exploited newly available remote sensing data to develop site classification maps (e.g., Romero and Rix, 2001; Wald and Allen, 2007; Yong *et al.*, 2008a; 2008b). As one of the earliest examples of studies utilizing satellite imagery to account for site conditions, Romero and Rix (2001) used Landsat 7 Enhanced Thematic Mapper Plus data to interpret the character of surficial sediments for estimating the potential of ground shaking in the Mississippi embayment. To infer soil rigidity, they manually applied digital image processing methods to distinguish the geological ages (Holocene or Pleistocene) of the sediments. Because the local brightness and textural variations in the spectral reflectance of sediments in satellite imagery is closely related to the relative ages of the sedimentary deposits, and because soil rigidity is reasonably well associated with the age of the rock unit, the investigators exploited this relationship to infer the site factor for estimating future ground motions in the absence of more precise geologic map information.

Through the use of Shuttle Radar Topography Mission (SRTM) (Farr *et al.*, 2007) DEMs, Wald and Allen (2007) presented a V_{S30} map for California-based only on topographic slope. In their approach, areas were first partitioned as active tectonic (California, Taiwan, Italy, and Utah) and stable continental (Memphis and Australia) regions. Then, for each region, they correlated available V_{S30} values with topographic slope (m/m). Because they did not expect to find a direct, physical relationship between slope and V_{S30}, Wald and Allen (2007) instead characterized the relationship in terms of discrete steps in shear-wave velocity values tied to the National Earthquake Hazards Reduction Program (NEHRP) V_{S30} boundaries. For validation of their slope-based V_{S30} proxy model, Wald and Allen (2007) used both statistical and empirical map-based analytic methods to assess their observations and concluded that their results were sufficiently accurate for a first approximation of site conditions and amplification.

Most recently, Yong *et al.* (2008a; 2008b) performed semiautomated analyses of spectra (visible to thermal-infrared electromagnetic frequency range) and 30-m resolution DEM (stereoscopic correlation method) data produced from the Advanced Spaceborne Thermal Emission and Reflection Radiometer satellite recordings to interpret the remotely sensed geologic and topographic information. Subsequently, these results were assembled into terrain types to estimate V_{S30} in a 60 km × 60 km area surrounding Islamabad, Pakistan. This study demonstrated that satellite data can help circumvent the problems associated with inconsistencies found in traditional geologic mapping. They also showed that the combination of multiple parameters, such as geology and topography, yields a more robust estimate than uniparametric models. However, results from earlier investigations (Yong *et al.*, 2005, 2006, 2007) indicated that a systematic and objective classification method that is both simple and yet robust for all terrain types is necessary for characterizing larger regions (e.g., Murray and Fonstad, 2007; Philips, 2007). To meet these requirements, we investigate the Iwahashi and Pike (2007) automated approach to develop a global terrain classification map that was also based on SRTM DEMs. We describe the details of this method, which is the basis for our study, in the following section.

In addition to the use of the V_{S30} parameter, the most recent models described previously share another commonality in their dependence on GIS to process and analyze the spatial relations between attributes of the respective parametric (geology or topography) frameworks and the distribution of V_S or V_{S30} measurement sites. For example, Petersen *et al.* (1997), Park and Elrick (1998), Wills *et al.* (2000), Wills and Clahan (2006), and Wald and Allen (2007) all used various forms of readily available digitized base-maps as their underlying frameworks and GIS to analyze and develop their respective V_{S30} proxy models.

Indeed, recent advances in GIS-related technologies have played a key role in the proliferation of map-based proxy models for estimating V_{S30}. Because these technologies are optimal for supporting fully systematic approaches, we exploit the power of GIS, and at the same time apply rigorous statistical methods to develop our terrain-based proxy model. Also, because geomorphology is the manifestation of near-surface geology and, because terrain types are primarily controlled by these same variations in the properties that make up the near-surface (e.g., Peterson, 1981; Harden, 1990; Ellis *et al.*, 1999; Saco *et al.*, 2006), we expect in principle, that our reliance of morphometrics (Oliver and Webster, 1986; Franklin, 1987; Dehn *et al.*, 2001; Minár and Evans, 2008; Pike *et al.*, 2009) in terrain models will meet the requirements for a consistent, objective, and quantitative framework that systematically encapsulates the parameters related to near-surface geological properties. Because our focus on site characterization is primarily related to local site conditions that are strongly controlled by soil and topographic effects, we first explore the effectiveness of using a terrain-based model as the underlying framework for a V_{S30} proxy-based approach.

Data

Terrain-Based Framework

Iwahashi and Pike (2007) (hereinafter IP07) introduced an automated topography classification scheme to develop a globally consistent terrain classification map. The IP07 approach employed an algorithm that divides continuous topography represented by elevations into a maximum possibility of 16 terrain classes through the use of a geometric signature consisting of three taxonomic criteria: slope gradient, local convexity, and surface texture. For the main part of this study, we use the same automated scheme and only adopt the California region out of the global dataset. In a later section, we discuss some results from preliminary tests

relating to the use of an IP07 terrain map covering the contiguous United States.

The first component of the three-part geometric signature is slope gradient, which is referred to by IP07 as steepness. Values of slope gradient are calculated to the nearest degree by the Earth Sciences Research Institute (ESRI) ArcInfo slope subroutine that applies gradient estimators while operating in a 3×3 array ($N \times N$ spatial matrix consisting of DEM grid cells) of neighboring elevations (Horn, 1981). Because slope gradient is a key factor in controlling geomorphic processes, this fundamental property of topography is the most influential parameter in virtually all automated terrain classification systems.

The second variable, local convexity, is surface curvature. This parameter was found to be necessary for discriminating among various low-relief features, such as alluvial fans, terraces, flood plains, and so forth. Here, a Laplacian operator is used to calculate the values for convex-upward areas (positive), as well as concave (negative) and planar (zero) regions. To quantify local convexity, the percentage of convex-upward cells within an arbitrary 10-cell radius is used. Broad valleys and foothills are characterized by low surface concavity, while lowland terrains such as terraces and alluvial fans are characterized by higher convexity values.

The third variable, surface texture, initially referred to as the frequency of ridges and valleys, or roughness (Iwahashi *et al.*, 2001), and later as pits and peaks (IP07), is used to quantify topographic spacing in the DEM. Here, the approach is to use a spatial convolution filtering technique (median filtering) typically applied in image edge-enhancement operations (Jensen, 2007) to aid in identifying the differences in values between the original DEM and the derived (median filtered) DEM. This technique captures the fine-versus-coarse fabric, that is, the grain of the pits and peaks. Surface texture is then determined by the number of surface pits and peaks within an optimal radius of 10 cells, a radius previously determined to be capable of adequately capturing all topographic crests and troughs regardless of amplitude (Iwahashi and Kamiya, 1995).

The design of the topographic classification scheme is based on an unsupervised nested-means approach (Scripter, 1970) where the final output of terrain classes reflects the statistical properties of the input geometric variables rather than preset criteria. Unlike supervised algorithms, for which predetermined criteria for each output class do not adapt equally well to all locations and DEM resolutions, the unsupervised approach treats topography as a continuous random surface, independent of any spatial or morphological orderliness imposed by fluvial activity and other geomorphic processes. Because the resultant terrain classes are not determined *a priori* by training examples of target physiographic types, IP07 acknowledged that these classes are only approximately equivalent to traditional terrain types such as mountains, high hills, plains with hills, or tablelands in the manner of Hammond's (1964) taxonomy. Nevertheless, on the basis of subsequent map overlays and statistical analyses, IP07 showed that their U.S. terrain classification map closely resembles the classic physiographic divisions of Fenneman and Johnson (1946) and for more recent work, the geomorphic systems in Graf (1987; p. 4).

Despite the unconstrained nature of the IP07 algorithm, output categories of 8, 12, or 16 classes of undefined terrain types must first be specified. For our California terrain model, based on the wide range in terrain types present, we choose the maximum possible output category of 16 classes for our V_{S30} proxy model. Using the nested-means approach, the algorithm begins by partitioning the DEM into classes of terrain forms on the basis of each of the three geometric variables. In this process, a set of three filters progressively divides the classes into half/half or twofold partitions of the DEM (Fig. 1).

In the initial step of the first filter, slope gradient is used to divide cells into two partitions on the basis of whether the value is larger or smaller than the mean steepness in the DEM area. Next, the partitions are subdivided in the same manner, by using the mean of the convexity in each of the two partitions, respectively. Then, each of the four partitions is further partitioned into eight divisions on the basis of the means of the texture in the four divided areas. At this intermediate stage, an eight-class partition has been created and is presented as the first of three possible output categories. In the second filter, classes 1–4 are retained while classes 5–8 are binarized into twofold partitions in the same manner, thus resulting in the second possible output category of 12 classes. The third filter, operating on the eight classes output from the second filter and using the same binarization and twofold thresholding methods for retention, produces the final and maximum output category of 16 undefined terrain classes. At the end of the automated DEM partitioning process, the map assignments of terrain nomenclature (mountains, terraces, alluvial fans or basins, etc.) are then empirically established on the basis of regional expert interpretations. In our terrain framework, for example, the physiographic taxonomy established for the United States (Hammond, 1954; 1964) is used to relate the equivalent classes to the terrain category (Fig. 2).

