FULL PAPER

Regional terrain-based V_{S30} prediction models for China

Yuting Zhang[1], Yefei Ren[1*], Ruizhi Wen[1], Hongwei Wang[1] and Kun Ji[2]

Abstract

Time-averaged shear-wave velocity to 30 m (V_{S30}) is commonly used in ground motion models as a parameter for evaluating site effects. This study used a collection of boreholes in Beijing, Tianjin, Guangxi, Guangdong, and three other municipalities and provinces, which were divided into three regions with reference to the seismic ground motion parameter zonation map of China, to establish V_{S30} prediction models based on terrain categories. Regional effects were verified by comparing morphometric parameter (topographic slope, surface texture, and local convexity) thresholds and terrain classification maps obtained from global digital elevation model (DEM) data and regional DEM data of the three regions. Additionally, V_{S30} prediction models for the three regions using both types of terrain classification maps were established and analyzed comparatively to provide credible regional V_{S30} models for China. Through analysis of the correlations between the measured V_{S30} values and the predicted V_{S30} values, calculation of the mean squared error and mean absolute percentage error in each region, and with consideration of the geological characteristics of the boreholes, the V_{S30} prediction models based on terrain classification maps from regional data were finally applied in developing regional V_{S30} models for China. Intercomparison of the V_{S30} prediction models for the three regions indicated that subregional consideration is necessary in terrain classification. Finally, a spatial analysis method adopting inverse distance weighting of the residuals was used to update the initial V_{S30} models. The developed V_{S30} models could be used both in developing regional ground motion models and in the construction of earthquake disaster scenarios.

Keywords Terrain classification, Morphometric parameters, V_{S30} prediction models, Regional features
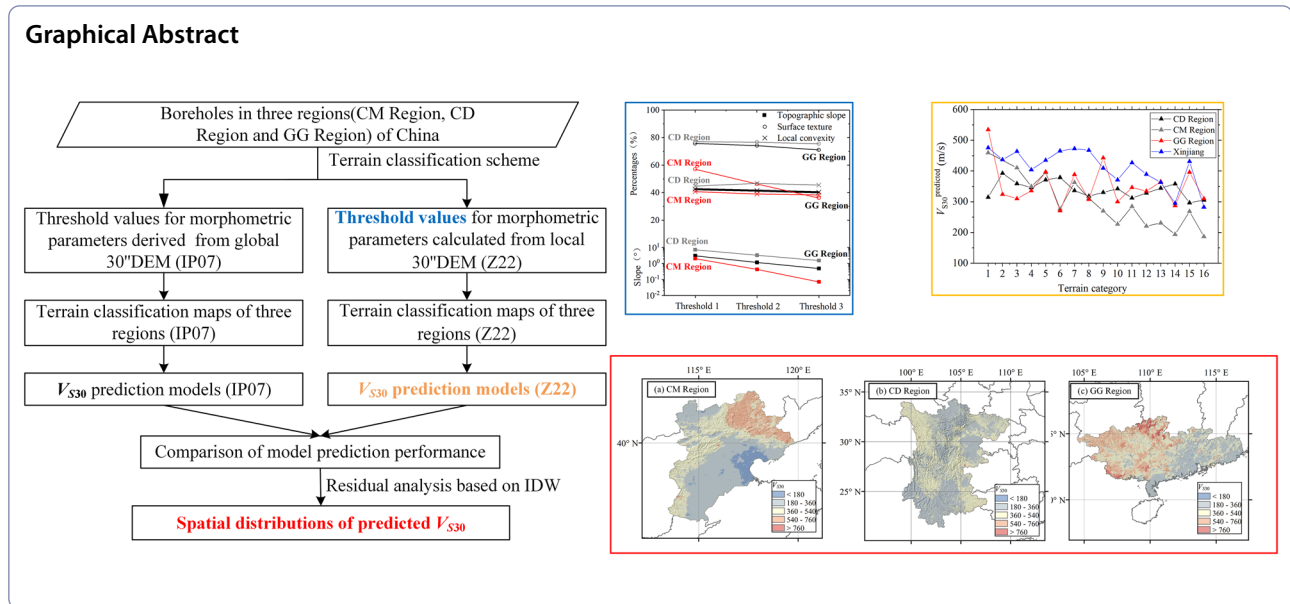
*Correspondence:
Yefei Ren
renyefei@iem.net.cn
Full list of author information is available at the end of the article

Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 2 of 19

Graphical Abstract

Introduction

It is recognized that damage to buildings is severely affected by strong ground motion, which is substantially influenced by local site conditions. The time-averaged shear-wave velocity to 30 m (V_{S30}) is used widely as an index of site classification in earthquake resistant design code and it is considered as an important parameter for measuring the amplification effect of ground motion. For example, in ground motion models (e.g., Abrahamson and Silva 2008; Boore and Atkinson 2008; Abrahamson et al. 2013; Boore et al. 2013) and ShakeMap (Wald et al. 2022), V_{S30} values were used to quantify the site amplification effects of ground motion. It is common to measure V_S profiles using in situ geophysical methods for critical projects (Ahdi et al. 2018). Although V_{S30} values can be calculated from site profile data obtained using invasive techniques such as downhole and crosshole testing and P–S suspension logging, and using non-invasive techniques such as surface wave inversion (Stokoe et al. 1994; Park et al. 1999) and P- and/or S-wave seismic reflection/refraction (Telford et al. 1990), obtaining V_{S30} values across an entire region remains a challenge. Therefore, to achieve regional site classification, available high-precision topographic, geologic, and geomorphic data are used.

Wald and Allen (2007) established a relationship between V_{S30} and topographic slope for active and stable tectonic regions based on 30arc-second digital elevation model (DEM) data. On this basis, much subsequent research has involved more detailed study in different countries or regions (Lee et al. 2001; Allen and Wald 2009; Lemoine et al. 2012; Xie et al. 2016; Zhou et al. 2022; Zhang et al. 2022). Age, lithology, genesis, and other properties of geological maps have also been used as indicators for estimating V_{S30} (e.g., Borcherdt 1994; Park and Elrick 1998; Wills 2000; Wills and Clahan 2006; Vilanova et al. 2018; Li et al. 2022). Furthermore, some studies proposed comprehensive consideration of topographic and geologic information to improve the accuracy of V_{S30} predictions (Thompson and Wald 2012; Thompson et al. 2014; Wills et al. 2015; Kwok et al. 2018; Li et al. 2019). However, owing to inadequate recognition of special areas such as volcanic plateaus, carbonate rocks, and continental glaciated terrain when using the slope-based method, and the subjective experience of the classification of the geology-based method, the terrain-based method (e.g., Yong et al. 2012; Yong 2016; Ahdi et al. 2017a, b; Stewart et al. 2014) has gradually become the focus of active research in recent years.

The terrain-based unsupervised classification method was proposed by Iwahashi and Pike (2007). Based on decision tree theory, the method was used to calculate three taxonomic criteria (topographic slope, surface texture, and local convexity) using DEM data and to partition the land into 16 undefined categories. The relationship between the undefined categories and the geology/geomorphology was verified, indicating the effectiveness of this method in terrain recognition. On this basis, Yong et al. (2012) used 853 V_{S30} data (including measured and inferred data) from California to establish a V_{S30} prediction model (the Y12 model) based on 16 terrain categories. They performed iterations of a random selection and applied the cross-validation technique to calculate the mean value of V_{S30}

Zhang *et al. Earth, Planets and Space* (2023) 75:72

Page 3 of 19

under each terrain category. They determined the best predicted value of V_{S30} for each terrain category when the iteration achieved the lowest value of the mean squared prediction error. The stability and accuracy of the terrain-based prediction model were verified through comparison with the slope-based method (Wald and Allen 2007) and the geology-based method (Wills 2000). Because most V_{S30} data considered in developing the Y12 model were not from measured data, only 503 V_{S30} measurements were used in developing the subsequent V_{S30} prediction model (the Y16 model) in Yong (2016). Nevertheless, the predictive capabilities of the Y16 and Y12 models were certified as statistically indistinguishable. The prediction results of the Y16 and Y12 models for National Earthquake Hazards Reduction Program (NEHRP) site classification were broadly consistent (approximately 85% of data were classified as Site Class C, 15% classified as Site Class D, and none classified as Site Class A, B, or E). This terrain-based method has been applied successfully in many cases. For example, Seyhan et al. (2014) considered the terrain-based method important for estimating V_{S30} values in the construction of the NGA-West2 site dataset, and evaluated its reliability against other methods that included the slope-based method (Allen and Wald 2009), geotechnical-based method (originally presented in Chiou et al. 2008), and geology-based method (Wills and Gutierrez 2008) to provide the optimal scheme for determining V_{S30} values. Stewart et al. (2014) and Foster et al. (2019) also applied this method in Greece and New Zealand, respectively, for developing V_{S30} predictions. Ahdi et al. (2017a) established the terrain-based V_{S30} prediction model for the Pacific Northwest region of North America, however, different to Yong et al. (2012), they used the logarithmic mean ($\mu \ln V_{S30}$) and standard deviation ($\sigma \ln V_{S30}$) of V_{S30} from all data in each terrain category to represent the predicted results. Contreras et al. (2018) divided the Chile borehole database into southern and northern regions and established separate prediction models to inspect the regional effects of this terrain-based method. The result showed that V_{S30} was slower in the southern region than in the northern region for the same terrain categories with abundant data, which indicated the existence of a remarkable regional effect.

The global terrain classification map based on 30 arc-second DEM data, provided by Iwahashi and Pike (2007), suggests that changes in global surface landforms are consistent. However, Zhang et al. (2022) recalculated the threshold values of three taxonomic criteria using local 30-arcsecond DEM data from Xinjiang and Hebei provinces in China, and reported large differences in the results between the two regions. Moreover, statistical

analysis of the V_{S30} values also showed clear differences in the lithological characteristics of the two regions. Therefore, to adequately consider regional effects, it is necessary to classify terrain categories and to establish V_{S30} prediction models for each region independently.

Based on the research of Yong et al. (2012) and Zhang et al. (2022), a collection of boreholes in this study from seven municipalities and provinces in China was divided into three regions: the GG Region (comprising Guangxi and Guangdong provinces), CD Region (comprising Sichuan and Yunnan provinces), and CM Region (comprising Beijing, Tianjin, and Hebei), with reference to the regional characteristics of seismic tectonic activity given by the seismic ground motion parameter zonation map of China. Threshold values for morphometric parameters (topographic slope, surface texture, and local convexity) and terrain classification maps obtained from the global DEM data provided by Iwahashi and Pike (2007) (hereafter, abbreviated as IP07) and recalculated from the regional DEM data of the three regions in this study (hereafter, abbreviated as Z22) were compared. Both types of terrain classification map were used to establish regional V_{S30} prediction models, which were analyzed comparatively to provide credible regional V_{S30} models for China. Through analysis of the correlations between the measured V_{S30} values and the predicted V_{S30} values at the boreholes, calculation of the mean squared error (MSE) and mean absolute percentage error (MAPE) in each region, and with consideration of the geological characteristics of the boreholes, the effectiveness of the proposed models in predicting V_{S30} values was validated. Finally, a spatial analysis method adopting inverse distance weighting of the residuals was used to update the initial V_{S30} models. The developed V_{S30} models could be used in some other research works such as developing regional ground motion models and constructing earthquake disaster scenarios.

