



Modeling of kappa factor using multivariate adaptive regression splines: application to the western Türkiye ground motion dataset

Tevfik Özgür Kurtuluş¹ · Fatma Yerlikaya-Özkurt² · Aysegül Askan³

Received: 15 September 2023 / Accepted: 29 February 2024 / Published online: 25 March 2024
© The Author(s) 2024

Abstract

The recent seismic activity on Türkiye's west coast, especially in the Aegean Sea region, shows that this region requires further attention. The region has significant seismic hazards because of its location in an active tectonic regime of North–South extension with multiple basin structures on soft soil deposits. Recently, despite being 70 km from the earthquake source, the Samos event (with a moment magnitude of 7.0 on October 30, 2020) caused significant localized damage and collapse in the Izmir city center due to a combination of basin effects and structural susceptibility. Despite this activity, research on site characterization and site response modeling, such as local velocity models and kappa estimates, remains sparse in this region. Kappa values display regional characteristics, necessitating the use of local kappa estimations from previous earthquake data in region-specific applications. Kappa estimates are multivariate and incorporate several characteristics such as magnitude and distance. In this study, we assess and predict the trend in mean kappa values using three-component strong-ground motion data from accelerometer sites with known V_{S30} values throughout western Türkiye. Multiple linear regression (MLR) and multivariate adaptive regression splines (MARS) were used to build the prediction models. The effects of epicentral distance R_{epi} , magnitude M_w , and site class (V_{S30}) were investigated, and the contributions of each parameter were examined using a large dataset containing recent seismic activity. The models were evaluated using well-known statistical accuracy criteria for kappa assessment. In all performance measures, the MARS model outperforms the MLR model across the selected sites.

Keywords Machine learning · Statistical methods · Seismic attenuation · Site effects · Multivariate adaptive regression splines (MARS)

1 Introduction

Studying variations in ground motion amplitudes as a function of distance and frequency is essential for accurately assessing seismic hazards and risk levels densely populated urban areas. In this regard, kappa (κ) is a well-known and practical parameter that characterizes the exponential decay of the high-frequency spectral attenuation of shear waves. It has

been used for various purposes, including stochastic simulations, regional adjustments of ground motion models (Atik et al. 2014; Bora et al. 2017), and correlations with site proxies, such as the average shear wave velocity of the top 30 m subsoil (V_{S30}) (van Houtte et al. 2014; Ktenidou et al. 2015; Cabas et al. 2017).

κ values exhibit regional characteristics, necessitating local κ estimates from past earthquake data for region-specific applications. κ studies regarding sites located in Türkiye include predictive models and estimations for the following regions. In Northwestern Türkiye, Askan et al. (2014) proposed a regional κ model for soil sites using a multivariate linear regression approach. Kurtulmus and Akyol (2015) analyzed ground motion data acquired from local small and moderate earthquakes to estimate regional κ values for central west Türkiye. Tanırçan and Dikmen (2018) predicted κ values for downhole arrays and bedrock sites in western Istanbul using small earthquake records. In their latest study, Biro et al. (2022) performed κ estimations for the eastern Anatolian region in Türkiye using data from a single fault segment to mitigate the source effects in κ values. Altındal and Askan (2022) proposed regional κ models for Eastern, Western, and Northwestern Türkiye using two different modeling approaches within a probabilistic framework. Sertcelik et al. (2022) determined κ values for 161 strong-ground motion stations operated by Türkiye's Disaster and Emergency Management Authority Presidential of Earthquake Department in the Marmara Region.

Since the pioneering work of Anderson and Hough (1984), understanding what causes the amplitudes of Fourier spectra of strong-ground motion to begin to decay with increasing frequency in the high-frequency range ($f > f_c$) remains a challenge in engineering seismology. Previous studies have alternating views on the source and site dependency of κ (Petukhin and Irikura 2000; Tsai and Chen 2000; Ktenidou et al. 2013). Significant research has been published focusing on deriving predictive models for κ as a function of earthquake magnitude, the type of source mechanism, source-to-site distance, and site class (Anderson and Hough 1984; Purvance and Anderson 2003; Motazedian and Atkinson 2005; Drouet et al. 2010; Iwakiri and Hoshiba 2012). In most of these applications, κ is assumed to be a linear function of distance and site effects, where a zero-distance kappa value (κ_0) is defined to eliminate path dependencies. Ktenidou et al. (2014) discussed that κ is mostly site dependent; however, to represent the physical phenomena comprehensively, parameters regarding the source and epicentral distance effects could also be included in the predictive models.

The relationship between the independent variables and κ may not be sufficiently described using traditional linear regression models because of considerable scattering in the individual κ_r values. Along with the potential nonlinearities of site response influencing κ at the recording site, a nonlinear variation of the observed κ with distance is possible. Sotiriadis et al. (2021) examined nonlinear regression analyses to relate κ , as computed for free-field and foundation motions, to magnitude, source-to-site distance, and V_{S30} . Recently, Gičev and Trifunac (2022) stated that the nonlinear behavior of soil and rock materials should be considered when predicting κ from greater than $M > 5$ events, and they proposed a physical model that κ can be associated with nonlinear site response. Moreover, various researchers have determined the average κ along the S-wave source-station paths using nonparametric inversion approaches (e.g. Anderson 1991; Fernández et al. 2010; Castro et al. 2022).