Regardless of the complexity and/or innovation associated with automated and systematic approaches, the accuracy of the derived model is inherently limited by the quality of the input data. For the topography of California, IP07 used DEMs of spatial resolution at 30 arcsec (approximately 1-km grid spacing or pixel size) derived from recordings measured by the Shuttle Radar Topography Mission (SRTM) (Farr *et al.*, 2007). Although higher resolutions of SRTM DEMs were also available (e.g., the SRTM recordings were sampled over a grid at spatial resolution of 1 arcsec by 1 arcsec or approximately 30 m $\times$ 30 m), we instead choose to adhere to the proven IP07 technique that produced reasonably accurate classifications (Iwahashi and Pike, 2007). At the same time, the use of the same 30-arcsec topography employed by Wald and Allen (2007) allows us to directly compare the two methods. By comparison, the 1:250,000-scale

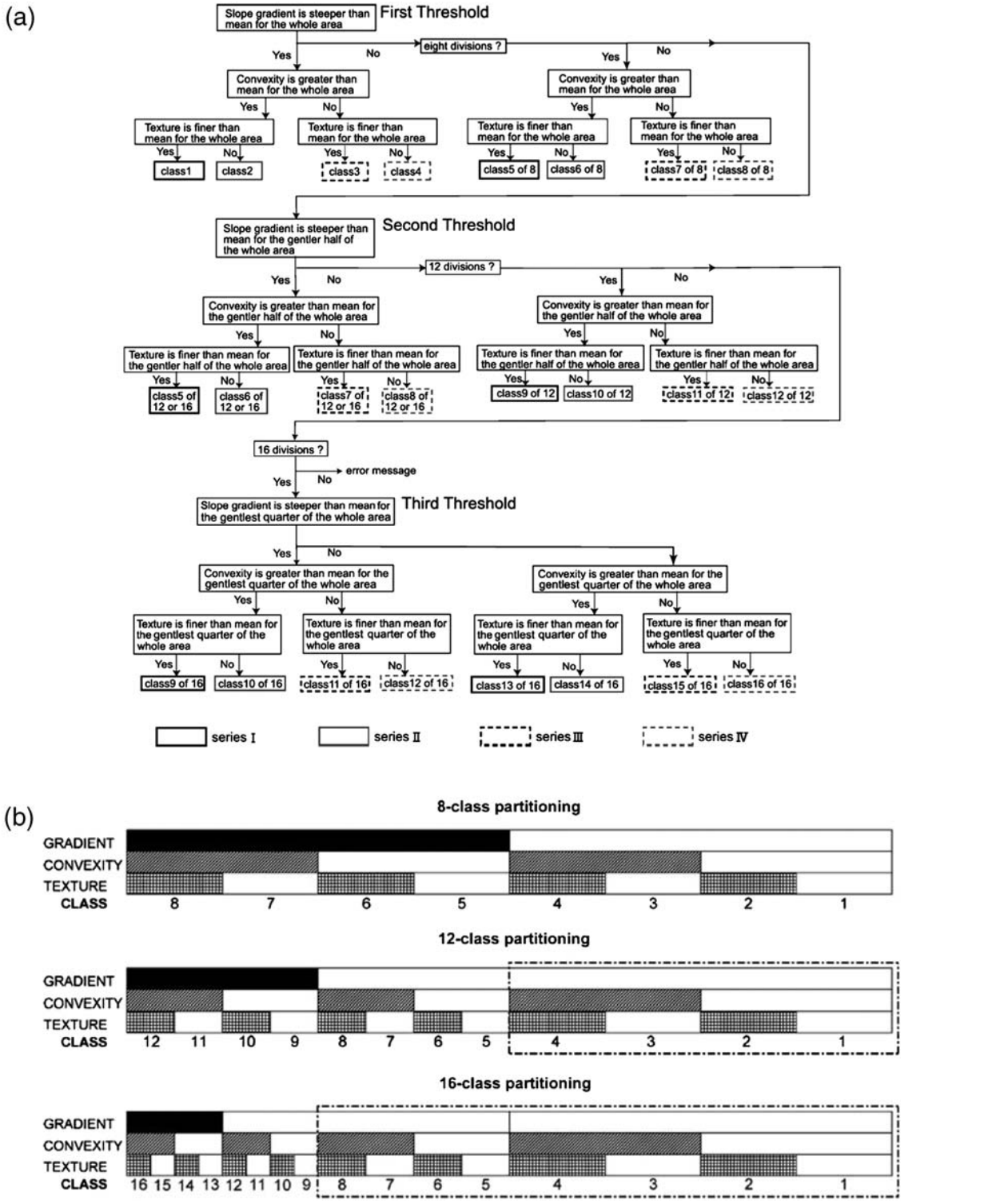

Figure 1. (a) Flow chart (reprinted from Iwahashi and Pike, 2007) for automated nested-means classification of topography; criteria are mean values of slope gradient, local convexity, and surface texture, which change to accommodate increasingly low-relief topography depending on number of classes requested (8, 12, or 16). Grid cells steeper than mean gradient of study area are distributed among classes 1–4 by first threshold calculation in all three options; for an 8-class output, the remaining cells fall into classes 5–8; for 12 classes, the latter cells are instead allotted among classes 5–12 by reduced parameter means (second threshold); for 16 classes, cells not placed in classes 1–8 are similarly parsed at the third threshold into classes 9–16. (b) Nested-means half/half partitioning shown diagrammatically (reprinted from Iwahashi and Pike, 2007) for gray-scaled images of three input variables and 8, 12, and 16 output classes; sequence of operations proceeds from top down. Shaded rectangles represent grid cells below the mean of variable in that row; open rectangles are cells above the mean; rectangles within two dot–dashed boxes are same as those immediately above. Rectangle size is symbolic only and does not represent number of grid cells.

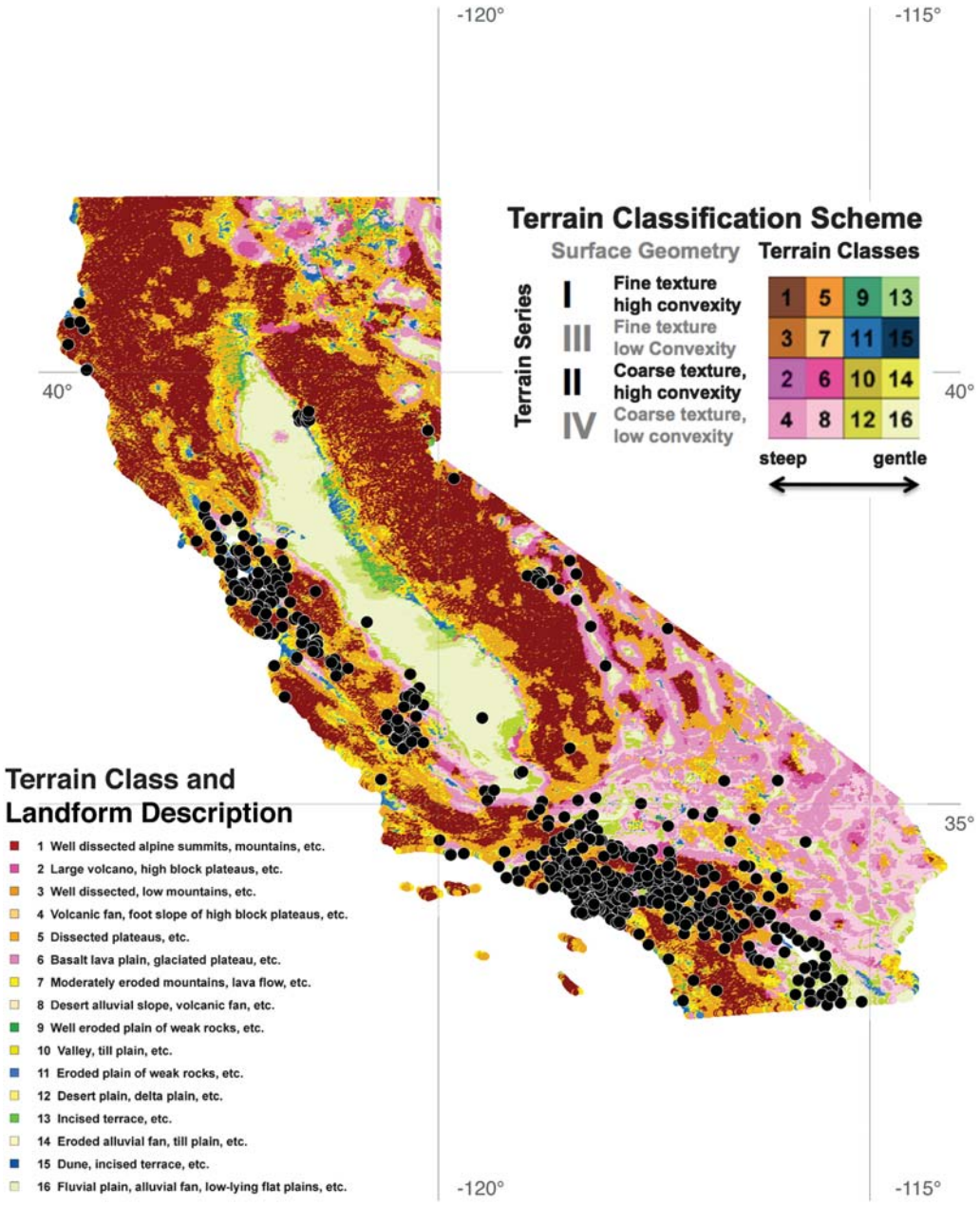

Figure 2. Terrain map of California (Iwahashi and Pike, 2007) with locations (black circles) of 853 V_{S30} values.

maps used by Wills *et al.* (2000) have (in principle) the best resolved details (pixel resolution of 125 m) with a detectable object-resolution of 250 m (Tobler, 1988).