Regional geomorphic features

The categories obtained according to the terrain-based classification method are not delineated casually but are related to local geomorphic or geological features. Therefore, to support qualitative analysis of the effect of the classification, the geomorphic features of each region are summarized in the following, and detailed geomorphological maps of the three regions (Cheng 2008) are shown in Fig. 1. A wide-area map indicating the locations of the CM, CD, and GG Regions in China is also shown in Fig. 1.

In the CD Region, the geomorphology of Sichuan Province is high in the west and low in the east, dipping from the northwest toward the southeast. The region can be divided into three parts: the Sichuan Basin in the
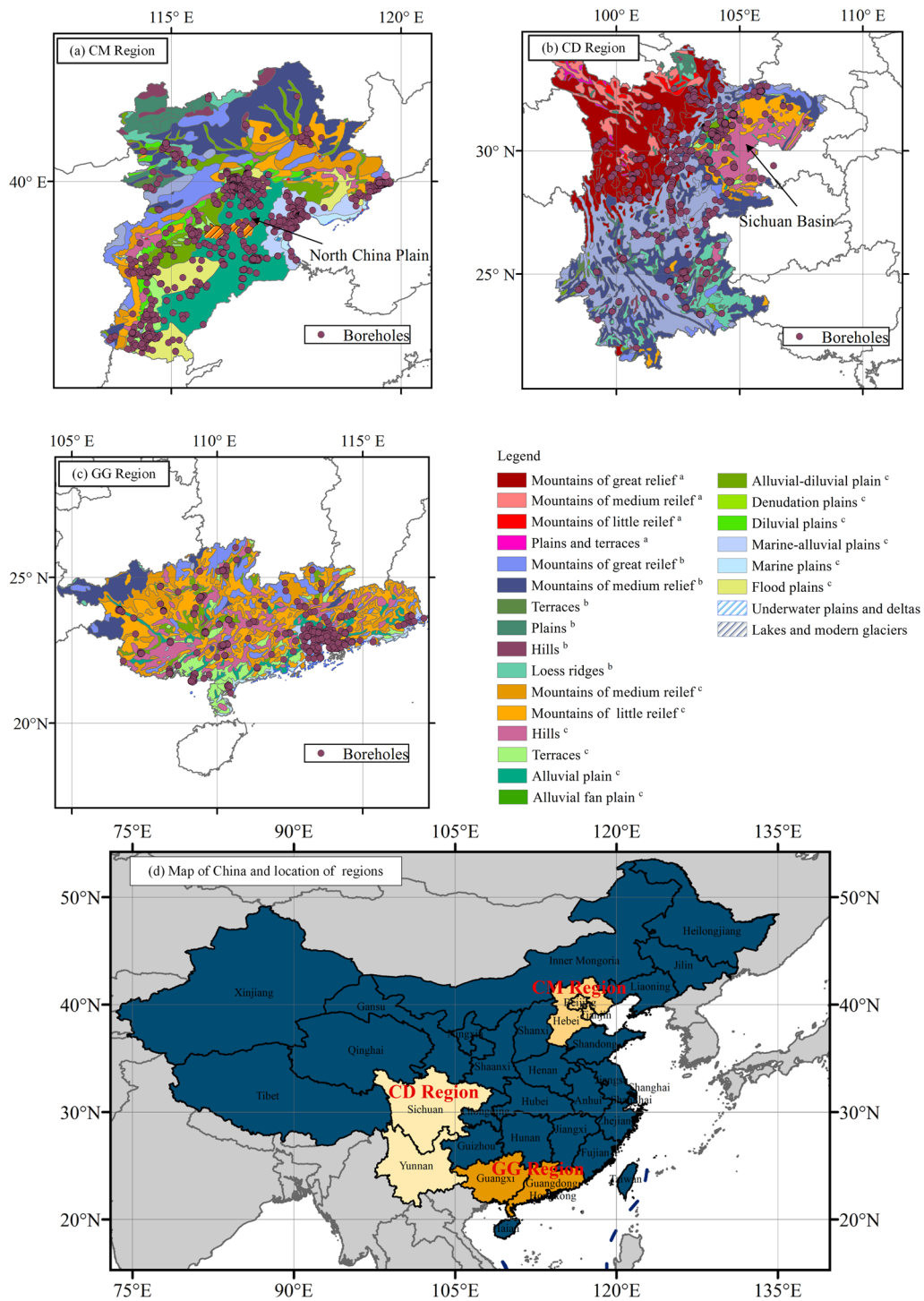
Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 4 of 19

Fig. 1 The geomorphological maps of CM Region, CD Region and GG Region: **a** for CM Region, **b** for CD Region and **c** for GG Region. Different colors in maps indicate the different geomorphological units, and the superscript a, b, c in legend mean the cases of low elevation (< 1000 m), medium elevation (1000–4000 m) and high elevation (> 4000 m), respectively. A wide-area map indicating the location of the CM Region, the CD Region, and the GG Region is shown in **d**

east, plateau mountains in the west, and mid-mountains, wide valleys and basins in the southwest (Xi 1991). The highest elevation is approximately 7213 m (in the 30 arc-second DEM data map used in this study), and the relative difference in elevation from the east to the west is almost 7000 m. The Sichuan Basin has specific

Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 5 of 19

landforms consisting of flat plains in the west, hills in the central area, and mountains and valleys in the east. The elevation of the hills is 500–1000 m. Overall, the mountains, plateaus, and hills in Sichuan Province account for approximately 92.7% of the entire land area (calculated from the data in Fig. 1). The geomorphology of Yunnan Province is high in the northwest and low in the southeast. The highest elevation is approximately 6042 m and the lowest value is near 92 m, i.e., a relative difference of 5950 m. Plains account for only 4.7% of the entire land area. The plateau area is undulating and most of the land surface is rugged with some areas of relatively gentle plateau plain.

In the GG Region, Guangxi Province is mostly surrounded by mountains and plateaus. Mountains, hills, and basins are scattered widely throughout Guangxi Province, and the mountains account for 63.9% of the entire land area. The basins and hills vary in size and interlace with each other. The plains account for only 12.1% of the total area. The geomorphology of Guangdong Province comprises mostly platforms and plains in the south and hills in the north, with valleys and basins distributed widely among the mountains. The strike of the mountains, which is generally northeast–southwest, is consistent with the underlying geological structure. Granite, sandstone, and metamorphic rocks are reasonably widespread.

The geomorphology of the CM Region is relatively high in the north and northwest, and relatively flat in the southeast and south (Ma et al. 2017). Plains are distributed widely in the area and wetlands are distributed only in the coastal region. Plateaus, mountains, and plains are arranged in order from the northwest to the southeast. The plains in the southeast can be further subdivided into alluvial plains, alluvial fan plains, denudation plains, diluvial plains, alluvial–dilluvial plains, marine plains and marine–alluvial plains. The elevation changes obviously across the entire region from approximately 1300–1700 m in plateau areas, to 500–1000 m in the mountains, to <200 m in plain areas, and to <4 m in coastal areas.

It can be seen that there are marked regional differences in geomorphic features among three regions; therefore, individual analysis is necessary to consider regional effects.

Borehole database

This study considered data obtained from 1948 boreholes in CM Region, 1597 boreholes in CD Region, and 1408 boreholes in GG Region, which were all collected from the reports of the probabilistic seismic hazard analyses of major engineering projects and buildings. The vast majority of the profile data were obtained using the

seismic downhole method, while only a few data were obtained using the P–S suspension log method at sites that did not conform to the requirements of the field operation. For these borehole sites, V_{S30} values can be calculated directly using shear wave velocity profile data for drilling depths >30 m and estimated using empirical extrapolation models for drilling depths <30 m. The models used in this study were the linear extrapolation model (Boore 2004) and the bottom constant velocity model (Kuo et al. 2011). We have found that V_{S30} values estimated using the linear extrapolation model are closer to measured V_{S30} values when the lithology changes uniformly, while V_{S30} values estimated using the bottom constant velocity model are closer to measured V_{S30} values when the lithology changes abruptly (bedrock appears within 30 m covered with soil). Therefore, the bottom constant velocity model was used in this study when bedrock was identified within 30 m, and the linear extrapolation model was used in other cases.

The median value (or the mean value of the middle of the distribution) of V_{S30} was adopted for boreholes in the same project that are close enough to be considered as having the same latitude and longitude. For example, in the same project as borehole No.
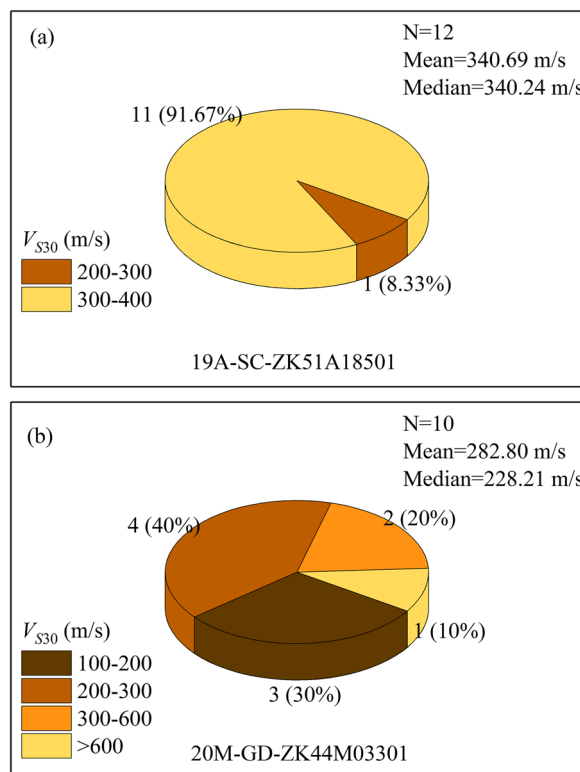

Fig. 2 Proportional pie chart of V_{S30} from boreholes with same latitude and longitude: **a** borehole No. 19A-SC-ZK51A18501, and **b** borehole No. 20M-GD-ZK44M03301

19A-SC-ZK51A18501, there are 11 other boreholes with the same latitude and longitude. The distribution of the V_{S30} values of the boreholes in this project is shown in Fig. 2a. The V_{S30} values of most of the boreholes are concentrated in the range of 300–400 m/s. In this case, the median value (340.24 m/s) of these data is close to the mean value (340.69 m/s). However, in another case (No. 20M-GD-ZK44M03301), there are nine other boreholes with the same latitude and longitude. The distribution of the V_{S30} values of the boreholes in this project is shown in Fig. 2b. The maximum and minimum values of V_{S30} of the boreholes are 607.85 and 177.21 m/s, respectively, and the mean and median values are 282.80 and 228.21 m/s, respectively. The V_{S30} values of most boreholes are concentrated in the range of 150–300 m/s (7 data), and only one borehole has a V_{S30} value that exceeds 600 m/s. Therefore, in comparison with the mean value, the median value of V_{S30} of the boreholes can avoid the influence of some extreme values. Through analysis of the borehole data, 483 V_{S30} values in the GG Region, 707 V_{S30} values in the CD Region, and 1237 V_{S30} values

in the CM Region were finally screened. Statistical histograms and cumulative percentages of the V_{S30} values in the three regions are shown in Fig. 3.

