Development of rule-based approaches for soil amplification prediction

Hayri Baytan Ozmen[1] and Esra Ozer[2,*]

[1]Department of Civil Engineering, Usak University, 64200 Usak, Turkey
[2]Department of Civil Engineering, Tokat Gaziosmanpasa University, 60250 Tokat, Turkey

SUMMARY: *Earthquakes are a natural disaster which causes loss of life and property. For this reason, earthquake professionals show great effort for reliable earthquake predictions. Earthquake attempts to guess with techniques rely on different modeling, approach, and certain precursors for seismic activity. These predictions in current technology can also improve by using various machine learning techniques. Because machine learning techniques such as association rule mining facilitate the interpretation of many complex problems, particularly for large databases. Furthermore, the techniques can develop more reliable approaches by combining previously acquired information. The aim of this study is to predict the key conditions influencing soil amplification from the data of the 8400 ground motion records for 100 different soil profiles. This is achieved by applying association rule mining on 31 different parameters related to soil profiles and the ground motion records with respect to amplification levels. The rules generated for predicting soil amplification are mathematically validated. Results from the proposed rule-based predictions for soil amplification show the most effective parameters (conditions) related to ground motion and soil such as frequency content, period and intensity. Additionally, probabilities of soil amplification and damping respect to soil type and peak ground acceleration were determined. These findings may provide valuable insights for future research on soil amplification.*

KEYWORDS: *soil amplification, association rule mining, earthquake; prediction, associate rules*

1   Introduction

The evaluation of soil amplification in response to seismic ground motion is a crucial aspect of earthquake engineering that helps in predicting how seismic waves interact with different types of soil layers. Numerous studies [1, 2, 3] have explored this topic, employing various analytical and experimental methodologies to assess ground motion parameters and their amplification effects.

One significant approach to evaluating soil amplification involves the analysis of ground motion data through different soil layering configurations and their respective dynamic properties. Malekmohammadi and Pezeshk [4] conducted a comprehensive study on ground motion site amplification factors, utilizing 1-D nonlinear analysis techniques to simulate seismic responses at various depths of soil deposits in the Mississippi Embayment. This study provided

a detailed understanding of how depth influences peak ground acceleration (PGA) and spectral periods. Similarly, Urzúa et al. [5] explored the efficacy of different velocity averaging methods in estimating the fundamental period of soil profiles, emphasizing the necessity to consider both shear wave velocity and site-specific conditions when evaluating ground motion amplification.

Ground motion modeling techniques, such as the equivalent linear approach, have also gained traction for their ability to account for soil nonlinearity and site-specific characteristics during ground response analysis. Yunita et al. [6] demonstrated how these models with the equivalent linear approximation could simulate seismic ground motion amplification across horizontally stratified layers. They underscore also the importance of accurately modeling soil behavior to derive realistic amplification factors. This is further supported by findings from Chen et al. [7], who illustrated the interplay between basin geometry and local site conditions, showing how these factors significantly affect ground motion amplification, especially in sedimentary valleys. Moreover, the quantification of site amplification factors is often derived from experimental observations of seismic responses across various regions. Chen et al. [7] illustrated the utilization of experimental ground-motion models (GMMs) and basin response analyses in predicting seismic site amplifications. It was emphasized that results from models and analysis which can be serve as reliable estimations for future seismic hazard assessments and risk analysis [7]. In a related study, Nagao [8] proposed a new proxy for incorporating deep subsurface characteristics into seismic evaluations, thus enhancing the predictive capabilities of models.

Kamiyama [9] highlighted that soil amplification may be critical because of nonlinear responses in soil layers during earthquake. Ceballo et al. [10], Loviknes et al. [11] and Sun and Kim [12] investigate the influence on soil amplification and structural damage of sediment thickness, site effect and frequency content. Prabowo et al. [13] investigated the variation of soil amplification with peak ground acceleration and seismic intensity parameters for a limited region. Jin [14] focused soil amplification of multi-layer soils and discussed Fourier spectral amplifications and site response analyses.

On the other hand, Torre et al. [15] fused observed linear basin amplification factors with 1-D nonlinear analyses to predict site responses effectively during strong ground motions, showcasing an innovative method that bridges experimental and analytical approaches. Additionally, Borghei et al. [16] have stated that the application of shaking table tests allows for real-time analysis of how various soil conditions—such as saturation levels—affect seismic responses in foundational structures.

Recent advances in technology have also enhanced the methodologies used for estimating site amplification. Ikram and Qamar [17, 18] investigated rules-based earthquake prediction. The results show that the integration of association rules provided significant insights into correlations between different geological factors, while predicate logic ensured that these insights could be systematically applied to make predictions. Additionally, Diana et al. [19] investigated rules-based damage evaluations. It was determined that the time needed to provide seismic vulnerability scenarios at city scale is significantly reduced, while accuracy is reduced by $<5\%$.