V_{S30} Values

In developing our terrain-based V_{S30} proxy model, we use shear-wave velocity data provided by Walter Silva (last revised July 2006; written commun., 2007) of Pacific Engineering and Analysis. As part of the Next Generation Attenuation (NGA) strong-motion database (Chiou *et al.*, 2008), the V_{S30} data consist of both measured and inferred values, that is, measurements from recorded waveforms and extrapolations from proxy- and/or map-based correlations, respectively. Hence, we will continue to purposefully refer to this mix of V_{S30} data as V_{S30} values or results and not as measurements to avoid any misrepresentation about the nature of the source data. Of the 853 values in the database, 644 (75%) were inferred. Measured V_{S30} data were calculated from 201 (24%) V_S records with (minimum) 30-m depth profiles (Fig. 3). Other data used to infer V_{S30} include measurements are from six (0.7%) V_S profiles of less than 20 m and two (0.2%) borehole records extending to the 9-m depth. Furthermore, this sparse subset of measurement-derived V_{S30} values are incompatible due to the diversity of measurement methods used (W. Silva, personal commun., 2009). This limitation was previously noted by a number of authors (e.g., Wills, 1998; Boore, 2006; Moss, 2008). Biases

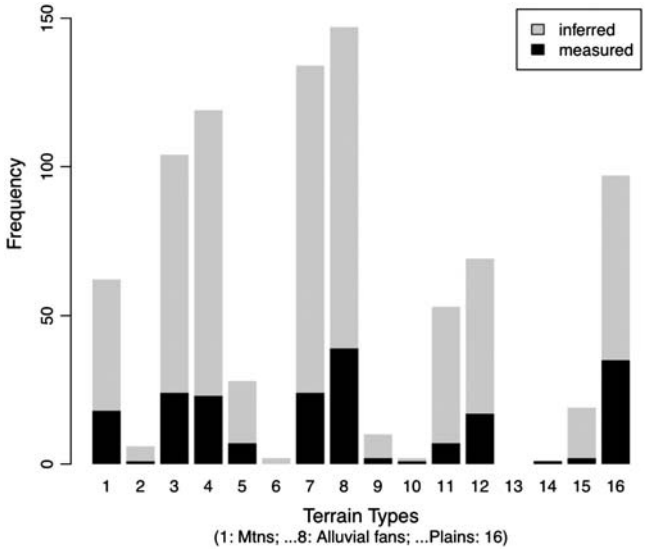

Figure 3. Barplot describing the combined frequency distribution of measured (black) and inferred (gray) V_{S30} values in each terrain-type. Note: No V_{S30} values (measured and/or inferred) are associated with terrain-type 13 described as "incised terraces by Iwahashi and Pike (2007).

between and imprecision associated with different methods control the overall epistemic uncertainties related to the results, and are directly attributable to the fundamental differences in principle that governs the respective geophysical method applied. For example, in a study comparing the variability of shear-wave velocity determinations from existing blind and comparative studies, Moss (2008) examined both the intramethod and intermethod variability and found a clear bias between invasive and noninvasive methods.

For V_{S30} values derived from proxy- and map-based inferences, concerns regarding data quality are generally related to the same issues as previously described for measured values. Despite these well-known issues, many of these data were used by Wills *et al.* (2000) to develop the site conditions map of California and later by Wald and Allen (2007). In effect, the Silva (personal commun., 2007) V_{S30} database used in this study is an updated version of the V_{S30} values used by both Wills *et al.* (2000) and Wald and Allen (2007). Wills *et al.* (2000) used 556 V_{S30} values; for their California map Wald and Allen (2007) used an updated database of 767 values as well as other values in active tectonic regions (e.g., Utah, 204; Taiwan, 387; and Italy, 43).

We use all 853 V_{S30} results from the updated (2007) Silva V_{S30} database to develop our terrain-based V_{S30} proxy model. To maintain a framework consistent with the Wills *et al.* (2000) and Wald and Allen (2007) approaches, we also do not appraise the quality of the V_{S30} values used in our study other than acknowledging the lack of quality assurance in the V_{S30} data, and addressing the limitations of V_{S30} as a measure of site conditions later. A description of the distribution of terrain categories classified in California and of the number of V_{S30} results in each terrain class is provided by histograms in Figures 3 and 4, respectively.

Methods

Developing the Terrain-Based V_{S30} Proxy Model

After establishing our terrain-based framework, we compare available V_{S30} values with our terrain classification map. Using the spatial locations for the 853 V_{S30} values, we determine the corresponding terrain-type class (Figs. 2, 5a,b). No V_{S30} values are associated with areas identified as our terrain category of class 13, which fit the description of incised terraces (Figs. 3, 4). Such a result might be expected because relatively few V_{S30} values are available at hard rock sites. V_{S30} values generally tend to be in major urban/metropolitan areas that are typically located on soft basin-type terrains (Fig. 3). Also, in the case of class 13, the terrace terrain-type has the second-lowest ranking number (2106 pixel cells) of occurrences in California representing $<1\%$ of the total cells (574,049) that make up the regionally extensive and morphologically diverse state (Fig. 6).

Following the association of V_{S30} values to the 15 site classes, we randomly select 556 values (same number used by Wills *et al.*, 2000) and recalculate the mean V_{S30} values for each of the 15 terrain classes. Here, the portion of V_{S30} values assigned within each terrain class is based on the distribution of the original 853 values we find in each of the 15 terrain categories (Figs. 3, 4). We use a standard cross-validation (nonreplacement) method to compare our prediction to the remaining 297 V_{S30} measurements by examining its mean-squared prediction error (MSPE):

$$\Delta = \frac{1}{N}\sum_{i=1}^{N}(V_{S30_i} - \text{prediction}_i)^2.$$

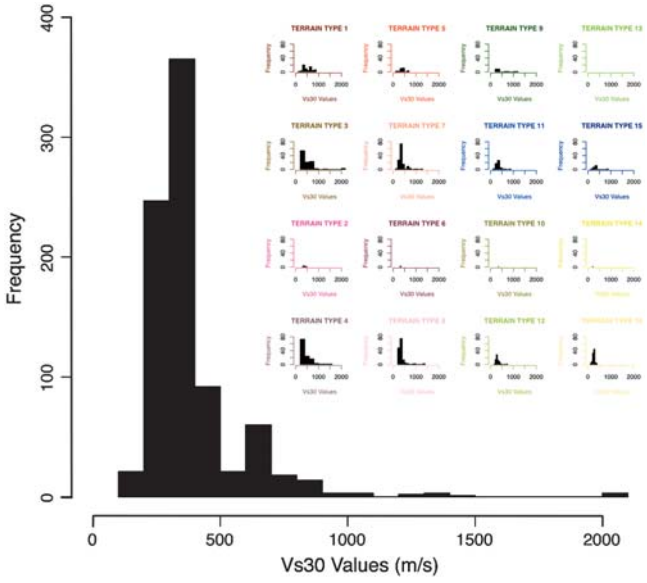

Figure 4. Histograms indicating the overall frequency distribution of V_{S30} values (large histogram) and the frequency distribution of V_{S30} values in each terrain-type (histograms arranged according to terrain classification scheme in Figure 2). Note: No V_{S30} values are associated with terrain types described by Iwahashi and Pike (2007) as incised terraces.

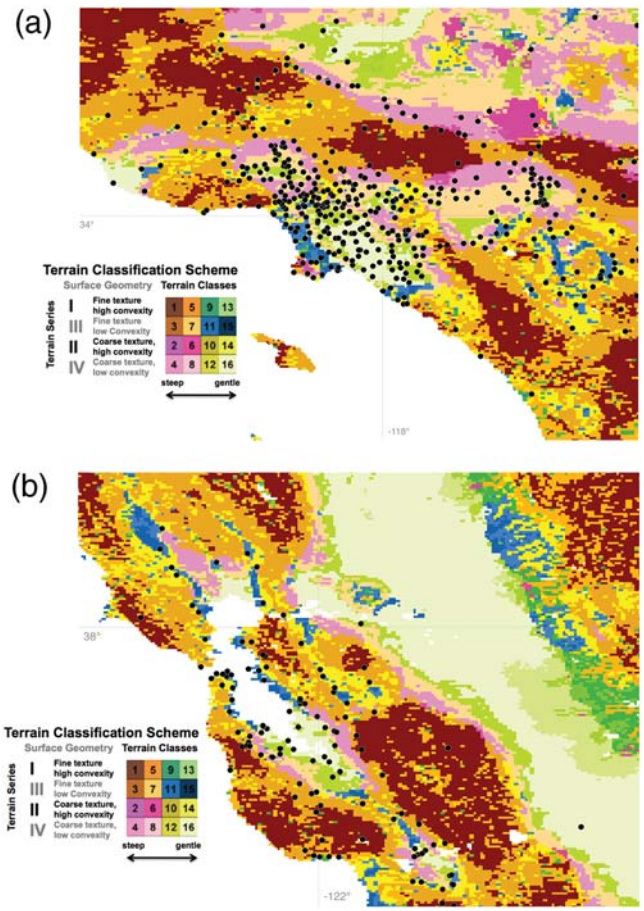

Figure 5. (a) Terrain map of California (Iwahashi and Pike, 2007) in the Los Angeles regions with locations (black circles) of V_{S30} values. (b) Terrain map of California (Iwahashi and Pike, 2007) in the San Francisco region with locations (black circles) of V_{S30} values.

The routine, from random selection through cross-validation, is iterated 1000 times. At the end of these runs, we select the model with the lowest MSPE (14,414) as our best predictor of V_{S30} (Fig. 7) for each terrain type (Fig. 8).