It can be seen from Fig. 3 that boreholes with $V_{S30} < 300$ m/s account for approximately 75% of all boreholes in the CM Region, while they account for only 39% and 38% of all boreholes in the GG Region and the CD Region, respectively. Generally, the characteristics of boreholes correspond to the regional lithology. A large area of alluvial plains with sand, gravel, sandy clay, and clay covers the CM Region, resulting in a relatively soft field. In contrast, hills and mountains with low–medium elevation in the GG Region, which are composed of sandstone, glutenite, slate, and granite, result in a relatively hard field. In the CD Region, the large area of mountains with marked relief and high elevation, and the environment of the Sichuan Basin, composed of a mainly pebble and gravel surface in the Chengdu Plain in the west of the basin (Wang et al. 2021), plentiful low-elevation hills in the middle of the basin, and ridges and valleys in the east of the basin (Xi 1991), also result in a hard field. Additionally, the boreholes with $V_{S30} > 700$ m/s account for
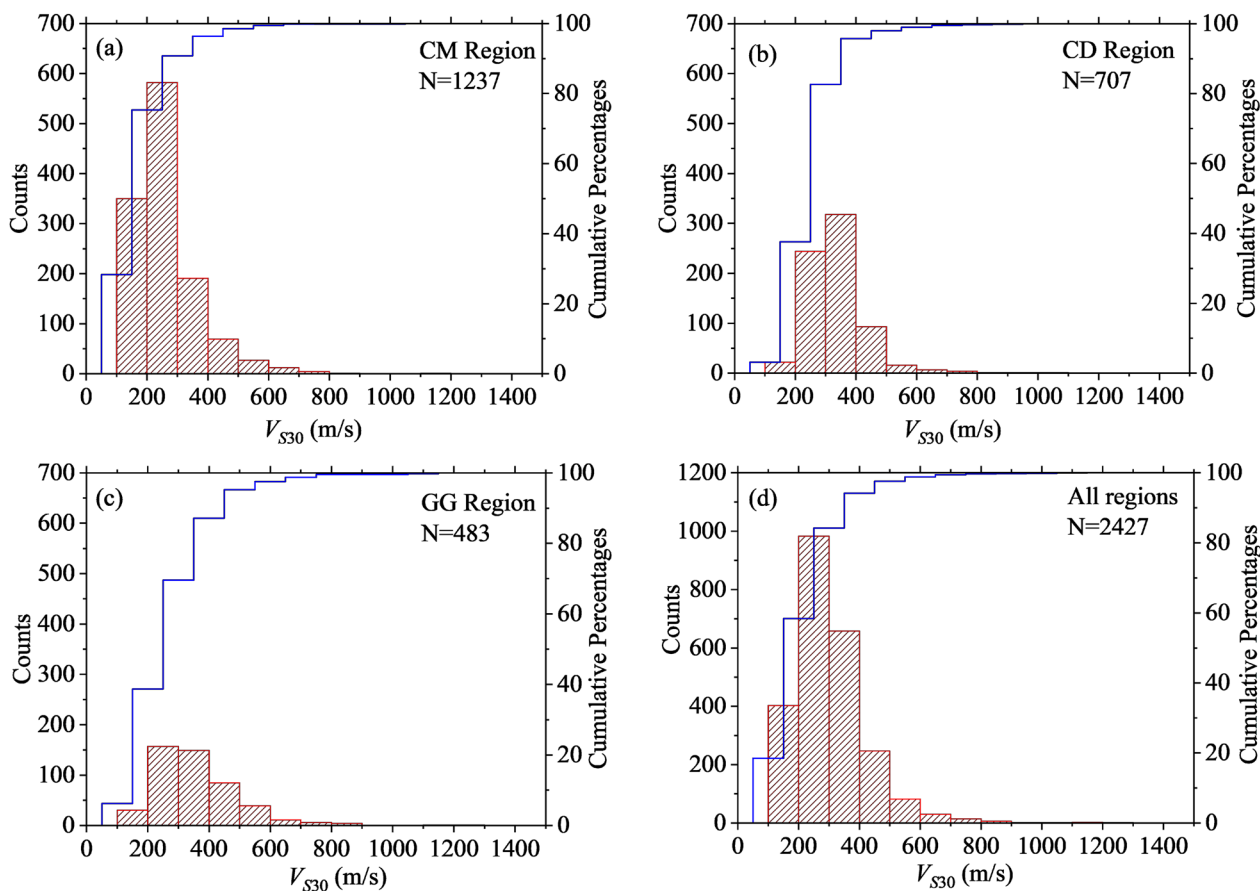

Fig. 3 Statistical histograms and cumulative percentages of V_{S30} in three regions (**a** CM Region, **b** CD Region, **c** GG Region, **d** all Regions)

< 1% of all boreholes. This is a consequence of the quality limitation of boreholes in engineering projects, i.e., most boreholes are drilled through a weathering layer that covers the underlying bedrock rather than being drilled directly into the bedrock.

V_{S30} prediction models

Terrain classifications

The terrain classification scheme proposed by Iwahashi and Pike (2007) using the three surface morphometric features of topographic slope, surface texture, and local convexity automatically identifies 16 terrain categories from DEM data. Mountains and basins can be identified by slope, terraces (convex-upward) and alluvial fans (convex-downward) can be identified by convexity, and alluvial fans with smooth surface and mountain slopes with rough surface can be distinguished by texture (Iwahashi et al. 2001).

Using V_{S30} values from California, Yong et al. (2012) proposed a method for estimation of V_{S30} based on the IP07 terrain classification map. Subsequently, Zhang et al. (2022) found significant differences between the threshold values used for terrain classification in Xinjiang and Hebei provinces of China, and highlighted that there would be a smoothing effect and inadequate recognition of local geomorphology when consistently using the global thresholds provided by Iwahashi and Pike (2007). Therefore, the terrain classification threshold values of the three regions were calculated using regional DEM

data, and the terrain categories of the three regions were reclassified accordingly.

Using 30arc-second DEM data (http://www.webgis.com/srtm30.html), and in accordance with the terrain classification scheme of Iwahashi and Pike (2007), the taxonomic thresholds of the three regions were calculated, as shown in Fig. 4a. The threshold values for Xinjiang and Hebei provinces provided by Zhang et al. (2022) and for the world provided by Iwahashi and Pike (2007) are compared in Fig. 4b.

Comparison of Fig. 4a, b reveals that the threshold values of topographic slope, surface texture, and local convexity in the CM Region are approximately the same as those reported for the Hebei province in Zhang et al. (2022) after the addition of the Beijing and Tianjin regions. The topographic slope of the CM Region is the lowest among the four local regions (i.e., the CM Region, CD Region, GG Region, and Xinjiang province) owing to the large area of plain landform in the CM Region. The texture values of both the CD Region and the GG Region are close to 75%, indicating similarity in surface roughness. However, the slope and convexity values are higher in the CD Region than in the GG Region, reflecting the surface characteristic of more mountains in the CD Region and more hills in the GG Region. Therefore, the CD Region has more upward-convex landforms and steeper topographic slopes than the GG Region. Overall, the geomorphological features of steepest topographic slope, roughest surface texture,
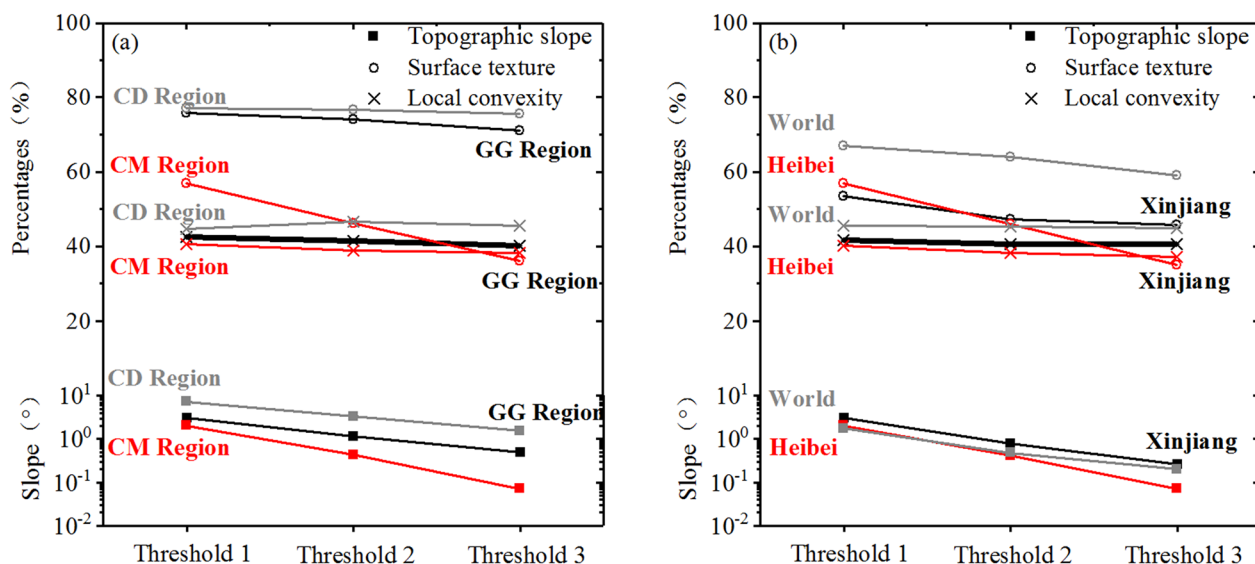