This study investigates change of soil amplification with 31 different ground motion and soil parameters. In the scope of study, data of the 8400 ground motion records for 100 different soil profiles were considered. Association rule mining (ARM) is used to find links between soil amplification and ground motion parameters. The results based on rules were presented as comparatively. The most effective parameters (conditions) related to ground motion and soil such as frequency content, period and intensity were determined. On the other hand, probabilities of soil amplification and damping respect to soil type and peak ground acceleration were

calculated. Although the previous studies have separately addressed the parameters related to ground motion and soil profiles that affect soil amplification, studies evaluating the combined effects of these parameters are quite limited. To the authors' knowledge, no study has yet been determined the parameters affecting soil amplification using the ARM method and obtained highly valid conditions/probabilities. In line with these results from ARM, it was thought that researchers may develop more robust models that address the complexities of soil interactions (amplification or damping) during seismic events by synthesizing experimental evidence with analytical frameworks.

2 Data properties, curation and details of categorization

2.1 Properties of considered ground motion record and soil parameters

In the study, it is aimed to evaluate the effect of soil profile and ground motion parameters on soil amplification. For this purpose, a total of 8400 ground motion record for 100 different soil conditions were considered. The ground motion records were selected Peer Ground Motion Database [20] and Itaca [21]. Detailed information on the soil conditions and ground motion records can be found in a study by Ozmen et al. [22]. The categorization of the soil conditions was carried out for a depth of 30 m through, taking into account the provisions of Turkish Building Earthquake Code (TBEC-2018) [23] and ASCE/SEI 7–16 [24]. Soil profiles were obtained from borehole geophysical studies conducted by Beyaz [25] in the field. On the other hand, a ground motion record set of 84 base rock was created. Care was taken to ensure that this set consistent with nature of ground motion and were various in frequency content, features and fault type. Peak ground acceleration (PGA) was used as the main parameter in selection of input acceleration records. The reason is that it is intensity indicator in many seismic codes and TBEC-2018. For this reason, the ground motion set has 0.0-0.6g bandwidth for PGA. By considering 0.05g increments in the relevant bandwidth, lower-density recordings were slightly larger than higher-density recordings. Record density is generally expected to increase in soil layers as one progresses from the bedrock to the surface. However, this increase can lead to an uneven distribution of the results obtained at the surface. Therefore, 32 ground motion records were considered for the 0-0.2g range, 28 ground motion records for the 0.2-0.4g range, and 24 ground motion records for the 0.4-0.6g range, all with 0.05g increments. While the intensity diversity considered with the PGA parameter, frequency content diversity considered with the dominant period (T_{pr}) and average period (T_m) values. Although real ground motion records were used in the ground motion record set as much as possible, alternative methods (deconvolutional, scaled, and synthetic) were also used due to the lack of sufficient records for some PGA intervals. It was pay attention to without altering the frequency content-intensity connection of the ground motion records for obtaining accurate results with alternative methods. The surface form of the base rock was obtained from ProShake 2.0 [26] software. In the light of the results obtained, the values of 31 parameters for each ground motion record and soil profile were determined. These parameters are explained in Table 1 and categorized with respect to ground motion record and soil properties.

The values in the base rock and surface for parameters represent input (*In*) and output (*Out*), respectively. The ratio of these values (*Out/In*) corresponds Amplification (*Amp.*). Shear wave velocities at four different depths (10, 20, 30, 40 m) are evaluated. Two different calculating approaches developed by Rathje et al. [27] and Sawada [28] for average period (T_m) values were considered. The formulizations of these are given in Eqs. (1)-(2), separately. The S_{ai}

and T_i parameter in the equations and throughout the text represents the spectral acceleration and period value of the i.period, respectively. Additionally, three different methods (weighted, sum, proposed) for site period (T_g) and four different period values (0.2, 1.0, 2.0 and 6.0s) for spectral acceleration ($S_a(T)$) were considered. The formulizations of site period developed by Sawada [28] are given in Eq. (3)-(6), separately. The V_i and H_i parameters in the equation and throughout the text represent the shear wave velocity and thickness of the i. layer, respectively. The related spectral period values are very important in TBEC-2018. Because they indicate building groups with different story heights or different regions in the spectral acceleration curve. Table 2 shows average (mean), standard deviation (std. dev.), minimum (min), quartile (25, 50, 75%) and maximum (max) values of each parameter, separately.

Table 1: Considered parameters for ground motion and soil properties

Parameter (Unit)	Component	Parameter (Unit)	Component	Parameter (Unit)	Component
-	-	In S_a (T) (g)	In S_a (T)=0.2	T_{pr} (s)	Out T_{pr}
			In S_a (T)=1.0		In T_{pr}
PGA (m/s)	In PGA		In S_a (T)=2.0		Amp.
	Out PGA		In S_a (T)=6.0	T_{m1} (s)	Out T_{m1}
	Amplification	Out S_a (T) (g)	Out S_a (T)=0.2		In T_{m1}
V_s (m/s)	V_{s10}		Out S_a (T)=1.0		Amp.
	V_{s20}		Out S_a (T)=2.0	T_{m2} (s)	Out T_{m2}
	V_{s30}		Out S_a (T)=6.0		In T_{m2}
	V_{s50}	Amp S_a (T) (-)	Out/In S_a (T)=0.2		Amp.
-	-		Out/In S_a (T)=1.0	T_g (s)	$T_{g\ weighted}$
			Out/In S_a (T)=2.0		$T_{g\ sum}$
			Out/In S_a (T)=6.0		$T_{g\ proposed}$