Results

Evaluation of Model Performances

To assess our terrain-based V_{S30} proxy model beyond inferential statistics, we evaluate its performance compared with the geologic unit-based model (Wills *et al.*, 2000) and the topographic slope-based model (Wald and Allen, 2007). Because both of the latter approaches effectively employed the same V_{S30} data we use for the same area (state of California), albeit in varying quantities, Wills *et al.* (2000) and Wald and Allen (2007) relied on approximately 65% and 91% (respectively) of the values from the updated V_{S30} database; the commonality in data source reasonably supports the basis for our direct comparison of the different models. Using the spatial join function in the ESRI ArcInfo (version 9.3) application program, we associate the 853 sites to the 7 categories of the Wills *et al.* (2000) model (data provided by C. Wills, written commun., 2006). On the basis of the site categories in the joined GIS-database tables, we translate every instance of their categories from class letter designations to corresponding mean V_{S30} values (Wills *et al.*, 2000). For the Wald and Allen (2007) model (see Data and Resources), we develop a simple algorithm using the nearest neighbor approach (Clark and Evans, 1954) to associate their coordinate-and-V_{S30}-value relations with the 853 V_{S30} sites. To identify a straightforward and common parameter between the Wills *et al.* (2000) geologic unit-based categories and our own model, we use their reported mean V_{S30} values for each of their seven expanded NEHRP-based classes. For the Wald and Allen (2007) topographic slope-based categories, mean V_{S30} values were not reported (or readily available) so we use the assigned V_{S30} values in their downloadable text-file. We use the Warnes (2004) bandplot (locally smoothed mean and standard deviation) function and a customized subroutine, all in the open-source statistical environment *R* (see Data and Resources), to consider the local mean and variance, and calculate the typical standard deviation (σ_{typical}) of each model (Fig. 9). As a result, we find the Wills *et al.* (2000) model has the highest σ_{typical}, approximately 195 m/s, while the Wald and Allen (2007) model and our model have approximately 142 m/s and 128 m/s, respectively. As part of our statistical analysis, we also calculate the median of the MSPE for each model and determine that all models perform significantly better than random (Fig. 10).

As an extension of our efforts to develop a site characterization map for California, we explore the utility of our approach for a V_{S30} map of the contiguous United States (Fig. 11; Yong *et al.*, 2010). Like the California map, the V_{S30} map of the United States is derived using the automated IP07 classification method that relies on the same taxonomic criteria (slope gradient, local convexity, and surface texture) developed from geomorphometry to also identify 16 terrain types from the same 1-km spatial resolution (SRTM30) DEMs. On the basis of the California model, we present our preliminary U.S. map by applying the same terrain classification approach and adopting the same V_{S30} relations determined for each terrain type in California. Although our database of 325 V_{S30} observations (for source, see Data and Resources) measured outside of California is currently too sparse to permit a rigorous validation of the terrain-based model, we plot the locally smoothed mean and standard deviation (Fig. 12) of the terrain-based and topographic slope models for a preliminary assessment of their performances. At the time of our investigation, no geology-unit based V_{S30} estimation model of the contiguous United States was readily available. When comparing the σ_{typical} of each model for their respective performances in California and the United States (Figs. 9, 12), we note the changes ($\Delta\sigma$) in their σ_{typical} and find that the terrain-based model ($\Delta\sigma = +19$ m/s) is more stable than the topographic slope-based model ($\Delta\sigma = 65$ m/s). We also find an appreciable improvement in the predictive performance of the terrain-based model

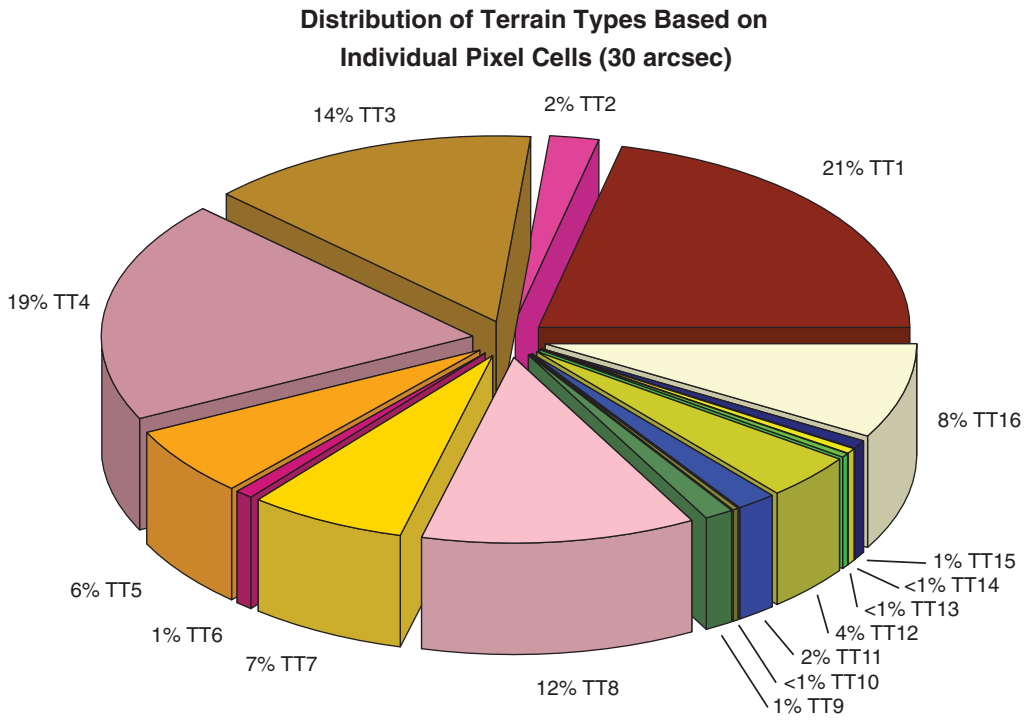

Figure 6. Three-dimensional pie chart describing terrain-types (TT) and corresponding percentage of occurrences (pixel cells) found in California.

($\sigma_{\text{typical}} = 147$ m/s) over the topographic sloped-based model ($\sigma_{\text{typical}} = 213$ m/s).

Discussion

Our primary motives for developing the terrain-based model are effectively twofold: (1) to more fully exploit the nature of geomorphology as a proxy for site amplification and (2) to advance an approach that is completely objective with results firmly supported by robust statistical methods. We acknowledge that a significant portion (76%) of the data used to develop our model are based on inferred V_{S30} values and that such a strong dependence fundamentally contradicts the established principles of model development. However, it is necessary to depend on these proxy-based estimates because there are simply not enough openly available

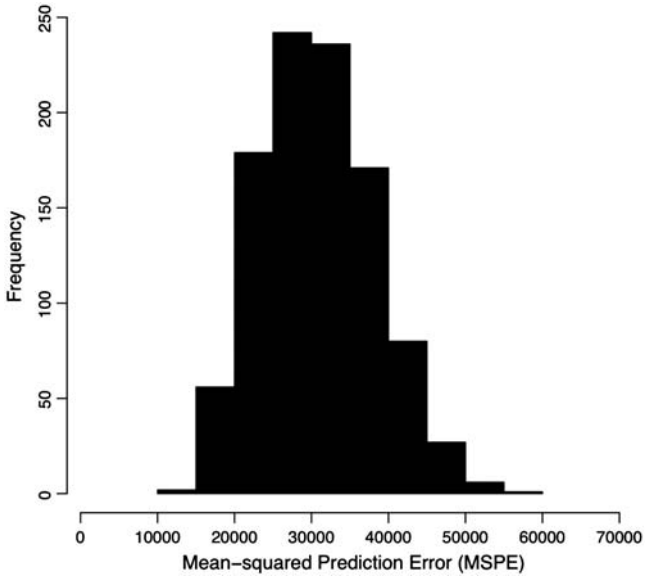

Figure 7. Histogram showing the distribution of MSPEs where the lowest score (14,414.01) is used as indicator for the model with best estimations of V_{S30}.

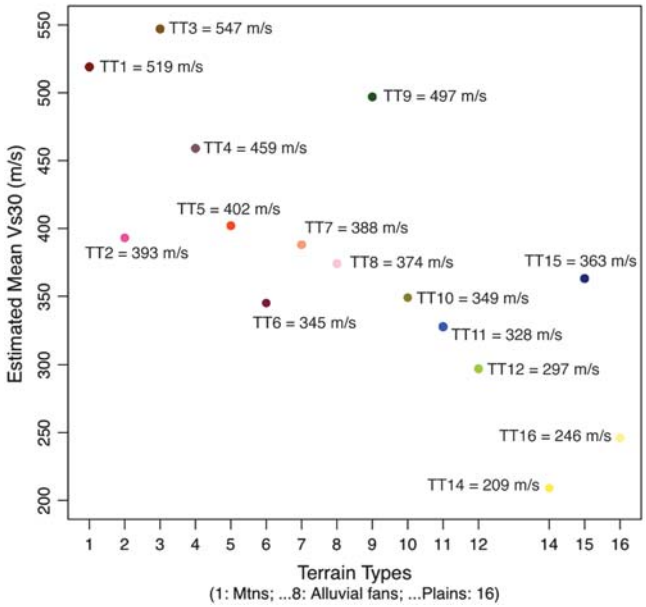

Figure 8. Plot describing mean V_{S30} values found in terrain-types. Note: No estimate is provided for Terrain-type 13.

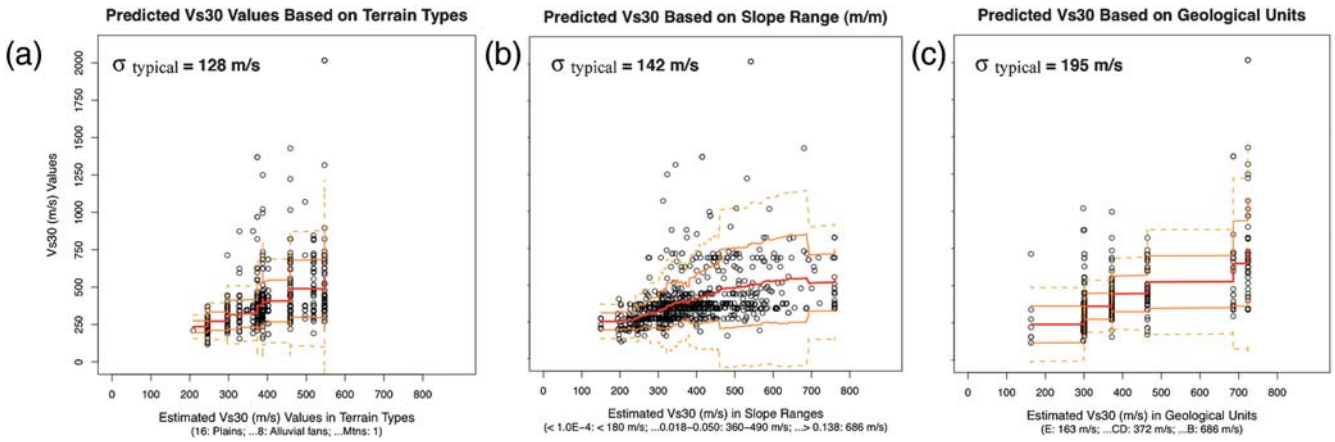

Figure 9. Plots describing the smoothed mean and standard deviation for 853 V_{S30} values based on both measured and inferred data (California) against the (a) terrain-based, (b) topographic-slope-based, and (c) geological unit-based estimation models. Red lines indicate the smoothed mean values, orange dashes indicate the one standard deviation from the mean and yellow dash indicates two standard deviations from the mean.

measured values to meet standard statistical requirements. As previously indicated, the reliance on inferential data has precedent in state-of-the-art approaches, for example, the geology-based and topographic slope-based models were also conditioned on effectively the same inference-based data. In the following section, we expand on these and other issues associated with our dependence on proxy-based models and the use of V_{S30} for estimating site amplification.