Fig. 4 Threshold values of the three morphometric parameters: **a** from the CM Region, GG Region and CD Region, and **b** from Xinjiang and Hebei provinces (Zhang et al.2022), and world (Iwahashi and Pike 2007)
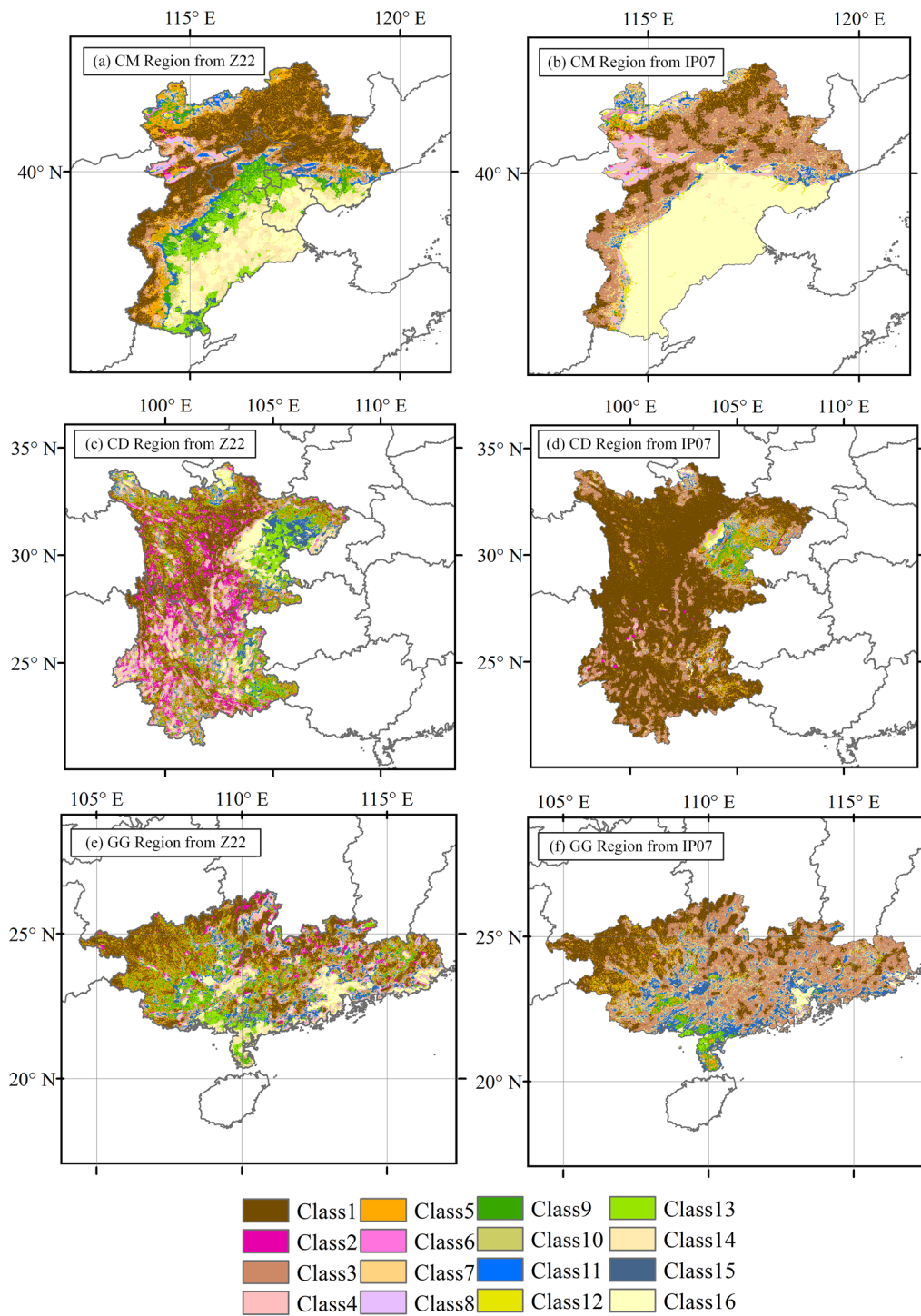
Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 8 of 19

Fig. 5 Terrain classification maps in three areas. **a**, **c** and **e** are obtained from Z22, and **b**, **d** and **f** are extracted from IP07

and most upward convexity all occur in the CD Region, reflecting the most complex changes of terrain.

The 16 terrain categories of each region were partitioned according to the classification scheme of Iwahashi and Pike (2007) and the calculated thresholds shown in Fig. 4a, and the results (Z22) are shown in Fig. 5. The results extracted from IP07 are also shown in Fig. 5 for comparison.

Zhang *et al. Earth, Planets and Space* (2023) 75:72

Page 9 of 19

For the CM Region, the classification of terrain categories can be more fully identified using the regional taxonomic thresholds. For example, in the North China Plain, the categories according to IP07 can be partitioned only as Class16, while the categories according to Z22 can be partitioned as Class11, Class13, Class14, and Class16, which can better distinguish the uneven and rough surfaces of areas with gentle low slopes that correspond to alluvial plains, alluvial–diluvial plains, and diluvial plains (Fig. 1). The grain size of the alluvial plains is coarser than that of the diluvial plains, which results in them having different characteristics of V_{S30} (Wills and Clahan 2006). Therefore, it is necessary to further partition the plain areas.

For the CD Region, IP07 unified most of western Sichuan and Yunnan into Class1, whereas the elevation of western Sichuan is actually higher than that of western Yunnan. Matsuoka et al. (2005) found that the V_{S30} value of the same geomorphic unit is affected by elevation. Therefore, further division and more detailed analysis of these areas are necessary. For the GG Region, owing to the complex terrain and varied geomorphic landforms, the mountains, hills, and plains are interlaced and the terrain categories obtained from Z22 and IP07 are relatively fragmented. Therefore, intuitive comparison of the two results is not possible and detailed analysis in combination with V_{S30} data should be conducted later.

Development of V_{S30} prediction models

Iterations of random selection and cross-validation methods were used by Yong et al. (2012) and Yong (2016) to establish the Y12 and Y16 V_{S30} prediction models, respectively, while Ahdi et al. (2017a) and Contreras et al. (2018) considered $\mu \ln V_{S30}$ and $\sigma \ln V_{S30}$ from all boreholes in each terrain category as representative. To determine the differences in the V_{S30} characteristics between the IP07 and Z22 terrain classification results, this study adopted the $\mu \ln V_{S30}$ and $\sigma \ln V_{S30}$ from all boreholes in each terrain category as representative as used by Ahdi et al. (2017a). The number of boreholes and the results of the predicted V_{S30} values for each terrain category from IP07 and Z22 are listed in Table 1, and the predicted V_{S30} values corresponding to the NEHRP site classification boundary are shown in Fig. 6.

It can be determined from Table 1 that from the IP07 results for the CM Region, no boreholes are listed under Class2 and Class6 and only 3 boreholes are listed under Class1, while the 706 boreholes listed under Class16 account for approximately 57% of the overall data. Additionally, only 4 boreholes are listed under Class9 and Class13. The Z22 results indicate that no boreholes are listed under Class2 (similar to the IP07 results) and that only 1 and 2 boreholes are listed under Class6 and

Class14, respectively, while over 200 boreholes are listed under Class9, Class13, and Class16.

In the CD Region, the IP07 results show that the boreholes under Class6 are empty, and that fewer than 3 boreholes are listed under Class2 and Class14. The terrain category with the largest number of boreholes (183 V_{S30} data) is Class1, and that with the largest standard deviation (0.34) is Class5 with 56 boreholes. The Z22 results show that each of the terrain categories has more than 5 boreholes, and that Class16 has the highest number (359 V_{S30} data). These boreholes are mainly located in the plain area of the Sichuan Basin, with similar components comprising sands and pebbles. The terrain category with the largest standard deviation (0.38) is Class13 with a total of 21 boreholes.

In the GG Region, the IP07 results indicate that no boreholes are listed under Class6 and Class10, only 1 borehole is listed under Class2 and Class4, and just 2 boreholes are listed under Class14. The exceptionally small numbers of boreholes under these terrain categories reflect both the uneven distribution of boreholes and the smoothing effect in IP07 with the application of global DEM data resulting in the absence of some terrain categories in local regions. The terrain category with the largest number of boreholes (122 V_{S30} data) is Class15. The Z22 results indicate that boreholes are listed under each of the terrain categories, but only 2 boreholes are listed under Class2. The terrain category with the largest number of boreholes (176 V_{S30} data) is Class16.

It can be seen from Fig. 6b that in the CD Region, the predicted V_{S30} values derived from IP07 under all categories are located in Site Class D of the NEHRP classification. The V_{S30} fluctuates between 280 and 360 m/s, with only a small range of variation. The predicted V_{S30} values derived from Z22 under the Class2, Class5, and Class6 categories are located in Site Class C of the NEHRP classification, and the other categories are in Site Class D, which is considered relatively reasonable in terms of site identification (the boreholes within each of NEHRP Site Class C and Site Class D account for 27.3% and 70.7% of all boreholes in the CD Region, respectively). With the exception of Class1 and Class16, the predicted V_{S30} values of each terrain category derived from Z22 are higher than or close to the predicted values derived from IP07. In the CM Region, only the predicted V_{S30} value under Class5 from Z22 is significantly higher than that from IP07, but the trends of the predicted V_{S30} values with the variation of terrain categories derived from the two terrain classification results are almost parallel. The predicted V_{S30} values under all terrain categories are located in Site Class D or Site Class C of the NEHRP site classification, and the predicted V_{S30} value under Class16 derived from Z22 is only 186.80 m/s, approaching that of Site Class E of the NEHRP site classification. In the GG

Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 10 of 19

Table 1 Predicted values and standard deviations of V_{S30} in logarithmic coordinates derived from I07 and Z22

	CD region (I07)			CD region (Z22)			CM region (I07)			CM region (Z22)			GG region (I07)			GG region (Z22)		
	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num	$\mu \ln V_{S30}$	$\sigma \ln V_{S30}$	Num
Class1	348.25	0.25	183	314.49	0.18	39	507.65	0.38	3	459.44	0.39	15	478.16	0.61	5	534.48	0.59	8
Class2	321.42	0.00	1	392.81	0.24	51	NaN	NaN	0	NaN	NaN	0	364.95	0.00	1	324.30	0.07	2
Class3	331.12	0.22	46	359.17	0.19	6	454.49	0.36	30	410.37	0.41	10	340.27	0.27	34	309.98	0.40	7
Class4	350.12	0.30	22	346.30	0.27	12	338.95	0.27	11	349.70	0.34	2	310.61	0.00	1	335.73	0.26	17
Class5	330.77	0.34	56	371.42	0.28	30	319.97	0.27	7	394.94	0.35	51	399.01	0.37	18	397.30	0.45	11
Class6	NaN	NaN	0	379.06	0.29	11	NaN	NaN	0	276.92	0.00	1	NaN	NaN	0	270.70	0.00	1
Class7	286.61	0.28	105	336.42	0.30	7	412.04	0.28	63	362.75	0.27	72	366.13	0.31	107	388.45	0.23	23
Class8	296.13	0.21	30	318.20	0.26	26	297.61	0.20	36	311.51	0.22	19	273.73	0.27	6	307.98	0.27	16
Class9	316.33	0.20	18	330.57	0.22	27	341.30	0.31	4	270.16	0.21	281	369.33	0.29	9	442.79	0.23	18
Class10	340.37	0.09	7	342.22	0.37	9	291.20	0.14	18	226.97	0.21	17	NaN	NaN	0	300.02	0.26	11
Class11	308.00	0.26	68	312.27	0.32	8	383.14	0.23	58	284.71	0.32	144	340.54	0.33	117	346.46	0.27	24
Class12	311.37	0.17	33	328.22	0.25	36	283.36	0.21	43	220.66	0.29	58	256.99	0.50	5	334.64	0.29	23
Class13	319.32	0.04	4	344.20	0.38	21	318.83	0.28	4	231.05	0.19	210	298.79	0.13	19	365.09	0.28	34
Class14	353.57	0.07	2	358.26	0.24	21	265.14	0.17	180	193.92	0.12	110	208.88	0.24	2	287.16	0.32	60
Class15	305.70	0.18	61	296.63	0.24	44	330.98	0.23	74	268.35	0.25	47	342.29	0.38	122	396.01	0.32	52
Class16	334.03	0.19	71	305.58	0.23	359	208.54	0.20	706	186.80	0.17	200	226.53	0.31	37	310.10	0.38	176