$$T_{M1} = \sum_{i=1}^{N} \frac{S_{ai}T_i^2}{T_i^2}, \tag{1}$$

$$T_{M2} = \sum_{i=1}^{N} \frac{S_{ai}T_i}{T_i}, \tag{2}$$

$$T_{g,sum} = \sum_{i=1}^{N} \frac{4H_i}{V_i}, \tag{3}$$

$$T_{g,weighted} = \frac{4H}{\sum_{i=1}^{N} \frac{V_iH_i}{H}}; \qquad H = \sum_{i=1}^{N} H_i, \tag{4}$$

$$T_{g,proposed} = \frac{3\sum_{i=1}^{N} S_i t_i^3 + \sqrt{9\left(\sum_{i=1}^{N} S_i t_i^3\right)^2 - 8\left(\sum_{i=1}^{N} S_i t_i^2\right)\left(\sum_{i=1}^{N} S_i t_i^4\right)}}{4\sum_{i=1}^{N} S_i t_i^2}, \tag{5}$$

$$t_i = \sum_{k=1}^{N} \frac{4H_k}{V_k}; \qquad S_i = \frac{V_{i+1} - V_i}{V_{i+1} + V_i}. \tag{6}$$

2.2 Association rules

The relationships based on correlation among a set of items or objects in a database identify with association rules [29]. A rule consists of a left-hand side (antecedent or condition) and

Table 2: Statistical properties of parameters

Parameters	Unit	Mean	Std. Dev.	Min	25%	50%	75%	Max
PGA	(g)	0.28	0.17	0.01	0.13	0.26	0.41	0.60
Amplification		1.14	0.53	0.00	0.79	1.05	1.38	4.69
V_{s10}	(m/s)	486.04	313.09	150.00	240.00	360.00	760.00	1365.83
V_{s20}	(m/s)	528.24	314.27	150.00	270.00	444.44	750.29	1429.78
V_{s30}	(m/s)	564.94	316.33	150.00	287.11	560.00	733.62	1452.44
V_{s50}	(m/s)	713.13	319.95	182.93	424.36	747.33	922.06	1471.10
S_a (T)=0.2s	(g)	0.46	0.37	0.00	0.16	0.33	0.72	1.66
S_a (T)=1.0s	(g)	0.17	0.16	0.00	0.02	0.12	0.29	0.62
S_a (T)=2.0s	(g)	0.10	0.12	0.00	0.01	0.06	0.17	0.42
S_a (T)=6.0s	(g)	0.02	0.03	0.00	0.00	0.01	0.03	0.15
Ratio T=0.2s		1.45	0.91	0.10	0.86	1.21	1.84	9.26
Ratio T=1.0s		2.10	1.30	0.03	1.25	1.93	2.65	71.86
Ratio T=2.0s		1.61	3.42	0.01	1.09	1.30	1.74	179.77
Ratio T=6.0s		1.43	14.11	0.02	1.01	1.05	1.23	1272.67
T_{pr} In	(s)	0.23	0.20	0.01	0.10	0.19	0.34	0.95
T_{pr} Out/In		10.02	32.82	0.01	1.00	1.17	2.52	450.00
$T_{g\ (weighted)}$	(s)	0.15	0.17	0.01	0.05	0.08	0.19	1.07
$T_{g\ (sum)}$		0.52	0.32	0.01	0.33	0.48	0.69	1.55
$T_{g\ (proposed)}$	(s)	0.14	0.22	0.00	0.02	0.06	0.17	1.22
T_{m1} In	(s)	0.44	0.30	0.10	0.18	0.34	0.71	1.20
T_{m1} Out/In		1.33	0.64	0.11	0.98	1.20	1.51	15.19
T_{m2} In	(s)	0.71	0.38	0.14	0.45	0.58	1.00	1.66
T_{m2} Out/In		1.10	0.36	0.11	0.91	1.04	1.24	13.48

a right-hand side (dependent part). Both the left-hand and right-hand sides consist of Boolean (true or false) conditions [30]. A rule indicates that if the left-hand side (antecedent(s)) is true, then the right-hand side (consequent(s)) is also true. A probabilistic rule indicates that, given that the left-hand side is true, the right-hand side will also be true with probability *p*. Probability *p* is the simple conditional probability that the right-hand side will also be true, given that the left-hand side is true [31]. Association rules are of the form X⇒Y. This association rule means that records in the database that satisfy rule X also satisfy the conditions of Y.

Rules are discrete in nature. Therefore, they are well-suited to modeling discrete and categorical variables. Association analysis is the search for association rules that represent frequently occurring attribute-valued conditions in a given dataset. Association analysis is widely used in market basket analysis. In market basket analysis, the goods purchased together by a consumer during their shopping process are determined. The output of market basket analysis is a set of associations related to consumer purchasing behavior. These associations are given in the form of a specific set of rules known as association rules. These association rules will help determine the appropriate product marketing strategy [32].

The mathematical model of the association rule was presented by Agrawal, Imielinski, and Swami in 1993 [33]. In this model, the set $I = i_1, i_2, .., i_m$ is called "products." *D* represents all transactions in the data set, and *T* represents each transaction of the products. *TID* is the unique identifier for each transaction.