Although geomorphology is (in principle) an appropriate proxy for capturing the key characteristics of both geology and topography (Peterson, 1981), one initial concern is the role that terrain types play in our model. Because the proxies are by definition estimates of the predictors (i.e., terrain type is the proxy for both geology and topography, which are in turn proxies for V_{S30}, which itself is a proxy for site amplification), this type of multitiered approach can inherently compound the uncertainties associated with each proxy. Despite these concerns, we show that the terrain-based approach is capable of estimating material (soil) properties that control site amplification.

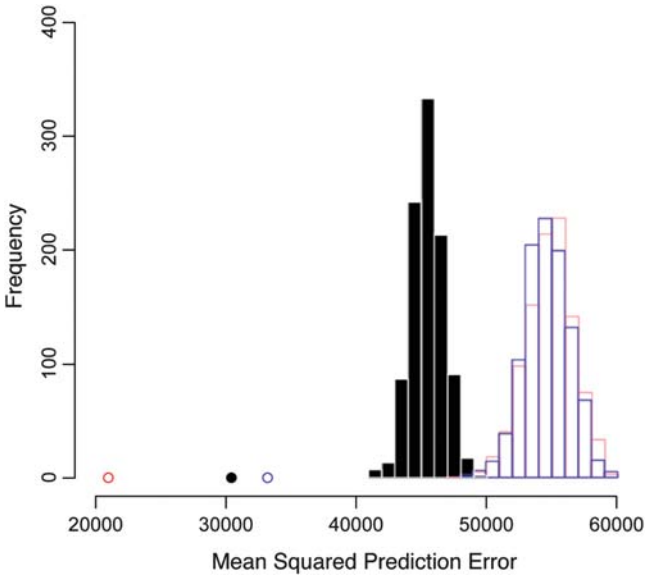

Figure 10. Histograms describing the frequency distributions (853 samples) of the calculated MSPE from randomization testing for the geological unit-based (red), terrain-based (black), and topographic-slope-based (blue) models and their actual (median) MSPE values (color-matching circles).

One utility of the terrain-based approach is that material properties, which are not directly related to a single proxy, such as slope, can be inferred. This is especially true when estimating V_{S30} in flat or steep topographic areas. For such instances, the combination of terrain nomenclature (Hammond, 1954, 1964) and the IP07 classification scheme provides a basis for differentiating material properties that do not always directly associate with individual proxies, such as topographic slope. For example, in a volcanic plain with flat topographic slope and high shear-wave velocity, topographic slope (alone) would predict high amplification (low V_{S30}). Furthermore, in active tectonic regions, lower shear-wave velocity (high amplification) with steep topographic slope consisting of young sediments, a slope-based model would predict high V_{S30}. In both cases, geology plays the key role for distinguishing the exceptions in the relationship between the nature of topographic slope and V_{S30}. A manual analysis of the same regions (e.g., Modoc Plateau, Mendocino, etc.) in each model indicates significant differences in V_{S30} predicted (Fig. 2; Wald and Allen, 2007). Actual observations of V_{S30} measurements from the regions are necessary to properly evaluate the performance of both models.

Clearly, the arbitrary nature of 30 meters as the key parameter for depth from the surface in the V_{S30} term is expected to have a strong first-order consequence on the effectiveness of the proxies. To address this issue, the use of more complicated parameterizations might be useful. Approaches worthy of further investigation include the one-quarter wavelength method (Joyner *et al.*, 1981), or extra soil

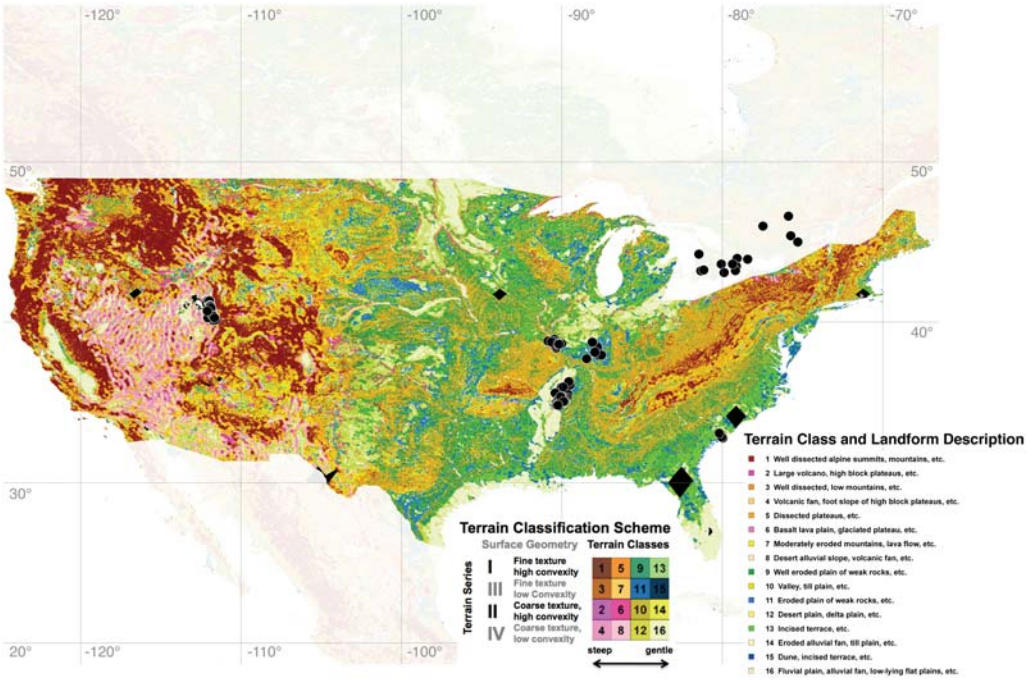

Figure 11. Terrain map of the contiguous United States (Iwahashi and Pike, 2007) with locations (black circles) of 325 V_{S30} measurement-based values, including 14 records from Canada. Areas of poor DEM quality were left unclassified and designated by the authors as black polygons.

site considerations, for example, sediment thickness coupled with shear-wave velocity (s_L) first proposed by Seed *et al.* (1976) and recently reviewed by Lee and Trifunac (2010). Another possible approach is to circumvent V_{S30} as a proxy by determining if the IP07 terrain scheme can be used to directly estimate site amplification; this will require the use of records from both permanent (e.g., California Integrated Seismographic Network) and temporary (e.g., USArray) seismographic station sites in order to meet minimum spatial sampling requirements. For now, in the absence of

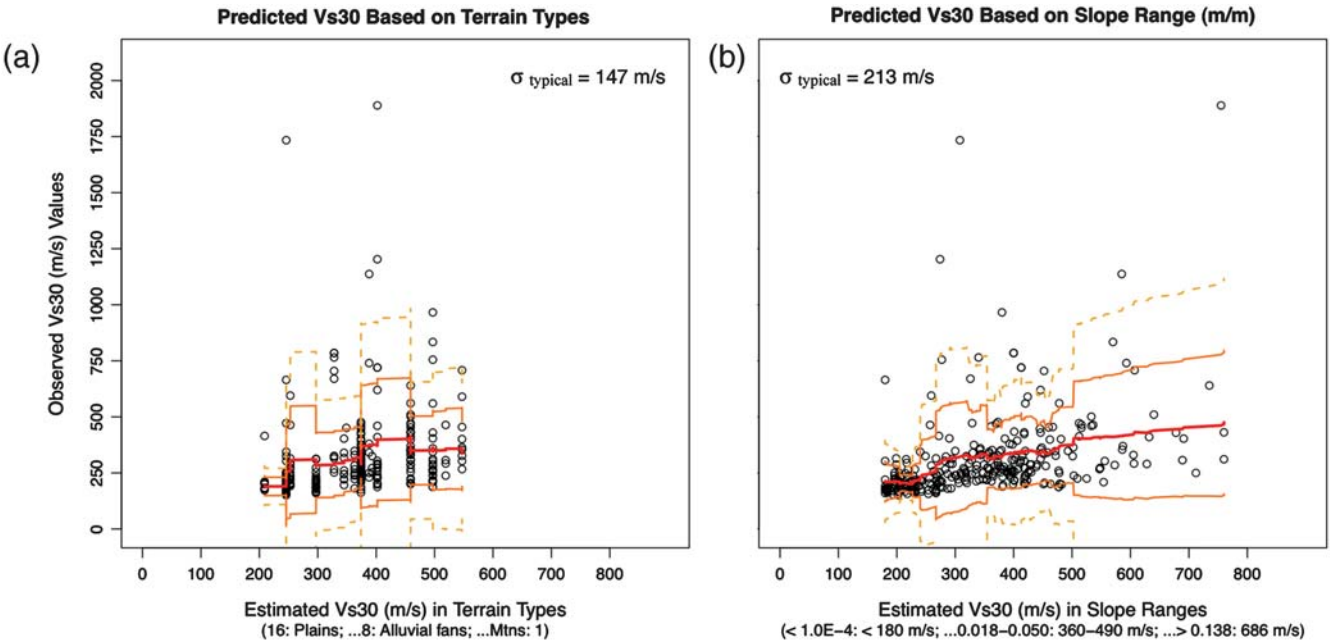

Figure 12. Plots describing the smoothed mean and standard deviation for V_{S30} values based on 325 measurements (contiguous United States, outside of California) against the (a) terrain-based and (b) topographic-slope based estimation models. Red lines indicate the smoothed mean values, orange dashes indicate the one standard deviation from the mean and yellow dash indicates two standard deviations from the mean.

a more robust site parameter, we follow accepted convention and proceed with V_{S30} as the proxy for material induced amplification.