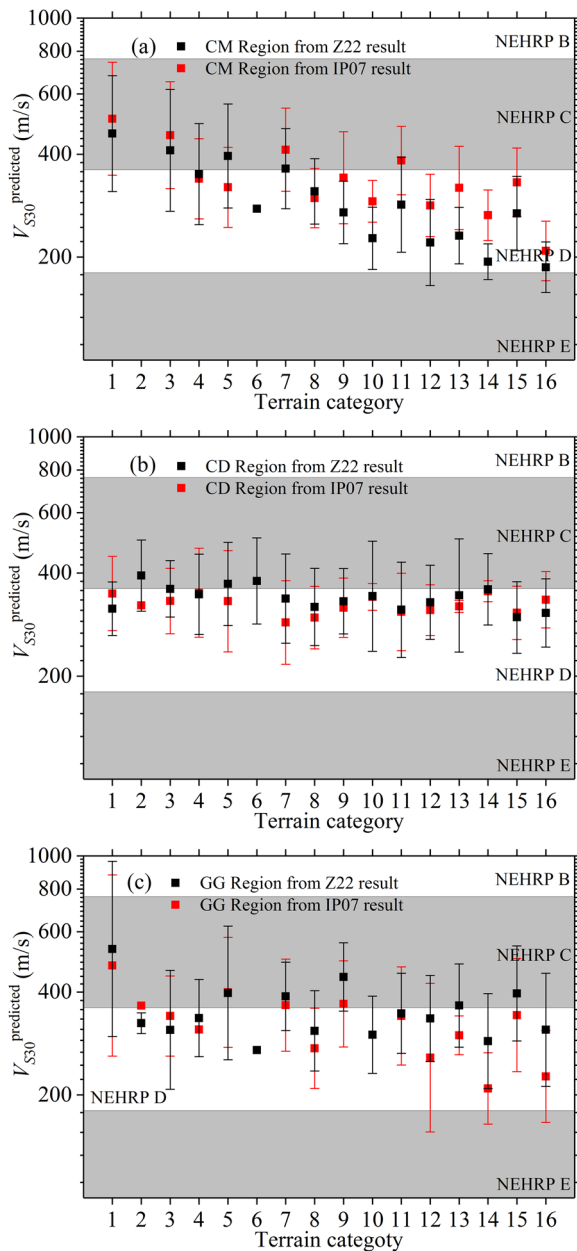"NAN" means there is no data under this terrain category

Fig. 6 Comparison of V_{S30} prediction models corresponding to NEHRP site classification. The predicted V_{S30} values derived from IP07 and Z22 classification results in CM Region, CD Region and GG Region

Region, the predicted V_{S30} values under all terrain categories derived from Z22 are higher than those from IP07, but the differences between the highest predicted values and the lowest predicted values of the two are close, i.e., 263.78 m/s from Z22 and 269.28 m/s from IP07.

To compare the validity of the predictions from Z22 and IP07, the linear relationship between the predicted V_{S30} values and the measured V_{S30} values was analyzed,

and the correlation coefficient is shown in Fig. 7. It can be seen that in the CD Region, the correlation coefficient between the predicted V_{S30} values and the measured V_{S30} values derived from Z22 is 0.32, while the equivalent correlation coefficient derived from IP07 is 0.24, i.e., significantly lower. A similar result is evident for the GG Region; however, in the CM Region, the correlation coefficient derived from IP07 (0.73) is higher than that derived from Z22 (0.65). A possible reason is that the boreholes with lower V_{S30} values are clustered in Class16 with a smaller standard deviation, resulting in a higher overall correlation.

To evaluate intuitively the performance of V_{S30} predictions between the two models in the three regions and to avoid the problem of offsetting positive and negative deviation, the MSE and MAPE values of the two models in the three regions were calculated, as shown in Table 2:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(V_{S30i}{}^{\text{measured}} - V_{S30i}{}^{\text{predicted}}\right)^2, \quad (1)$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\left(V_{S30i}{}^{\text{measured}} - V_{S30i}{}^{\text{predicted}}\right)\middle/ V_{S30i}{}^{\text{measured}}\right|, \quad (2)$$

where n is the total number of boreholes in each region, $V_{S30i}{}^{\text{predicted}}$ is the predicted V_{S30} value at the i-th borehole site, and $V_{S30i}{}^{\text{measured}}$ is the measured V_{S30} value at the i-th borehole site.

Table 2 shows that the MAPEs of both models are < 20% in the CD Region and the CM Region, whereas the MAPE is > 25% in the GG Region. The MAPEs in the three regions indicate that the performance of V_{S30} prediction is best for the CM Region, second best for the CD Region, and poorest for the GG Region, which is consistent with the number of boreholes in each region; i.e., there is a sufficient number of 1237 boreholes in the CM Region but only 483 boreholes in the GG Region. Therefore, it is anticipated that the performance of the model could be improved in future with incorporation of updated data.

It can be inferred from the MSE and MAPE values that the Z22 models show better performance in the CD Region and the GG Region, but worse performance in the CM Region in comparison with that of the IP07 model. However, it can be seen from Table 1 that 706 boreholes are classified into Class16 in the IP07 classification results. These boreholes are located on the North China Plain (in Fig. 1), which represents a complex Quaternary sedimentary environment. Detailed analysis of these boreholes is provided in the following section to
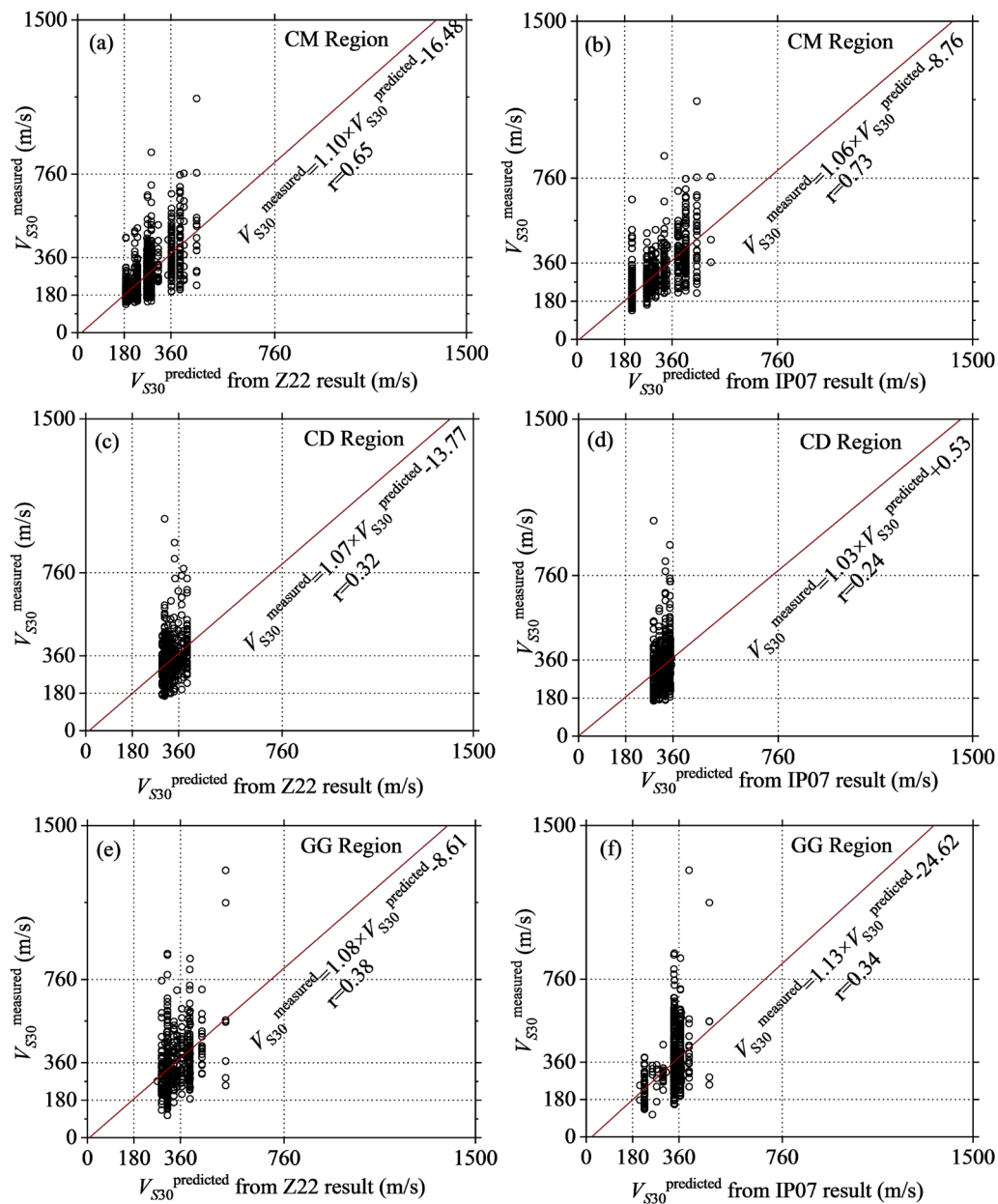
Fig. 7 The linear relationship between the predicted V_{S30} values and the measured V_{S30} values. The prediction models derived from the Z22 and IP07 classification results: **a** and **b** for the CM Region, **c** and **d** for the CD Region, and **e** and **f** for the GG Region

Table 2 MSE and MAPE values for the Z22 and IP07 models used in the three regions

Testing value	Model	CM region	CD region	GG region
MSE	IP07	4774.92	8380.63	17,397.96
	Z22	5978.65	8001.58	16,846.66
MAPE	IP07	15.7%	18.3%	26.8%
	Z22	17.2%	18.0%	26.3%

determine the rationality of classifying these boreholes into the same category under the IP07 classification results.