The association rule can be defined as follows:

$$A_1, A_2, ..., A_m \Rightarrow B_1, B_2, ..., B_n.$$

In this expression, A_i and B_j are the actions or objects performed. This rule states that when actions or objects "A_1, A_2, ..., A_m" occur, often actions or objects "B_1, B_2, ..., B_n" are involved in the same event or action [34].

The association rule is generated to satisfy user-specified minimum support and confidence thresholds. The support for an item set is the percentage of transactions containing the item set among all transactions. If the association rule for item sets A and B is denoted as "$A \Rightarrow B$," then the support is defined as follows.

support $(A \Rightarrow B) = $ (number of rows containing A and B)/ (total number of rows).

The confidence value of the $A{\Rightarrow}B$ association rule is the percentage of transactions containing A that also contain B. For example, if a rule has 85% confidence, 85% of the product sets containing A also contain B. If the confidence value is 100%, the rule is true in all data analyses, and these rules are called "exact". Given the data rows associated with the task, the $(A \Rightarrow B)$ confidence is defined as follows.

confidence$(A \Rightarrow B) = $ (number of rows containing A and B)/(number of rows containing A).

Association rule mining has two stages: finding all frequent items and generating strong association rules from these frequent items. The Apriori algorithm [35], used for the first stage of association rule mining, is the most popular and classical algorithm for frequent item mining. In this algorithm, features and data are evaluated using Boolean association rules [30]. If a k-item set (item set with k elements) is denoted by c, its items (products) are denoted as $c[1], c[2], c[3], ..., c[k]$, and are ordered from smallest to largest such that $c[1] < c[2] < c[3] < ... < c[k]$ [35]. Frequently occurring itemsets are denoted by the L character, and candidate itemsets are denoted by the C character [36].

In the Apriori algorithm, the database is scanned multiple times to find frequent itemsets. First, the dataset to be mined is scanned to determine the number of transaction records in which the items are included. Then, a process begins by assuming that the items are equal to or greater than the minimum support value and are designated as the L_1 frequent 1-itemset.

The loop structure established within the code creates a new cluster in the first stage, similar to the binary combination of elements in the L_1 frequent itemset ($L_1 \infty L_1$). This operation is called join. The clusters formed by this operation are called candidate item sets and are symbolized by the letter C. Because each element of this candidate itemset consists of two elements, it is designated by the symbol C_2. This candidate set is pruned using the apriori-gen function. It examines whether subsets of elements from set C_2 are in the L_1 itemset. Elements from the subsets not included in L_1 are deleted from the C_2 candidate set. The dataset is scanned again using the Apriori algorithm to determine the number of transaction records through which the elements of candidate set C_2, which have undergone the pruning process, have passed. The elements of candidate set C_2 whose value is equal to or greater than the minimum support value constitute the L_2 frequent itemset.

In the next stage, the loop creates a new candidate itemset using a triple combination of elements from set L_2, symbolized by C_3. As in the first stage, this set undergoes pruning, and

the L_3 frequent itemset is formed with the elements remaining above the minimum support level. The loop continues, increasing the number of elements at each iteration. This process continues until no new frequent itemset is found. When the algorithm terminates the pruning process on the C_k candidate item set, the items in the L_{k-1} frequently occurring item set are considered to be the most frequently purchased items.

2.3   Categorization of parameters respect to soil amplification

The previous section stated that association rules are quite suitable for modeling discrete and categorical variables. Therefore, the discrete parameters were categorized in this section. The categorization was based on the Amplification parameter related to the PGA value, which is the main subject of the study. In general, the Amplification value can be interpreted as follows: A value of the Amplification parameter less than 1 indicates damping; a greater value indicates amplification. For a detailed categorization, the quartile values in Figure 1 of the Amplification parameter were also considered. The quartiles (Q) served as categorization boundaries (threshold values). The threshold values based on quartiles (Q) for some parameters were determined as in Table 3. The 25% and 75% quartile values for the Amplification parameter were calculated as 0.79 and 1.38, respectively. In the categorization of the relevant parameter, if the amplification value is below 0.8, it is categorized as "Severe Damping," between 0.80 and 1.00 as "Damping", between 1.00 and 1.35 as "Amplification," and above 1.35 as "Severe Amplification". The reason for using the upper threshold value of 1.35 instead of 1.38 is to ensure a more balanced distribution in the number of data falling into the relevant categories.
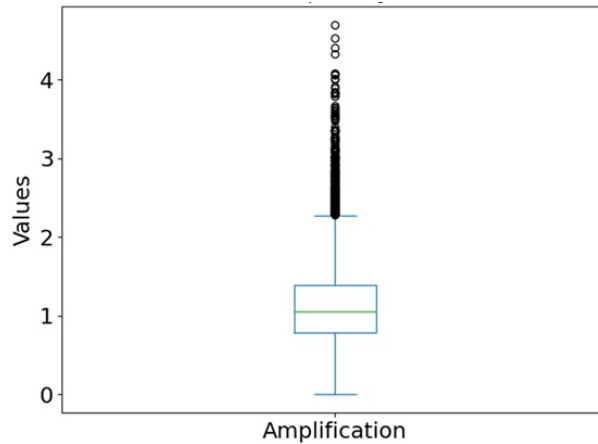