When considering the uncertainties associated with the V_{S30} term, in addition to other vexing factors such as artifacts associated with spatially referenced data, not to mention circularity issues relating to the comingling of measured- and inferred- (map-based) values found in the V_{S30} database, we recognize that these limitations will tend to confound our dependence on strict adherence to our statistically rigorous systematic approach. Rather, it is in the face of these uncertainties that we consider an objective and systematic approach based firmly on statistics to be most warranted when investigating the predictive capability of the IP07 terrain model to estimate V_{S30}. By statistically determining that each model performs significantly better than random (Fig. 10), we confirm that all of the proxies have a strong correlation to V_{S30}. In an attempt to further explore the utility of each proxy, we rank the results (Fig. 10) and find the geological unit-based model has the lowest median MSPE score, followed by the terrain-based and the topographic slope-based models, respectively. An important caveat when using rank to compare the models against each other is that despite the commonality of the Silva V_{S30} database used in each model, the differences in the methods used to develop each model have a strong predisposed effect on the resultant rank order. These effects remain influential even if we consider only a few factors from each model. For example, when developing their geological unit-based model, Wills *et al.* (2000) trained and tested on the same 556 V_{S30} values, essentially using 100% of the available data, only 65% of the 853 values available in the updated (2007) Silva database. On the basis of their own validation test, the 25% error rate is mostly attributable to the plotting of clustered data points on low-resolution maps. Nevertheless, their model should have the strongest advantage when tested for statistical significance because their empirical approach allows for manual intervention in the lumping of the geologic units, and thus positively affects the apparent degree of correlation to V_{S30} values. Although we use all available V_{S30} values from the Silva database, our approach is strictly systematic: we randomly select 556 samples and reserve the remaining 297 values for cross-validation while iterating this process 1000 times to find the lowest MSPE to establish our estimates. Like Wills *et al.* (2000), we also use 65% of the updated Silva database to construct our model; rather, it is the effect of our rigorous cross-validations that has the strongest influence on our secondary position in the ranking. Wald and Allen (2007) also employed empirical methods when correlating slope to the NEHRP categories; that is, to improve resolution where possible, the official NEHRP groupings were further subdivided into narrower velocity windows. Unlike either aforementioned models, their California samples only consist of 767 V_{S30} values from the Silva database. By relying on only about half (55%) of their overall samples (approximately 90% of the California samples in the Silva database used in this study) as the dependent variable when developing the topographic-slope model, our test of the Wald and Allen (2007) active tectonic regional estimates against the California-centric Silva V_{S30} values places their model at the least advantageous position as demonstrated by the calculated MSPE (alone) and the lowest (apparent) position observed in our ranking. Hence, we intend these observations only for discussion and not present them as part of our results, because we do not consider these comparisons to be a robust evaluation of model performances.

We also consider standard regression analysis methods to evaluate the performance of the models by themselves and against each other. Here, we encounter another fundamental statistic-related problem inherent in all three models. The respective proxies of the geologic unit-, topographic slope-, and terrain-based models do not readily fit in any formalized description of a statistical variable; that is, although the proxies are defined as names (nominal characteristic), at the same time their associations to V_{S30} are strongly dependent on numeric values tied to predefined (discrete variable) ranges on a common continuous scale with meaningful hierarchical order (ordinal characteristic). For example, the Wills *et al.* (2000) and Wald and Allen (2007) models are based on classes that are modified versions of the original NEHRP categories (e.g., B, C, D, and E) where the respective proxies are tied to V_{S30} ranges. In our terrain-based model, the 16 classes represent the gradation of terrain types from mountains through alluvial fans to flat basins. Although these classes are not related to the discrete steps in the NEHRP categories as implemented by the others, our estimates of V_{S30} in each class are averaged on the basis of the portion of the 853 V_{S30} values that spatially coregister within each individual class. From the perspective of statistical analyses, regardless of the nature of the relationship of the proxies to V_{S30}, the nominal character (nonnumeric) of the respective proxies precludes direct comparison.

It is in the utility of the standard deviation (σ) where we find a meaningful, yet simple, statistical measure for evaluating the performance of each model and the comparisons against each other. Here, based on the 853 values from the Silva database, we look for the typical amount of dispersion (σ_{typical}) in the V_{S30} values with respect to each model's proxy-based estimates. Because our terrain-based estimates depend on the average of the V_{S30} values for the respective terrain types, we choose the equivalent average V_{S30} values described by Wills *et al.* (2000) for their seven respective classes to calculate the σ_{typical} of their geologic unit-based model. Although the topographic slope-based approach initially estimated V_{S30} on the basis of the median velocity of the subdivided NEHRP classifications, these results are not fully described in Wald and Allen (2007), and we find no means to calculate the average (nor median) of their V_{S30} estimates for each class because the complete data set is no longer available (D. J. Wald, written commun., 2009). In the face of these deficiencies, we instead rely on the latest and most readily available estimates from their USGS web site (see Data and Resources) for

comparison with the Silva V_{S30} data. Because we consider their estimates to possess the most statistically advantageous traits on the basis of the continuous nature and sheer abundance (1401) of the V_{S30} data used in their analyses, we expect Wald and Allen's (2007) estimations to yield the least σ_{typical}; hence, they rank as the best performing model when compared against the discretized proxies in geologic unit- and terrain-based models. Instead, as noted earlier, we find our terrain-based model to yield the lowest σ_{typical} (approximately 128 m/s), followed closely by the topographic slope-based models (approximately 142 m/s) and the geologic unit-based models (approximately 195 m/s), respectively.

Conclusions

To varying degrees of success, individual proxies such as mapped surface geology (Wills *et al.*, 2000) and topographic slope (Wald and Allen, 2007) have been used to develop V_{S30} (the time-averaged shear-wave velocity in the upper 30 m) predictions for California. This type of broad-brush (single-parameter) approach to site characterization has been demonstrated to be useful for some purposes, especially in the absence of measured V_{S30} values (Frankel *et al.*, 2010; Thompson *et al.*, 2010). However, clearly no single proxy can fully account for variations in material properties that control V_{S30}. In an attempt to improve site characterization, we develop a method based on geomorphometry (slope gradient, local convexity, and surface texture) to better predict their material properties and thus estimate V_{S30}. We use results from the automated terrain classification method based on taxonomic criteria derived from geomorphometry to systematically identify 16 terrain types from a 1-km spatial resolution (SRTM30 data) DEM of California (Iwahashi and Pike, 2007). Using 853 V_{S30} values, we assign V_{S30} values that are the average of values within each terrain type. We then compare these values with the single-proxy prediction models to determine relative performances. As expected, the overall results indicate that our approach performs better than the characterizations based on single proxies. In each case, we note that the reliance on inconsistent geologic data sets (Wills *et al.*, 2000), or a simple parameter, such as the 1-km resolution SRTM30 DEM-derived slope gradient (Wald and Allen, 2007), are less discriminating in accounting for variations in site conditions. Furthermore, our preliminary studies also indicate the terrain-based framework to be more effective than topographic-slope alone when comparisons are extended to the contiguous United States (Yong *et al.*, 2010). Despite statistical tests that indicate our method to be, at worst case, a marginal improvement over the next-best method, we also note that by employing objective and systematic methods, as well as a multifaceted framework that encapsulates geology and topographic slope, our approach holds the most promise for improving the prediction of material (soil) properties when estimating site amplification. As new and spatially distributed measurement-based V_{S30} data become available, we expect to be able to improve our predictive capabilities.

Data and Resources

GIS data of worldwide automated terrain classification map from SRTM30 can be downloaded at http://gisstar.gsi.go.jp/terrain/front_page.htm (last accessed February 2009). The Wald and Allen (2007) model from the USGS is an evenly sampled xyz file, available at http://earthquake.usgs.gov/research/hazmaps/interactive/vs30/predefined.php (last accessed July 2008). The Warnes (2004) bandplot function (locally smoothed mean and standard deviation) in the statistical environment R is available at www.r-project.org/ (last accessed April 2009).

The updated NGA database flat file (including measurement-based V_{S30} observations at 325 recording sites outside of California), referred in this article as the NGA strong-motion database, can be downloaded at http://peer.berkeley.edu/nga/flatfile.html (last accessed January 2010).

Acknowledgments

We greatly appreciate the helpful comments from associate editor Martin C. Chapman and the two anonymous reviewers. We are also grateful to David M. Boore, Eric M. Thompson, Brad T. Aagaard, Christopher D. Stephens, and Dominic Assimaki for constructive feedback that greatly improved our manuscript. We are in debt to Chris H. Cramer, Jack K. Odum, Kris L. Pankow, Walter Silva, and Robert A. Williams for providing the V_{S30} data. Special thanks to Walter Silva and Masashi Matsuoka who provided many important insightful discussions about the role of V_{S30} in site amplification. Lisa Christiansen and Stan Schwartz were instrumental in preprocessing the GIS data. Lastly, we thank Richard J. Pike and Chris J. Wills for early discussions on the development of our project.