Cross-validation of subgroups considering geological characteristics

In the GG Region and the CD Region, the performance of the Z22 classification results has been proven better than that of IP07, but this is not the case for

Table 3 Subgroups based on geology of Class16 in IP07 in CM region

Subgroup	Number	Code	Normal distribution (0) or not (1)
Alluvial	37	G-u7	0
Lacustrine–paludal	1	G-u6	–
Lacustrine	11	G-u5	0
Alluvial–pluvial	518	G-u4	1
Marine	13	G-u3	1
Aeolian	14	G-u2	0
Alluvial-marine	100	C-u1	0
Other units	12	G-u0	0

the CM Region. It can be seen from Table 1 that the IP07 and Z22 terrain classification results might lead to concentration of boreholes under a specific terrain category. For example, the Z22 result lists 359 boreholes under Class16 in the CD Region; similarly, the IP07 result lists 706 boreholes under Class16 in the CM Region. After further consideration, we found that these 706 boreholes are located in different Quaternary sedimentary units. Considering the differences in the V_{S30} values among geological units caused by different genesis processes (e.g., sedimentation, weathering and cementation), the subgroups of these 706 boreholes based on geological condition were classified in detail, as listed in Table 3. The subgroups were re-coded, and cross-validation was performed to investigate whether there was statistical distinction among the geological units, and to determine whether it was reasonable to classify these 706 boreholes into the same terrain category. Two testing processes were performed: a parametric method (*F*-test and *t*-test) based on normally distributed data, and a nonparametric test method (*K–S* test) adopted when the data did not follow the normal distribution (Massey 1951). The normality test results for each geological unit are also shown in Table 3. The tests were performed in natural logarithmic units.

The homogeneity of variance (i.e., the *F* statistic) and the two-sample *t*-test were used to determine whether two geological units have statistically distinct standard deviations or means. Because the basic hypothesis of the *t*-test is that the two sets of data have the same variance, we first performed the homogeneity of variance test. When the results satisfied the same variance, we performed the two-sample *t*-test. The statistics *F* and *t* can be computed as follows:

$$F = \frac{s_1^2}{s_2^2}, \tag{3}$$

where s_1 and s_2 is the standard deviations of V_{S30} in geological unit 1 and 2, respectively; and

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \tag{4}$$

where $\overline{x_1}$ and $\overline{x_2}$ is the mean V_{S30} in geological unit 1 and 2, and n_1 and n_2 is the number of boreholes in geological unit 1 and 2, respectively.

The statistic of the *K–S* test is *D*:

$$D = \max_x (F(X_1) - F(X_2)), \tag{5}$$

where $F(X_1)$ and $F(X_2)$ is the cumulative distribution function of V_{S30} in geological unit 1 and 2, respectively.

By comparing the statistic to the empirical distribution, the significance level (*p*) is computed. If they have a value of $p \le 0.05$, the two geological units are considered distinct.

As can be seen in Table 3, the alluvial–pluvial unit (G-u4) and the marine unit (G-u3) failed the normality test. Therefore, we performed the *K–S* test on them, and derived a *p*-value of 1.38e−09. Thus, the results identified a statistical difference between G-u4 and G-u3.

Because only one data value was obtained from the lacustrine–paludal unit (G-u6), we ignored the statistical tests for this subgroup. Furthermore, owing to the lack of detailed sedimentary information on G-u0, this subgroup was also ignored. Thus, the hypothesis of the normal distribution was verified and the homogeneity of variance was tested only for the alluvial (G-u7), lacustrine (G-u5), aeolian (G-u2), and alluvial-marine units (G-u1). The *t*-test was also performed on those units that satisfied the same variance hypothesis to verify whether there were differences in the mean values among them; the results are listed in Table 4. The results in Table 4 show

Table 4 *F*-test results (p_1) and *t*-test results (p_2) between subgroup pairs

	p_1	Distinct standard deviations?	p_2	Distinct means?	Distinct (0) or not distinct (1)
Gu1-gu2	0.5972	1	1.97e−11	0	0
Gu1-gu5	0.0183	0	–	–	0
Gu1-gu7	2.15e−13	0	–	–	0
Gu2-gu5	0.2010	1	0.0023	0	0
Gu2-gu7	0.0023	0	–	–	0
Gu5-gu7	0.1202	1	4.18e−08	0	0

the statistical differences in V_{S30} between these geological units. Therefore, the concentration of boreholes in Class16 under IP07 classification would lead to failure in identifying different geological units with statistical distinct V_{S30} values.

For the CD Region in the Z22 results, there is also a concentration of borehole data under Class16. According to geological information, it can only be divided into a Holocene unit (253 data), Pleistocene unit (44 data), and other units (62 data). Because the data from the Holocene unit and the Pleistocene unit failed the normality test, the K–S test was performed, and the derived p-value of 0.2076 proved the lack of statistical difference between the two units. These boreholes are mainly located in the plain area of the Sichuan Basin, with similar lithology of mixed sands and pebbles, and the V_{S30} values showed no significant differences despite the different sedimentary ages.

Discussions

Comparison of V_{S30} prediction models

The V_{S30} prediction model results derived from Z22 in the CD Region, GG Region, and CM Region and the V_{S30} prediction model results of Zhang et al. (2022) for the Xinjiang province are shown in Fig. 8. It can be seen that the predicted V_{S30} value under Class1 for the CD Region (314.49 m/s) is significantly lower than that predicted for the other three regions, despite the fact that the boreholes are located in mountain areas with higher elevations and steeper slopes than in the other regions. It might be related to the characteristics of the engineering boreholes involved in the study, i.e., only a few boreholes had values of > 500 m/s, or it might reflect problems inherent in the application of the decision tree theory. Comparison of the correlation between the V_{S30} values and slopes in the four regions revealed that the correlation was least significant

in the CD Region (Fig. 9). Because the classification scheme based on the decision tree method takes slope as an important taxonomic threshold, V_{S30} cannot be distinguished by slope in the CD Region. Therefore, it is insufficient to achieve V_{S30} prediction models that rely only on terrain. Instead, geological age, lithology, and other information available in geological maps can be used as supplementary conditions to provide more reasonable estimates in conjunction with measurements from boreholes. On this basis, application of spatial interpolation and geostatistical Kriging interpolation can improve the accuracy of V_{S30} prediction and reveal the trend of V_{S30} across a region.

Figure 8 also shows that among the four regions the Xinjiang province has relatively high predicted V_{S30} values under all terrain categories, and that the CD Region has minimal variation in the predicted V_{S30} values with the various terrain categories. The difference in V_{S30} value between the GG Region and the CD Region in Class1 (219.99 m/s) is the largest among the four regions under the same terrain. The highest predicted V_{S30} values are for Class1 in the GG Region, and the lowest predictions are for Class16 in the CM Region. Significant differences in the predicted V_{S30} values between different regions under the same terrain category demonstrate the necessity for delineating terrain by region.

Spatial distributions of predicted V_{S30}

The predicted V_{S30} value at any point in the four regions can be obtained using the model established in the above, but there is substantial limitation in that all the predicted V_{S30} values in Site Class C and Site Class D of the NEHRP site classification are < 540 m/s, which is obviously impractical. Therefore, we calculated the residuals of the boreholes between the measured V_{S30} values and the predicted V_{S30} values, analyzed the spatial variation trend of the residuals, and updated the initial V_{S30} prediction results through multiplying the residual results with the V_{S30} prediction model results. The residuals were calculated using the following equation:

$$R_i = V_{S30,i}{}^{\text{measured}} \Big/ V_{S30,i}{}^{\text{predicted}}, \tag{6}$$

where R_i represents the residual at the i-th borehole (1237 data in CM Region, 707 data in CD Region, and 483 data in GG Region), $V_{S30,i}{}^{\text{predicted}}$ represents the predicted V_{S30} value at the borehole points, and $V_{S30,i}{}^{\text{measured}}$ represents the measured V_{S30} value at the borehole points.

Commonly used interpolation methods can be classified into exact methods [e.g., inverse distance weighting (IDW)] and inexact methods (e.g., Kriging and trend surface analysis) (Li and Heap 2014). The exact interpolation
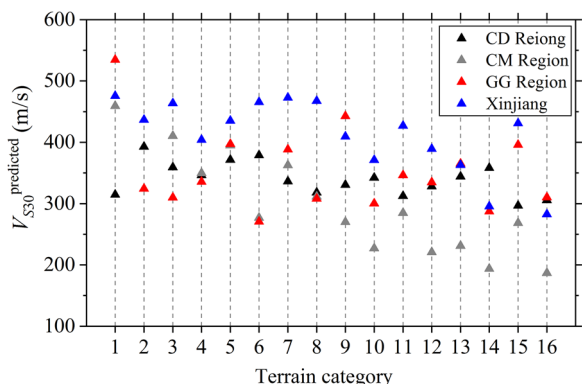

Fig. 8 Comparison of the V_{S30} prediction model results for the four regions [the Xinjiang province (Zhang et al.2022), the CD Region, the GG Region, and the CM Region]
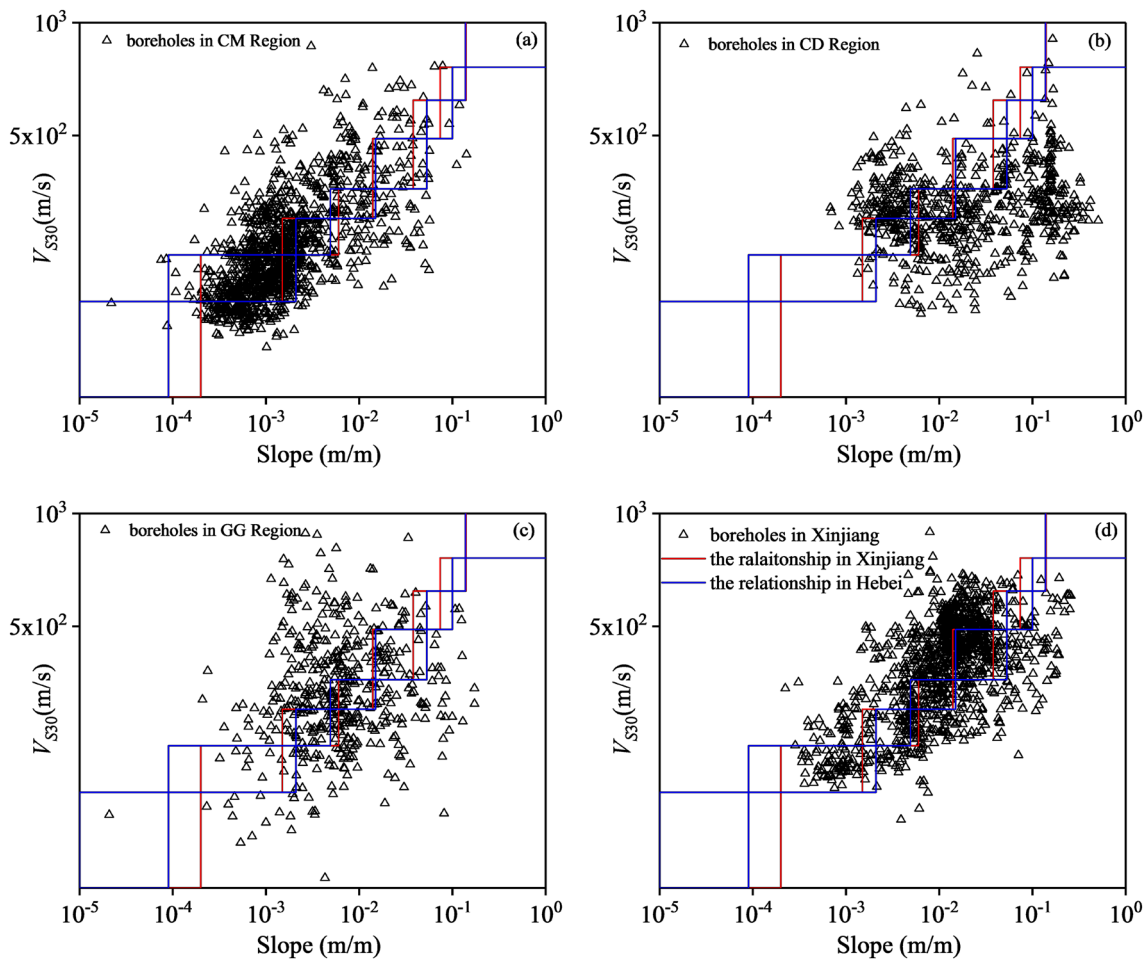
Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 15 of 19