Figure 1: Boxplot of Amplification parameters for PGA

Another categorization method was subjected to the other parameters. Since the focus of the study was soil amplification, using of a decision tree (DT) compatible with Amplification categorization was preferred instead of quartile values for the categorization of the other parameters. In this way, it was thought that more valid rules regarding amplification may be obtained. The threshold values based on decision tree (DT) for some parameters were calculated as in Table 3. For the other parameters, if the middle threshold value is below the average values of that parameter, it is categorized as Very Low-Low-Moderate-High, and if it is above the average values, it is categorized as Low-Moderate-High-Very High. Different from these, the local site conditions were categorized with "Site Classes/Types (such as *ZB, ZC, ZD, ZE* and *ZF*)" respect to TBEC-2018 seismic design code. The *ZA* soil type ($V_{s30} > 1500$ m/s) was not taken into account since it corresponds approximately to the bedrock level.

Table 3: Threshold values based on quartile and decision tree for parameters

Parameters	Unit	Q/DT	Threshold values			Parameters	Unit	Q/DT	Threshold values		
			Limit 1	Limit 2	Limit 3				Limit 1	Limit 2	Limit 3
PGA	(g)	Q	0.13	0.26	0.41	S_a (T)=2.0	(g)	Q	0.01	0.06	0.17
		DT	0.07	0.16	0.44			DT	0.01	0.19	0.41
Amplification	(-)	Q	0.79	1.05	1.38	T_{pr} In	(s)	Q	0.10	0.19	0.34
		DT	0.80	1.00	1.35			DT	0.04	0.16	0.75
V_{s30}	(m/s)	Q	287.11	560.00	733.62	$T_{g\ (sum)}$	(s)	Q	0.33	0.48	0.69
		DT	278.55	465.87	1209.49			DT	0.06	0.33	0.92
S_a (T)=0.2	(g)	Q	0.16	0.33	0.72	$T_{g\ (proposed)}$	(s)	Q	0.02	0.06	0.17
		DT	0.14	0.20	0.37			DT	0.02	0.07	0.17
S_a (T)=1.0	(g)	Q	0.02	0.12	0.29	T_{m2} In	(s)	Q	0.45	0.58	1.00
		DT	0.00	0.03	0.13			DT	0.52	1.29	1.52

3 Generation of association rules and evaluations

The Apriori algorithm was applied to 31 parameters related to soil amplification, considering 8400 ground motion records for 100 different soil profiles. Based on the characteristics of the ground motion records and the soil profile, total of 17477 rules derived from association rule mining analysis. From these rules, the "Amplification" (correspond to 1.00-1.35 times amplification) or "Severe amplification" (correspond to more than 1.35 times amplification) cases, which are the focus of the study as consequents, were given in Table 4. When the table is examined, the rules obtained may be expressed as follows:

1. If V_{s30} is High (>1209.49) and soil type ZB, the amplification is 86.5% "Amplification".

2. If V_{s30} is High (>1209.49) and soil type ZB, the amplification is 89.9% "Amplification" or "Severe Amplification".

3. If PGA is below 0.068g, the probability of the "Severe Amplification" (exceeding 1.35) is 63%.

4. The amplification risk is generally higher on ZF soil types than ZB soil types. The "Severe Amplification" on ZF soil types is calculated as 56.6%. If V_{s30} is low (279-466) on ZF soil types, the risk increases from 56.6% to 73.4%. If T_{m1} is also moderate (0.894-1.048), the percentage increases to 76.9%.

5. If T_{m1} In is low (<0.145), the probability of Severe Amplification condition is 52.5%.

6. If T_{m1} In is low, the probability of Severe Amplification or Amplification condition is 83.8%.

7. If T_{m1} In is low (<0.145) and PGA (<0.068) is very low, the probability of Severe Amplification condition is 72.2%.

8. If T_{m1} In is low and PGA is very low, the probability of Severe Amplification or Amplification condition is 93.4%.

9. If $T_{g\ (proposed)}$ is moderate (0.073-0.171) and T_{m1} In is moderate (0.145-0.894), this may indicate a resonance situation, and the probability of Severe Amplification condition is 48.4%.

10. If $T_{g\ (proposed)}$ is moderate and T_{m1} In is moderate, this may indicate a resonance situation, and the probability of Severe Amplification or Amplification condition is 84.9%.