References

Aki, K. (1988). Local site effects on strong ground motion, in *Proc. Earthq. Eng. and Soil Dyn II: Recent Advances in Ground-Motion Evaluation*, ASCE, Park City, Utah, 27–30 June, 103–155.

Akinci, A., L. Malagnini, and F. Sabetta (2010). Characteristics of the strong motions from the 6 April 2009 L'Aquila earthquake, Italy, *Soil Dynam. Earthquake Eng* **30,** 320–335.

Ameri, G., M. Massa, D. Bindi, E. D'Alema, A. Gorini, L. Luzi, S. Marzorati, F. Pacor, R. Paolucci, R. Puglia, and C. Smerzini (2009). The 6 April 2009 M_w 6.3 L'Aquila (Central Italy) earthquake: Strong-motions observations, *Seismol. Res. Lett.* **80,** no. 6, 951–966.

Anderson, J. G., Y. Lee, Y. Zeng, and S. Day (1996). Control of strong motion by the upper 30 meters, *Bull. Seismol. Soc. Am.* **86,** no. 6, 1749–1759.

Assimaki, D., G. Gazetas, and E. Kausel (2005). Effects of local soil conditions on the topographic aggravation of seismic motion: Parametric investigation and recorded field evidence from the 1999 Athens earthquake, *Bull. Seismol. Soc. Am.* **95,** no. 3, 1059–1089.

Boore, D. M. (2004). Can site response be predicted?, *J. Earthq. Eng.* **8,** no. 1, 1–41.

Boore, D. M. (2006). Determining subsurface shear-wave velocities: A review, in *Proc. 3rd International Symposium on the Effects of Surface Geology on Seismic Motion*, Grenoble, France, 30 August – 1 September.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1993). Estimation of response spectra and peak accelerations from western North American earthquakes: An interim report, part 1, in *U.S. Geol. Surv. Open-File Rept. 93-509*, 69 pp.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1994). Estimation of response spectra and peak accelerations from western North American

earthquakes: An interim report, part 2, in *U.S. Geol. Surv. Open-File Rept. 94-127*, 40 pp.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1997). Equations for estimating horizontal response spectra and peak acceleration from western north American earthquakes: A summary of recent work, *Seismol. Res. Lett.* **68,** no. 1, 128–153.

Borcherdt, R. D. (1970). Effects of local geology on ground motion near San Francisco Bay, *Bull. Seismol. Soc. Am.* **60,** no. 1, 29–61.

Borcherdt, R. D. (1994). Estimates of site-dependent response spectra for design (methodology and justification), *Earthquake Spectra* **10,** 617–653.

Borcherdt, R. D., and G. Glassmoyer (1992). On the characteristics of local geology and their influence on ground motions generated by the Loma Prieta earthquake in the San Francisco bay region, California, *Bull. Seismol. Soc. Am.* **82,** no. 2, 603–641.

Borcherdt, R. D., C. M. Wentworth, A. Janssen, T. Fumal, and J. Gibbs (1991). Methodology for predictive GIS mapping of special study zones for strong ground shaking in the San Francisco bay region, in *Proc. 4th Int. Conf. on Seismic Zonation*, Earthquake Engineering Research Institute, Oakland, California 54**,** 5–552.

Chiou, B., R. Darragh, N. Gregor, and W. Silva (2008). NGA project strong-motion database, *Earthquake Spectra* **24,** no. 1, 23–44.

Clark, P. J., and F. C. Evans (1954). Distance to nearest neighbor as a measure of spatial relations in populations, *Ecol.* **35**, 445–453.

Dehn, M., H. Gärtner, and R. Dikau (2001). Principles of semantic modeling of landform structures, *Comp. and Geosci.* **27,** 1005–1010.

Dobry, R., R. D. Borcherdt, C. B. Crouse, I. M. Idriss, W. B. Joyner, G. R. Martin, M. S. Power, E. E. Rinne, and R. B. Seed (2000). New site coefficients and site classification system used in recent building seismic code provisions, *Earthquake Spectra* **16,** no. 1, 41–67.

Ellis, M. A., A. L. Densmore, and R. S. Anderson (1999). Development of mountainous topography in the Basin Ranges, USA, *Basin Res.* **11,** 21–41.

Farr, T. G., P. A. Rosen, E. Caro, R. Crippen, R. Duren, S. Hensley, M. Kobrick, M. Paller, E. Rodriquez, L. Roth, D. Seal, S. Shaffer, J. Shimada, M. Werner, M. Oskin, D. Burbank, and D. Alsdorf (2007). *The Shuttle Radar Topography Mission*, Jet Propulsion Laboratory/ California Institute of Technology, Pasadena, California, 43 pp.

Fenneman, N. M., and D. W. Johnson (1946). Physical division of the United States, *U.S. Geol. Surv. Special Map*, scale 1:7,000,000.

Field, E. H. (2000). A modified ground-motion attenuation relationship for southern California that accounts for detailed site classification and a basin-depth effect, *Bull. Seismol. Soc. Am.* **90,** no. 6BS209–S221.

Frankel, A. D., D. L. Carver, and R. A. Williams (2002). Nonlinear and linear site response and basin effects in Seattle for the *M* 6.8 Nisqually, Washington, earthquake, *Bull. Seismol. Soc. Am.* **92,** no. 6, 2090–2109.

Frankel, A. D., S. Harmsen, C. Mueller, E. Calais, and J. Haase (2010). Documentation for initial seismic hazard maps for Haiti, in *U.S. Geol. Surv. Open-File Rept. 2010-1067*, 12 pp.

Franklin, S. E. (1987). Geomorphometric processing of digital elevation models, *Comp. and Geosci.* **13,** no. 6, 603–609.

Fumal, T. E. (1978). Correlations between seismic wave velocities and physical properties of near-surface geologic materials in the southern San Francisco Bay region, California, in *U.S. Geol. Surv. Open-File Rept. 78-1067*, 114 pp.

Fumal, T. E., and J. C. Tinsley (1985). Mapping shear-wave velocities in near-surface geological materials, in *Evaluating Earthquake Hazards in the Los Angeles Region—An Earth Science Perspective*,J. I. Ziony (Editor), in *U.S. Geol. Surv. Profess. Pap. 81-1360*, 127–150.

Fumal, T. E., J. F. Gibbs, and E. F. Roth (1981). *In-situ* measurements of seismic velocities at 19 locations in the Los Angeles, California, region, in *U.S. Geol. Surv. Open-File Rept. 81-399*, 123 pp.

Fumal, T. E., J. F. Gibbs, and E. F. Roth (1982). *In-situ* measurements of seismic velocities at 22 locations in the Los Angeles, California region, in *U.S. Geol. Surv. Open-File Rept. 82-833*, 140 pp.

Fumal, T. E., J. F. Gibbs, and E. F. Roth (1984). *In-situ* measurements of seismic velocities at 16 locations in the Los Angeles, California, region in *U.S. Geol. Surv. Open-File Rept. 84-683*, 109 pp.

Gao, S., H. Liu, P. M. Davis, and L. Knopoff (1996). Localized amplification of seismic waves and correlation with damage due to the Northridge earthquake: Evidence for focusing in Santa Monica, *Bull. Seismol. Soc. Am.* **86,** no. 1B, S209–S230.

Gibbs, J. F., T. E. Fumal, and E. F. Roth (1980). *In-situ* measurements of seismic velocities at 27 locations in the Los Angeles, California region, in *U.S. Geol. Surv. Open-File Rept. 80-378*, 169 pp.

Graf, W. L. E. (1987). *Geomorphic Systems of North American; Centennial Special Volume 2*, Geological Society of America, Inc., Boulder, Colorado, 643 pp.

Hammond, E. H. (1954). Small-scale continental landform maps, *Ann. Assoc. Am. Geogr.* **44,** 33–42.

Hammond, E. H. (1964). Analysis of properties in land form geography: An application to broad-scale landform mapping, *Ann. Assoc. Am. Geogr.* **54,** 11–19.

Harden, J. W. (1990). Soil development on stable landforms and implications for landscape studies, *Geomorphology* **3,** 391–398.

Holzer, T. L., M. J. Bennett, T. E. Noce, and J. C. Tinsley (2005). Shear-wave velocity of surficial geologic sediments in northern California: Statistical distributions and depth dependence, *Earthquake Spectra* **21,** no. 1, 161–177.

Horn, B. K. P. (1981). Hill shading and the reflectance map, *Proc. IEEE* **69,** no. 1, 14–47.

Hough, S. E., J. R. Altidor, D. Anglade, D. Given, M. G. Janvier, J. Z. Maharrey, M. Meremonte, B. S. L. Mildor, C. Prépetit, and A. Yong 2010. Localized damage caused by topographic amplification during the 2010 *M* 7.0 Haiti earthquake, *Nature Geoscience* **3,** 778–782 (http://www.nature.com/ngeo/journal/vaop/ncurrent/pdf/ngeo988.pdf, last accessed October 2010).

Hough, S. E., P. A. Friberg, R. Busby, E. H. Field, K. H. Jacob, and R. D. Borcherdt (1990). Sediment-induced amplification and the collapse of the Nimitz Freeway, *Nature* **344,** 853–855.

Iwahashi, J., and I. Kamiya (1995). Landform classification using digital elevation model by the skills of image processing—mainly using the Digital National Information, *Geoinformatics* **6,** no. 2, 97–108 (in Japanese with English abstract).

Iwahashi, J., and R. J. Pike (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature, *Geomorphology* **86,** 409–440.

Iwahashi, J., I. Kamiya, and M. Matsuoka (2010). Regression analysis of V_{S30} using topographic attributes from a 50-m DEM, *Geomorphology* **117,** 202–205.

Iwahashi, J., S. Watanabe, and T. Furuya (2001). Landform analysis of slope movements using DEM in Higashikubiki area, Japan, *Comput. Geosci.* **27,** 851–865.