Fig. 9 Correlations between V_{S30} and slope in the four regions: **a** the CM Region, **b** the CD Region, **c** the GG Region, and **d** the Xinjiang province. Polygonal boundary enclosed by straight lines represent V_{S30} and slope ranges provided by Zhang et al. (2022) corresponding to Xinjiang and Hebei province

method considers that the measured value is equal to the predicted value at the sampling points, whereas the inexact interpolation method considers that the measured value differs from the predicted value at the sampling points; it can avoid the effect of severe peaks or troughs in the results; i.e., it can handle outliers effectively.

We calculated the Kriging interpolation and IDW interpolation results and found that the variation trend of both methods was similar in most areas. However, in some areas, the IDW method retained the original high (or low) value data, while the Kriging method smoothed out these high (or low) values. Therefore, the IDW method was used in this study to analyze the spatial variation trend of the residuals. The formula for IDW can be expressed as follows:

$$c = \sum_{i=1}^{n} \frac{1}{d^{\alpha}} c_i, \tag{7}$$

where c is the estimated value of the residuals at an unknown point in space, c_i is the calculated value of the residuals at the i-th borehole, d is the distance from the i-th borehole, and α is the power value.

The IDW method mainly depends on the power value of the inverse distance. The power value can control the influence of the sampling points on the interpolation based on the distance from the output point. By defining a higher power value, the nearest point can be further emphasized. Therefore, as the power value increases, the adjacent data will be most affected, the surface will become more detailed (less smooth), and the interpolation will gradually approach the value of
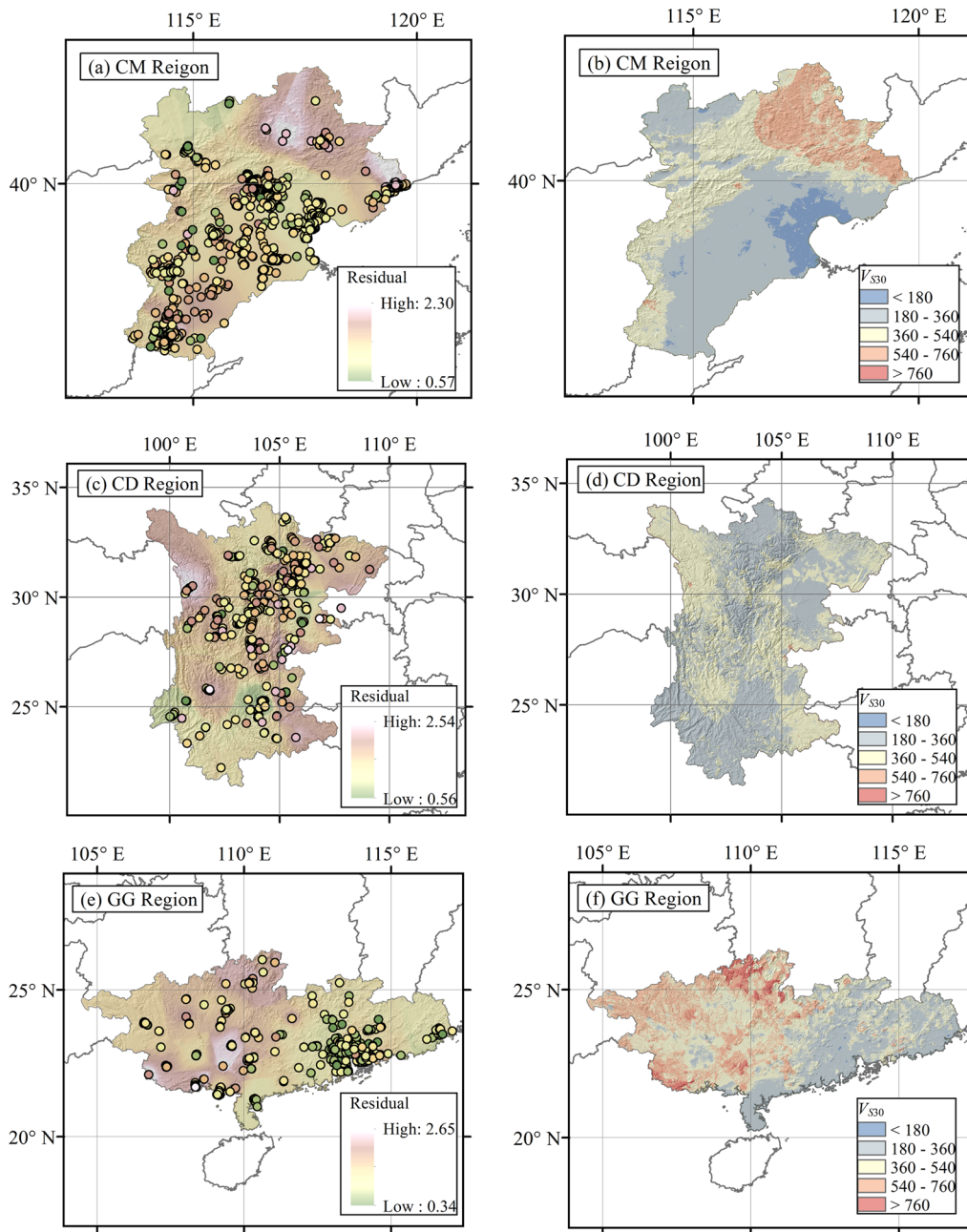
Fig. 10 Spatial variation trends of the residuals and the updated V_{S30} prediction results in three regions. Spatial variation trends of the residuals obtained using the IDW interpolation method for: **a** the CM Region, **c** the CD Region, and **e** the GG Region. Final prediction models obtained by multiplying the initial prediction models by the residuals for: **b** the CM Region, **d** the CD Region, and **f** the GG Region

the nearest sampling point. Conversely, a smaller power value will have greater impact on the surrounding points further away, resulting in a smoother surface.

The power value α was selected using cross-validation method (Li et al. 2021). Overall, 60% of data were randomly selected as the prediction group to build the interpolation model, and the remaining 40% of data were

selected as the detection group used to test the prediction capacity of the interpolation model by calculating the MAPE value. We calculated the MAPEs under the case of $\alpha = 1$, 2 and 3 for each region, and set α with the lowest MAPE value as the best power value in each region.

Zhang *et al. Earth, Planets and Space* (2023) 75:72

Page 17 of 19

Finally, the α value was set as 1 in the CM Region and the CD Region, and set as 2 in the GG Region.

The spatial variation trends of the residuals obtained by the IDW method for the CM Region, CD Region, and GG Region are shown in Fig. 10a, c, and e, respectively. It can be seen that the variation of the residuals in the GG Region is the most remarkable, with a maximum value of 2.65 and a minimum value of 0.34. The variation of the residuals in the CM Region is relatively moderate, with a maximum value of 2.30 and a minimum value of 0.57. The residuals were used to update the initial regional V_{S30} prediction results (Fig. 5, Table 1), and the final V_{S30} prediction maps for the CM Region, CD Region, and GG Region are shown in Fig. 10b, d, and f, respectively. In comparison with the initial V_{S30} prediction results, the updated results based on the residuals can identify the coastal areas of Site Class E ($V_{S30} < 180$ m/s) and the northern mountain area with $V_{S30} > 540$ m/s in the CM Region, Furthermore, the updated results can also identify the region with shear wave velocity of > 540 m/s, which represents mountains, plateaus, and platforms in the western parts of the GG Region, which is more consistent with the actual site conditions.

Conclusion

This study used terrain-based methods (Iwahashi and Pike 2007; Yong et al. 2012; Zhang et al. 2022) to establish regional V_{S30} prediction models for China. Considering regional differences in morphometric features, the boreholes considered in the study were divided into three regions (the CD Region, CM Region, and GG Region), with reference to the fifth seismic zoning map of China, and models for prediction of V_{S30} were established. To verify the regionality of this method, we recalculated the terrain classification threshold values of topographic slope, surface texture, and local convexity and reclassified the landforms into 16 categories for the three regions using regional DEM data. The terrain classification results for the three regions were compared with results extracted from the IP07 global classification map. Results proved that using regional DEM data instead of global DEM data can distinguish uneven and rough surfaces in the region of gentle low slopes in the CM Region and the region of mountains with steep slopes and high elevations in the CD Region.

The V_{S30} prediction models for the three regions derived using the IP07 and Z22 terrain classification maps were compared to evaluate the prediction performance of these two results. By comparing the correlations between the measured and predicted V_{S30} values under 16 categories and calculating the MSE and MAPE values from all boreholes in each region, we found that

the V_{S30} prediction models based on the Z22 results performed better in the CD Region and the GG Region. Although the Z22 results did not perform well in the CM Region, analysis of the borehole geological information revealed that the high concentration of boreholes under Class16 could not distinguish different sedimentary units with significant statistical differences of V_{S30}. The use of global thresholds in the CM Region did not allow for a detailed division of the plains, whereas the use of regional thresholds could resolve this problem.

Moreover, the necessity for delineating terrain by region was proven by means of the significant differences in the predicted V_{S30} values between different regions under the same terrain category. Finally, by analyzing the spatial variation trends of the residuals between the measured V_{S30} values and the predicted V_{S30} values at the boreholes, the initial regional V_{S30} prediction model was updated, and the updated model was proven to be more consistent with the actual site conditions.

Analysis also revealed that the effectiveness of the model is closely related to the statistical characteristics of the V_{S30} of collected boreholes. The insignificant correlations between V_{S30} and slope in the CD Region confirmed that it is insufficient to use V_{S30} prediction models that rely only on terrain. The geological age, lithology and other information available in geological maps can be used as additional supplementary data to provide a more reasonable estimation.