Table 4: The rules related to amplification obtained from associating rule mining analysis

Antecedents	Consequents	Antecedent support	Consequent support	Support	Confidence
V_{s30} (m/s)_High, ZB	Amplification	0.061	0.306	0.052	0.865
V_{s30} (m/s)_High, ZB	Severe Amplification	0.061	0.268	0.002	0.034
V_{s30} (m/s)_High, ZB	Severe Amplification or Amplification	0.061	0.574	0.054	0.899
PGA (g)_Very_Low	Severe Amplification	0.143	0.268	0.090	0.630
ZF	Severe Amplification	0.091	0.268	0.051	0.566
ZF, V_{s30} (m/s)_Low	Severe Amplification	0.061	0.268	0.044	0.734
ZF, V_{s30} (m/s)_Low, T_{m1} In_Moderate	Severe Amplification	0.048	0.268	0.037	0.769
T_{m1} In_Low	Severe Amplification	0.095	0.268	0.050	0.525
T_{m1} In_Low	Amplification	0.095	0.306	0.030	0.313
T_{m1} In_Low	Severe Amplification or Amplification	0.095	0.574	0.080	0.838
T_{m1} In_Low, PGA (g)_Very_Low	Severe Amplification	0.048	0.268	0.034	0.722
T_{m1} In_Low, PGA (g)_Very_Low	Severe Amplification or Amplification	0.048	0.574	0.044	0.934
$T_{g\ (proposed)}$_Moderate, T_{m1} In_Moderate	Severe Amplification	0.161	0.268	0.078	0.484
$T_{g\ (proposed)}$_Moderate, T_{m1} In_Moderate	Amplification	0.161	0.306	0.059	0.366
$T_{g\ (proposed)}$_Moderate, T_{m1} In_Moderate	Severe Amplification or Amplification	0.161	0.574	0.137	0.849

In addition, the amplified (amplification + severe amplification) and damped (damping + severe damping) probabilities respect to the PGA values as the amplitude indicator of ground motion and soil type, are calculated as in Figure 2 and Table 5, separately. It may be said that amplified probabilities are generally greater than damping probabilities. As the PGA value increases, the amplification ratio decreases in each soil type, while the damping ratio increases. However, the situation differed for ZF, and no specific trend was observed. This supports site-specific design requirements for the ZF soil class, as stated in TBEC-2018.
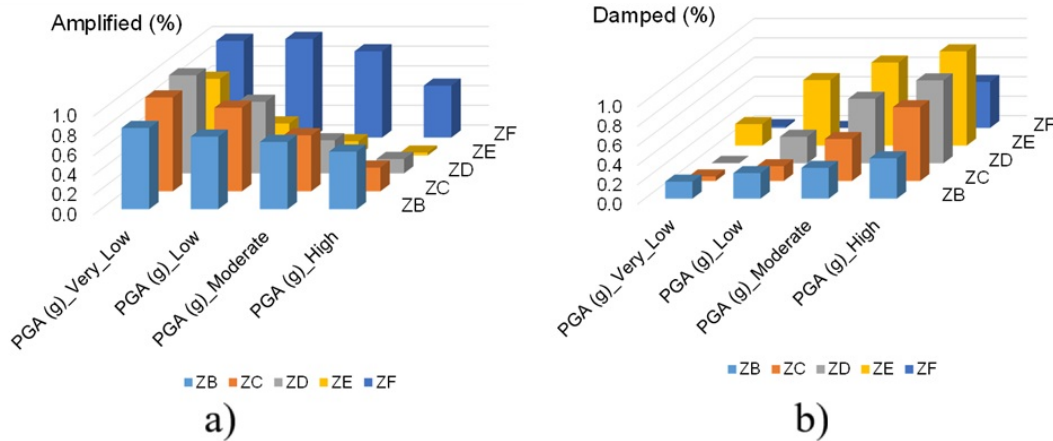

Figure 2: a) Total of amplified percentages b) Total of damped percentages

Table 5: Probabilities of amplified and damped respect to soil type and peak ground acceleration

Soil Type	Amplified				Damped			
	PGA (g)_ Very_Low	PGA (g)_ Low	PGA (g)_ Moderate	PGA (g)_ High	PGA (g)_ Very_Low	PGA (g)_ Low	PGA (g)_ Moderate	PGA (g)_ High
ZB	0.82	0.74	0.68	0.58	0.18	0.26	0.32	0.42
ZC	0.95	0.85	0.57	0.24	0.05	0.15	0.43	0.76
ZD	1.00	0.73	0.34	0.15	0.00	0.27	0.66	0.85
ZE	0.78	0.32	0.14	0.03	0.22	0.68	0.86	0.97
ZF	0.98	1.00	0.88	0.53	0.02	0.00	0.13	0.47

The change in amplified probabilities for same soil type with respect to the PGA value is calculated as minimum 24% (on ZB soil type) and maximum 85% (on ZD soil type). The change in damping probabilities with respect to the PGA value is equal to the change in amplified percentages. In the PGA-Very Low category, which is the most critical in terms of amplification, the soil class with the highest amplification was calculated in ZD (100%) soil type, while the lowest amplification was calculated in ZE (78%) soil type. In the PGA-Very High category, which is the most critical in terms of damping, the soil class with the highest damping was calculated in ZE (97%) soil type, while the lowest damping was calculated in ZB (42%) soil type. In line with these, the lowest amplification ratio and the highest damping ratio were calculated ZE soil type.

For PGA values greater than 0.16g (moderate or high), amplification ratios show a decreasing trend from more rigid to more flexible soils, while damping ratios show an increasing trend. The ZF soil class does not exhibit this trend. This may indicate that more stable results may be obtained for higher intensity ground motions depending on the soil type. Additionally, no specific trend was detected for PGA values less than 0.16g.