Jensen, J. R. (2007). *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall, Upper Saddle River, New Jersey, 526 pp.

Joyner, W. B., R. E. Warrick, and T. E. Fumal (1981). The effect of Quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979, *Bull. Seismol. Soc. Am.* **71,** no. 4, 1333–1350.

Lee, V. W., and M. D. Trifunac (2010). Should average shear-wave velocity in the top 30 m of soil be used to describe seismic amplification?, *Soil Dynam. Earthquake Eng.* **30,** 1250–1258.

Lee, S., Y. Chan, D. Komatitsch, B. Huang, and J. Tromp (2009). Effects of realistic surface topography on seismic ground motion in the Yangminshan region of Taiwan based upon the spectral-element method and LiDAR DTM, *Bull. Seismol. Soc. Am.* **99,** 2A, 681–693.

Magistrale, H., S. Day, R. W. Clayton, and R. Graves (2000). The SCEC southern California reference three-dimensional seismic velocity model version 2, *Bull. Seismol. Soc. Am.* **90,** no. 6B S65–S76.

Matsuoka, M., and S. Midorikawa (1995). GIS-Based Integrated Seismic Hazard Mapping for a Large Metropolitan Area, *Proc. fifth Intern. Conf. on Seismic Zonation* **2,** 1334–1341.

Matsuoka, M., K. Wakamatsu, K. Fujimoto, and S. Midorikawa (2006). Average shear-wave velocity mapping using Japan engineering

geomorphogic classification map, *J. Struc. Mech. Earthq. Eng.* **23,** no. 157s–68s.

Minár, J., and I. S. Evans (2008). Elementary forms for land surface segmentation: The theoretical basis of terrain analysis and geomorphological mapping, *Geomorphology* **95,** 236–259.

Moss, R. E. S. (2008). Quantifying measurement uncertainty of 30-meter shear-wave velocity, *Bull. Seismol. Soc. Am.* **98**, no. 3, 1399–1411.

Murray, B., and M. A. Fonstad (2007). Preface: Complexity (and simplicity) in landscapes, *Geomorphology* **91,** 173–177.

Oliver, M. A., and R. Webster (1986). Semi-variograms for modeling the spatial pattern of landform and soil properties, *Earth Surface Processes and Landforms* **11,** 491–504.

Olsen, K. B., S. M. Day, J. B. Minster, Y. Cui, A. Chourasia, M. Faerman, R. Moore, P. Maechling, and T. Jordan (2006). Strong shaking in Los Angeles expected from southern San Andreas earthquake, *Geophys Res. Lett.* **33,** L07305.

Park, S., and S. Elrick (1998). Predictions of shear-wave velocities in southern California using surface geology, *Bull. Seismol. Soc. Am.* **88,** no. 3, 677–685.

Peterson, F. F. (1981). Landforms of the basin and range province; defined for soil survey, *Nevada Agricultural Experiment Station, University of Nevada Technical Bulletin* **28,** Max C. Fleischmann College of Agriculture, Reno, Nevada, 52 pp.

Petersen, M., W. Bryant, C. Cramer, M. Reichle, and C. Real (1997). Seismic ground-motion hazard mapping incorporating site effects for Los Angeles, Orange, and Ventura counties, California: A geographical information system application, *Bull. Seismol. Soc. Am.* **87,** no. 1, 249–255.

Philips, J. D. (2007). The perfect landscape, *Geomorphology* **84,** 159–169.

Pike, R. J., I. S. Evans, and T. Hengl (2009). Geomorphometry: A brief guide in *Geomorphometry: Concepts, Software, Applications*, T. Hengl, H. I. Reuter (Editors), *Series Developments in Soil Science*, **33,** Elsevier, Amsterdam, 3–30.

Romero, S. M., and G. J. Rix (2001). Ground motion amplification of soils in the upper Mississippi embayment, *Mid-America Earthquake Center at School of Civil and Environmental Engineering at Georgia Institute of Technology*, Report No. GIT-CEE/GEO-01-1, 46–112.

Saco, P. M., G. R. Willgoose, and G. R. Hancock (2006). Spatial organization of soil depths using a landform evolution model, *J. Geophys. Res.* **111,** F02016, doi 10.1029/2005JF000351.

Scripter, M. W. (1970). Nested-means map classes for statistical maps, *Ann. Assoc. Am. Geogr* **60,** 385–393.

Seed, H. B., C. Ugas, and J. Lysmer (1976). Site-dependent spectra for earthquake-resistant design, *Bull. Seismol. Soc. Am.* **66,** no. 1, 221–243.

Singh, S. K., and M. Ordaz (1993). On the origin of long coda observed in the lake-bed strong-motion records of Mexico City, *Bull. Seismol. Soc. Am.* **83,** no. 3, 1298–1306.

Spudich, P., M. Hellweg, and W. H. K. Lee (1996). Directional topographic site response at Tarzana observed in aftershocks of the 1994 Northridge, California, earthquake: Implications for mainshock motions, *Bull. Seismol. Soc. Am.* **86,** no. 1B, S193–S208.

Thompson, E. M., L. G. Baise, R. E. Kayen, Y. Tanaka, and H. Tanaka (2010). A geostatistical approach to mapping site response spectral amplifications, *Eng. Geol.* **114,** 330–342.

Tinsley, J. C., and T. E. Fumal (1985). Mapping Quaternary sedimentary deposits for areal variations in shaking response, in *Evaluating Earthquake Hazards in the Los Angeles Region—An Earth Science Perspective* J. I. Ziony (Editor), *U.S. Geol. Surv. Profess. Pap. 1360*, 101–126.

Tobler, W. R. (1988). Resolution, resampling and all that, *Building Databases for Global Science*, H. Mounsey, R. Tomlinson (Editors), **12,** Taylor & Francis, London, 9–137.

Vibration Instruments Company (VIC) (1993). Shear wave velocity profiling of instrumented sites in the greater Los Angeles area using surface wave measurements, in *Report to the University of Southern California, Department of Civil Engineering*, Vibration Instruments Company, Tokyo, Japan.

Wakamatsu, K., M. Matsuoka, K. Hasegawa, S. Kudo, and M. Sugiura (2004). GIS-based engineering geomorphologic map for nationwide hazard assessment, *11th Intl. Conf. Soil Dynamics & Earthquake Eng. and 3rd Intl. Conf. Earthquake Geotech. Eng. Proceedings*, University of California, Berkeley, California, 7–9 January 2004, 879–886.

Wald, D. J., and T. I. Allen (2007). Topographic slope as a proxy for seismic site conditions and amplification, *Bull. Seismol. Soc. Am.* **97,** no. 5, 1379–1395.

Warnes, G. R. (2004). *Plot x-y points with locally smoothed mean and standard deviation*, http://rss.acs.unt.edu/Rdoc/library/gplots/html/bandplot.html (last accessed April 2009).

Wills, C. J. (1998). Differences in shear-wave velocity due to measurement methods: A cautionary note, *Seismol. Res. Lett.* **69,** no. 3, 216–221.

Wills, C. J., and K. B. Clahan (2006). Developing a map of geological defined site-condition categories for California, *Bull. Seismol. Soc. Am.* **96**, no. 4A, 1483–1501.

Wills, C. J., and W. Silva (1998). Shear-wave velocity characteristics of geologic units in California, *Earthquake Spectra* **14,** no. 3 533–556.

Wills, C. J., M. Petersen, W. A. Bryant, M. Reichle, G. J. Saucedo, S. Tan, G. Taylor, and J. Treiman (2000). A site-conditions map for California based on geology and shear-wave velocity, *Bull. Seismol. Soc. Am.* **90,** no. 6B, S187–S208.

Yamazaki, F., K. Wakamatsu, J. Onishi, and K. T. Shabestari (2000). Relationship between geomorphological land classification and site amplification ratio based on JMA strong motion records, *Soil Dynam. Earthquake Eng.* **19,** 41–53.

Yong, A., S. E. Hough, M. J. Abrams, H. M. Cox, C. J. Wills, and G. W. Simila (2008a). Site characterization using integrated imaging analysis methods on satellite data of the Islamabad, Pakistan, region, *Bull. Seismol. Soc. Am.* **98,** no. 6 2679–2693.

Yong, A., S. E. Hough, M. J. Abrams, and C. J. Wills (2008b). Preliminary results for a semi-automated quantification of site effects using geomorphometry and ASTER satellite data for Mozambique, Pakistan, and Turkey, *J. Earth Syst. Sci.* **117,** no. S2 797–808.

Yong, A., S. E. Hough, A. Braverman, and J. Iwahashi (2010). A terrain-based V_{S30} estimation map of the contiguous United States, *Seismol. Res. Lett.* **81,** no. 2, 294, Abstract.

Yong, A., S. Hough, H. Cox, K. Tiampo, A. Braverman, J. Harvey, S. Hook, K. Hudnut, and G. Simila (2006). Balloons to satellites: A century of progress in geotechnical site characterization, *Seismol. Res. Lett.* **77,** no. 2, 302, Abstract.

Yong, A., S. Hough, K. Hudnut, S. Healy, J. Giberson, and G. Simila (2005). Geotechnical site characterization using remote sensing data: A feasibility study, *Seismol. Res. Lett.* **76,** no. 2, 241, Abstract.

Yong, A., S. E. Hough, C. Wills, and M. Abrams (2007). Site characterization using satellite imagery: A progress report, *Seismol. Res. Lett.* **78,** no. 2, 272, Abstract.

U.S. Geological Survey
525 South Wilson Ave.
Pasadena, California 91106, USA
yong@usgs.gov
(A.Y., S.E.H.)

Geospatial Information Authority of Japan
Tsukuba, Ibaraki, Japan
(J.I.)

Jet Propulsion Laboratory/California Institute of Technology
Pasadena, California, USA
(A.B.)

Manuscript received 25 September 2010