Abbreviations

V_{S30}	Time-averaged shear-wave velocity to 30 m
NEHRP	National Earthquake Hazards Reduction Program
Y12 model	V_{S30} Prediction model provided by Yong et al. (2012)
Y16 model	V_{S30} Prediction model provided by Yong (2016
DEM	Digital elevation model
IP07	Threshold values for morphometric parameters and terrain classification maps obtained from the global DEM data provided by Iwahashi and Pike (2007)
Z22	Threshold values for morphometric parameters and terrain classification maps recalculated from the regional DEM data of the three regions in this study
CM Region	Comprising Beijing, Tianjin, and Hebei
CD Region	Comprising Sichuan and Yunnan provinces
GG Region	Comprising Guangxi and Guangdong provinces
$V_{S30}^{predicted}$	The predicted V_{S30} values at the borehole points
$V_{S30}^{measured}$	The measured V_{S30} values at the borehole points
MSE	The mean squared error
MAPE	Mean absolute percentage error
IDW	The inverse distance weighting

Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 18 of 19

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, No. 29 Xuefu Road, Harbin 150080, Heilongjiang, People's Republic of China. [2]College of Civil and Transportation Engineering, Hohai University, No. 1 Xikang Road, Nanjing 213022, Jiangsu, People's Republic of China.

References

Abrahamson NA, Silva WJ, Kamai R (2013) Update of the AS08 ground-motion prediction equations based on the NGA-West2 data set. Pacific Earthquake Engineering Research Center, University of California, Berkeley, PEER Report, 2013-04

Abrahamson NA, Silva WJ (2008) Summary of the Abrahamson & Silva NGA ground-motion relations. Earthq Spectra 24(1):67–97

Ahdi SK, Stewart JP, Ancheta TD, Kwak DY, Mitra D (2017a) Development of V_S profile database and proxy-based models for V_{S30} prediction in the Pacific Northwest region of North America. Bull Seismol Soc Am 107:1781–1801

Ahdi SK, Stewart JP, Kwak DY, Ancheta TD, Mitra D (2017b) Proxy-based V_{S30} prediction in Alaska accounting for limited regional data. In: Paper presented at the 3rd Int. Conf. on performance-based design in earthquake geotechnical engineering (PBD-III), Vancouver, Canada, 16–19 July 2017

Ahdi SK, Sadiq S, Ilhan O, Bozorgnia Y, Hashash A, Kwak DY, Park CB, Yong A (2018) Development of a United States community shear wave velocity profile database. Geotechnical earthquake engineering and soil dynamics V GSP, vol 291. American Society of Civil Engineers, Reston, pp 330–339

Allen TI, Wald DJ (2009) On the use of high-resolution topographic data as a proxy for seismic site conditions (V_{S30}). Bull Seismol Soc Am 99(2A):935–943

Boore DM (2004) Estimating V_{S30} (or NEHRP site classes) from shallow velocity models (depths < 30m). Bull Seismol Soc Am 94:591–597

Boore DM, Atkinson GM (2008) Ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods between 0.01 s and 10.0 s. Earthq Spectra 24(1):99–138

Boore DM, Stewart JP, Seyhan E, Atkinson GM (2013) NGA-West2 equations for predicting response spectral accelerations for shallow crustal earthquakes. Pacific Earthquake Engineering Research Center, University of California, Berkeley, PEER report 2013-05

Borcherdt RD (1994) Estimates of site-dependent response spectra for design (methodology and justification). Earthq Spectra 10(4):617–653

Cheng WM (2008) 1:4,000,000 geomorphological map of China. National Earth System Science Data Center, National Science & Technology Infrastructure of China. http://www.geodata.cn

Chiou B, Darragh R, Gregor N, Silva W (2008) NGA project strong-motion database. Earthq Spectra 24(1):23–44

Contreras V, Ruz F, Ahdi S K, et al. (2018) V_S profile database and proxy based model for V_{S30} prediction in Chile for NGA-subduction. In: Paper presented at the 11th national conference in earthquake engineering, Earthquake Engineering Research Institute, Los Angeles, California, 25–29 June 2018

Foster KM, Bradley BA, Mcgann CR, Wotherspoon LM (2019) A V_{S30} map for New Zealand based on geologic and terrain proxy variables and field measurements. Earthq Spectra 35(4):1865–1897

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

Iwahashi J, Watanabe S, Furuya T (2001) Landform analysis of slope movements using DEM in Higashikubiki Area, Japan. Comput Geosci 27(7):851–865

Kuo CH, Wen KL, Hsieh HH, Chang TM, Lin CM, Chen CT (2011) Evaluating empirical regression equations for V_S and estimating V_{S30} in northeastern Taiwan. Soil Dyn Earthq Eng 31(3):431–439

Kwok LA, Stewart JP, Kwak DY, Sun PL (2018) Taiwan-specific model for V_{S30} prediction considering between-proxy correlation. Earthq Spectra 34(4):1973–1993

Lee CT, Cheng CT, Liao CW, Tsai YB (2001) Site classification of Taiwan free-field strong-notion stations. Bull Seismol Soc Am 91(5):1283–1297

Lemoine A, Douglas J, Cotton F (2012) Testing the applicability of correlations between topographic slope and V_{S30} for Europe. Bull Seismol Soc Am 102(6):2585–2599

Li J, Heap AD (2014) Spatial interpolation methods applied in the environmental sciences: a review. Environ Modell Softw 53:173–189

Li XJ, Jing BB, Liu C, Yin JM (2019) Site classification method based on geomorphological and geological characteristics and its application in China. Bull Seismol Soc Am 109(5):1843–1854

Li PF, Zhang XC, Yan L, Hu JF, Li D, Dan Y (2021) Comparison of interpolation algorithms for DEMs in topographically complex areas using airborne LiDAR point clouds. Trans Chin Soc Agric Eng 37(15):146–153

Li M, Rathje EM, Cox BR, Yust M (2022) A Texas-specific V_{S30} map incorporating geology and V_{S30} observations. Earthq Spectra 38(1):521–542

Ma Z, Xie HL, Lin LJ, Hu QY, Qian Y, Zhang SR, Wang GL, Li JG, Tan CX, Guo HP, Zhang FC, Zhao CR, Liu HW (2017) The environmental geological conditions of Land resources in the Beijing-Tianjin-Hebei region. Geol China 44(5):857–873

Massey FJ (1951) The Kolmogorov–Smirnov test for goodness of fit. J Am Stat Assoc 46(253):68–78

Matsuoka M, Wakamatsu K, Fujimoto K, Midorikawa S (2005) Nationwide site amplification zoning using GIS-based Japan engineering geomorphologic classification map. In: Proc. 9th Int. Conf. on Struct. safety and reliability of engineering systems and structures, pp 239–246

Park S, Elrick S (1998) Predictions of shear-wave velocities in southern California using surface geology. Bull Seismol Soc Am 88(3):677–685

Park CB, Miller RD, Xia J (1999) Multichannel analysis of surface waves. Geophysics 64(3):800–808

Seyhan E, Stewart JP, Ancheta TD, Darragh RB, Graves RW (2014) NGA-West2 site database. Earthq Spectra 30(3):1007–1024

Stewart JP, Klimis N, Savvaidis A, Theodoulidis N, Zargli E, Athanasopoulos G, Pelekis P, Mylonakis G, Margaris B (2014) Compilation of a local V_S profile database and its application for inference of V_{S30} from geologic- and terrain-based proxies. Bull Seismol Soc Am 104(6):2827–2841

Stokoe KH II, Wright SG, Bay JA, Roesset JM (1994) Characterization of geotechnical sites by SASW method. In: Woods RD, Balkema AA (eds)

Zhang *et al. Earth, Planets and Space*   (2023) 75:72

Page 19 of 19

Geophysical characterization of sites, Rotterdam, The Netherlands. International Science Publisher, New York, pp 15–25

Telford WM, Geldart LP, Sheriff RE (1990) Applied geophysics, 2nd edn. Cambridge University Press, New York, p 770

Thompson EM, Wald DJ, Worden CB (2014) A map for California with geologic and topographic constraints. Bull Seismol Soc Am 104(5):2313–2321

Thompson EM, Wald DJ (2012) Developing V_{S30} site-condition maps by combining observations with geologic and topographic constraints. In: Paper presented at 15th world conference on earthquake engineering, Lisbon, Portugal, 24–28 September 2012

Vilanova SP, Narciso J, Carvalho JP, Lopes I, Quinta-Ferreira M, Pinto CC, Moura R, Borges J, Nemser ES (2018) Developing a geologically based V_{S30} site-condition model for Portugal: methodology and assessment of the performance of proxies. Bull Seismol Soc Am 108(1):322–337

Wald DJ, Allen TI (2007) Topographic slope as a proxy for seismic site conditions and amplification. Bull Seismol Soc Am 97(5):1379–1395

Wald DJ, Worden BC, Thompson EM, Hearne M (2022) ShakeMap operations, policies, and procedures. Earthq Spectra 38(1):756–777

Wang SM, Zhang M, Lin Gang WuTF, Song TW, Pei QM, Wu XY (2021) Construction of quaternary 3D geological model and analysis of buried faults in the urban area of Chengdu. Acta Geol Sin 95(8):2601–2612

Wills CJ (2000) A site-conditions map for California based on geology and shear wave velocity. Bull Seismol Soc Am 90(6B):187–208

Wills CJ, Clahan KB (2006) Developing a map of geologically defined site condition categories for California. Bull Seismol Soc Am 96(4):1483–1501

Wills CJ, Gutierrez CI (2008) Investigation of geographic rules for improving site-conditions mapping, final technical report, USGS/NEHRP award number 07HQGR0061, California Geological Survey, Sacramento, California

Wills CJ, Gutierrez CI, Perez FG, Branum D (2015) A next generation V_{S30} map for California based on geology and topography. Bull Seismol Soc Am 105(6):3083–3309

Xi CT (1991) The features and evolution of the landforms pattern in Sichuan. J Southwest China Teach Univ (Nat Sci) 16(3):384–388

Xie JJ, Zimmaro P, Li XJ, Wen ZP, Song YS (2016) V_{S30} empirical prediction relationships based on a new soil-profile database for the Beijing Plain area, China. Bull Seismol Soc Am 106(6):2843–2854

Yong A (2016) Comparison of measured and proxy-based V_{S30} values in California. Earthq Spectra 32(1):171–192

Yong A, Hough SE, Iwahashi J, Braverman A (2012) A terrain-based site conditions map of California with implications for the contiguous United States. Bull Seismol Soc Am 102(1):114–128

Zhang YT, Ren YF, Wen RZ, Wang DR, Ji K (2022) A method of site parameter estimation based on decision tree theory considering terrain features. Chin J Geophys 65(02):698–710

Zhou J, Li XJ, Tian XM, Xu GY (2022) New framework of combining observations with topographic slope to estimate V_{S30} and its application on building a V_{S30} map for mainland China. Bull Seismol Soc Am 112(4):2049–2069

Publisher's Note