4 Conclusions

In the study, ground motion parameters and soil profiles are investigated for soil amplification with association rule mining analysis. For this purpose, 8400 ground motion records in surface form for 100 different soil profiles are considered. Effect on soil amplification of 31 different parameters related to ground motion frequency content, intensity and soil properties for each record were investigated. The categorization is based on Amplification parameters related to PGA, which represents intensity indicator of ground motion record. For a categorization compatible with the amplification categorization, the decision tree method was also used in the categorization of other parameters. The remarkable outcomes are summarized below:

- As shear wave velocity (V_{s30}) decreases, the risk of "Severe Amplification" increases. While the risk of severe amplification is 3.4% on ZB soil ($760 < V_{s30} < 1500$ m/s), it increased to 56.6% on ZF soil.

- The risk ratio of "Severe Amplification" is sensitive to the acceleration magnitude (PGA). If the PGA value is below 0.07g, the probability of the amplification exceeding 1.35 times is calculated as 0.63.

- The frequency content of the acceleration is another effective parameter on soil amplification. If T_{m1} In is low (<0.145), the probability of Severe Amplification condition is 52.5%. This ratio increased to 83.8% for Severe Amplification or Amplification conditions.

- Thanks to association rule mining, the effects of the frequency content of the accelerogram and the frequency content of the soil on the soil amplification were evaluated together. If $T_{g\ (proposed)}$ is moderate (0.073-0.171) and T_{m1} In is moderate (0.145-0.894), this may indicate a resonance situation, and the probability of Severe Amplification is calculated as 48.4%. This ratio increases to 84.9% for Severe Amplification or Amplification conditions.

- On the other hand, the amplified (amplification + severe amplification) and damped (damping + severe damping) percentages respect to the PGA values and soil types were calculated, separately. It was concluded that that amplified percentages are generally greater than damped percentages.

- As the PGA value increases, the amplification ratio decreases in each soil class, while the damping ratio increases. The ZF soil class does not exhibit this trend. For this reason, ZF soil class may prove special design needs respect to site-condition.

- The change in amplified and damped percentages respect to the PGA value is calculated between 24-85%.

- The soil class with the highest amplification ratio was calculated in ZD (100%) soil type, while the lowest amplification ratio was calculated in ZE (78%) soil type.

- The soil class with the highest damping ratio was calculated in ZE (97%) soil type, while the lowest damping ratio was calculated in ZB (42%) soil type.

- In line with these, the lowest amplification ratio and the highest damping ratio were calculated ZE soil type.

- For PGA values greater than 0.16g (moderate or high), amplification ratios show a decreasing trend from more rigid to more flexible soils, while damping ratios show an increasing trend. This may indicate that more stable results may be obtained for higher intensity ground motions depending on the soil class.

Acknowledgments

References

[1] Firat, S., Isik, N. S., Arman, H., Demir, M., & Vural, I. (2016). Investigation of the soil amplification factor in the Adapazari region. *Bulletin of Engineering Geology and the Environment, 75*(1), 141-152.

[2] Beresnev, I. A., Wen, K. L., & Tein Yeh, Y. (1995). Nonlinear soil amplification: its corroboration in Taiwan. *Bulletin of the Seismological Society of America, 85*(2), 496-515.

[3] Pitilakis, D., & Petridis, C. (2022). Fragility curves for existing reinforced concrete buildings, including soil–structure interaction and site amplification effects. *Engineering Structures, 269*, 114733.

[4] Malekmohammadi, M., & Pezeshk, S. (2015). Ground motion site amplification factors for sites located within the Mississippi embayment with consideration of deep soil deposits. *Earthquake Spectra, 31*(2), 699-722.

[5] Urzúa, A., Dobry, R., & Christian, J. (2017). Is harmonic averaging of shear wave velocity or the simplified rayleigh method appropriate to estimate the period of a soil profile?. *Earthquake Spectra, 33*(3), 895-915.

[6] Yunita, H., Setiawan, B., Saidi, T., & Abdullah, N. (2018). Site response analysis for estimating seismic site amplification in the case of banda aceh - Indonesia. *Matec Web of Conferences, 197*, 10002.

[7] Chen, G., Jin, D., Zhu, J., Jian, S., & Li, X. (2015). Nonlinear analysis on seismic site response of fuzhou basin, china. *Bulletin of the Seismological Society of America, 105*(2A), 928-949.

[8] Nagao, T. (2020). Seismic amplification by deep subsurface and proposal of a new proxy. *Engineering Technology & Applied Science Research, 10*(1), 5157-5163.

[9] Kamiyama, M. (1992), "Non-Linear Soil Amplification Identified Empirically from Strong Earthquake Ground Motions", *Journal of Physics of the Earth, 40*(1), 151-173.

[10] Moreno Ceballo, R., González Herrera, R., Paz Tenorio, J. A., Aguilar Carboney, J. A., & Del Carpio Penagos, C. U. (2019). Effects of sediment thickness upon seismic amplification in the urban area of Chiapa de Corzo, Chiapas, Mexico. *Earth Sciences Research Journal, 23*(2), 111-117.

[11] Loviknes, K., Cotton, F., & Weatherill, G. (2024). Exploring inferred geomorphological sediment thickness as a new site proxy to predict ground-shaking amplification at regional scale: application to Europe and eastern Türkiye. *Natural Hazards and Earth System Sciences, 24*(4), 1223-1247.

[12] Sun, C. G., & Kim, H. S. (2017). GIS-based regional assessment of seismic site effects considering the spatial uncertainty of site-specific geotechnical characteristics in coastal and inland urban areas. *Geomatics, Natural Hazards and Risk, 8*(2), 1592-1621.

[13] Prabowo, U. N., Sehah, S., Ferdiyan, A., & Sismanto, S. (2023). Quaternary Deposit Response to Earthquakes in Pemalang City Based on Peak Ground Acceleration, Earthquake Intensity, and Microtremor Method. *Indonesian Journal on Geoscience, 10*(3), 407-417.

[14] Jin, Y., Jeong, S., Moon, M., & Kim, D. (2024). Analysis of the dynamic behavior of multi-layered soil grounds. *Applied Sciences, 14*(12), 5256.

[15] de la Torre, C. A., Bradley, B. A., Kuncar, F., Lee, R. L., Wotherspoon, L. M., & Kaiser, A. E. (2024). Combining observed linear basin amplification factors with 1D nonlinear site-response analyses to predict site response for strong ground motions: Application to Wellington, New Zealand. *Earthquake Spectra, 40*(1), 143-173.

[16] Borghei, A., Ghayoomi, M., & Turner, M. (2020). Centrifuge tests to evaluate the effect of depth of water table on seismic response of shallow foundations on silty sands. In *E3S Web of Conferences* (Vol. 195, p. 01005). EDP Sciences.

[17] Ikram, A., & Qamar, U. (2014). A rule-based expert system for earthquake prediction. *Journal of Intelligent Information Systems, 43*(2), 205-230.

[18] Ikram, A., & Qamar, U. (2015). Developing an expert system based on association rules and predicate logic for earthquake prediction. *Knowledge-Based Systems, 75*, 87-103.

[19] Diana, L., Thiriot, J., Reuland, Y., & Lestuzzi, P. (2019). Application of association rules to determine building typological classes for seismic damage predictions at regional scale: The case study of Basel. *Frontiers in Built Environment, 5*, 51.

[20] PEER (2016). Pacific Earthquake Engineering Research Center [Internet]. 2016. Available from: `http://peer.berkeley.edu/smcat/index.html`

[21] Itaca-2016. Italian Accelometric Archive [Internet]. 2016. Available from: `http://itaca.mi.ingv.it/`

[22] Ozmen, H. B., Yilmaz, H., & Yildiz, H. (2019). An acceleration record set for different frequency content, amplitude and site classes. *Research on Engineering Structures & Materials, 5*(3), 321-333.

[23] Turkish Building Earthquake Code (TBEC-2018) (2018). Republic of Turkey Prime Ministry Disaster and Emergency Management Authory Presidential of Earthquake Department, Ankara, Turkey (in Turkish).

[24] Asce American Society of Civil Engineers, Minimum Design Loads for Buildings and Other Structures. ASCE, American Society of Civil Engineers, Reston, VA, USA, 2016.

[25] Beyaz, T. (2004). Zemin etkisinden arındırılmış deprem kayıtlarına göre Türkiye için yeni bir deprem enerjisi azalım bağıntısının geliştirilmesi. *Ankara Üniversite Fen Bilimleri Enstitüsü Doktora Tezi, 224s.*, Ankara.

[26] Edu Pro Civil System, "ProShake-V2.0". `www.proshake.com`

[27] Rathje, E. M., Abrahamson, N. A., & Bray, J. D. (1998). Simplified frequency content estimates of earthquake ground motions. *Journal of Geotechnical and Geoenvironmental Engineering, 124*(2), 150-159.

[28] Sawada, S. (2004, August). A simplified equation to approximate natural period of layered ground on the elastic bedrock for seismic design of structures. In *Proceeding of the 13th World Conference on Earthquake Engineering 2004*(1-6), # 1100.

[29] Bayardo Jr, R. J., & Agrawal, R. (1999, August). Mining the most interesting rules. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data mining* (pp. 145-154).

[30] Gao, W., (2004). *A Hierarchical Document Clustering Algoritm*. MSc Thesis, Dalhousie University, Halifax, Nova Scotia.

[31] Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT Press. (Adaptive Computation and Machine Learning Series).

[32] Roiger RJ & Geatz, M. (2003). *Data Mining: A Tutorial-Based Primer*. Boston MA: Addision Wesley.

[33] Agrawal, R., Imieliński, T., & Swami, A. (1993, June). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data* (pp. 207-216).

[34] Zhu, H., (1998). *On-Line Analytical Mining of Association Rules*. MSc. Thesis, Simon Fraser University, Ottawa, Canada.

[35] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In *Proceedings of the 20th International Conference on Very Large Databases (VLDB '94)* (pp. 487–489). Santiago, Chile.

[36] Sever, H., & Oğuz, B. (2002). Veri tabanlarında bilgi keşfine formel bir yaklaşım Kısım I: Eşleştirme sorguları ve algoritmalar. *Bilgi Dünyası, 3*(2), 173-204.