An effective approach to estimate the complex nature of the κ parameter, influenced by multiple interrelated parameters, involves the use of robust data-driven methodologies such as Multivariate Adaptive Regression Splines (MARS). MARS is a nonparametric, multistage regression method that captures nonlinear interactions between variables in several dimensions

(Friedman 1991). The primary advantage of MARS is its ability to accurately estimate the contributions of the input factors to the response variable. The algorithm automatically identifies the most salient qualities within the dataset and incrementally incorporates terms into the model. Each term represents a primary or interaction effect of a piecewise linear function that is included in the model if it improves the model's goodness of fit. The approach iteratively incorporates additional terms until the model reaches a point where further improvement is no longer observed. MARS is a well-known statistical methodology that has been used in numerous susceptibility evaluations of natural disasters such as floods, landslides, and wildfires (e.g. Felicísimo et al. 2013; Conoscenti et al. 2015; Hai et al. 2023). Its effective application in several domains of geotechnical engineering, such as determining the elastic modulus of rocks (Samui 2013), analyzing soil liquefaction (Zhang et al. 2015; Zheng et al. 2020), assessing soil and wall characteristics in braced excavation (Zhang et al. 2017), and evaluating slope reliability (Deng et al. 2021), has been well documented. To gain a thorough understanding of the MARS technique and its practical applicability to geotechnical problems, please refer to Zhang (2020). The recent trend involves integrating AI-based techniques and other data-driven methods with the MARS methodology to achieve higher accuracy, improved interpretability, and valuable insights in diverse scientific fields (e.g. Zhang et al. 2021, 2022; Asare et al. 2023; Vaheddoost et al. 2023; Hong et al. 2024). Therefore, it may be possible to mitigate the inherent limitations of a singular technique, such as insufficient memory, overfitting, and uncertainty, by using a convenient hybrid approach.

The recent seismic activity on the west coast of Türkiye, including the Aegean Sea region, indicates that a closer focus is necessary on this region. Located in an active tectonic regime of North–South extension with multiple basin structures on soft soil deposits, the Aegean Sea region in Türkiye has a high seismic hazard. Recently, as a combination of basin effects and building vulnerability, on October 30, 2020, the Samos event ($M_w=7.0$) caused localized major damage and collapses in the Izmir city center despite the 70 km distance from the earthquake source. Despite this activity, studies on site characterization and site response modeling, including local velocity models and κ estimates, are still limited in this region (Akyol et al. 2013; Kurtulmus and Akyol 2015; Pamuk et al. 2018, 2019).

This study attempts to develop predictive κ models for western Türkiye with a particular focus on the Aegean Sea region. We used three-component strong-ground motion data from accelerometer stations with known V_{S30} values throughout western Türkiye, including Aydın, Denizli, Izmir, Kutahya, Manisa, and Mugla in the Aegean region, as well as Balikesir and Canakkale provinces. Prediction models were built using multiple linear regression (MLR) and MARS techniques. According to the literature, MARS has not been previously used in the modeling of high-frequency spectral attenuation of shear waves. The impacts of epicentral distance R_{epi} , magnitude M_w , and V_{S30} (site class) are explored, and the contributions of each parameter are analyzed using a large dataset that includes recent seismic activity. In κ assessment, models are validated using well-known accuracy metrics such as correlation coefficient (r), R -square, adjusted R -squared (Adj. R^2), mean absolute percentage error (MAPE), and mean squared error (MSE).

2 Study area and tectonic framework

The Aegean Sea and the basin-and-range system in the western part of Türkiye formed as a result of a long series of tectonic regimes, including the collision between the Arabian and Eurasian plates and the northward movement of the African lithosphere beneath Eurasia

(Jolivet et al. 2013; Şengör and Yazıcı 2020). The collision– and subduction–related mechanisms have led to the Anatolian block’s westward escape along the right–lateral North Anatolian and the left–lateral East Anatolian Fault Zones (NAFZ and EAFZ, respectively) and the subsequent African slab roll–back and trench retreat along the Aegean–Cyprian Subduction Zone (Şengör et al. 1985; Dewey et al. 1986; Bozkurt 2001; Le Pichon and Kreemer 2010; Barbot and Weiss 2021). As a result, the continental crust has expanded and volcanism has occurred in the overlying Aegean–west Anatolian region (Fig. 1) (Jolivet et al. 2013). Although the origin and evolutionary history of the principal extensional features remain a matter of debate, western Türkiye is characterized by a complex array of grabens and basins bounded by large–scale extensional detachments and high–angle normal faults. (Brun and Sokoutis 2007; Çiftçi et al. 2010). The horst–graben architecture arises in different geographical trends, including EW–trending grabens (e.g. from the Gulf of Edremit to the north and toward the Gulf of Gökova in the south), NE–trending depressions (e.g. Gördes basin, Acıpayam and Burdur grabens), and intervening horsts, respectively (Bozkurt and Mittweide 2005; Çiftçi et al. 2010). All basins contain the Menderes core complex, Lycian nappes, and Tauride platform basements and are filled with lower–middle Miocene to Holocene continental deposits of more than 3000 m in some locations (Çiftçi et al. 2010; Özkaptan et al. 2021).

As one of the areas with the highest intensity and frequency of seismic activity and the most severe seismic and geological hazards on the Turkish mainland, numerous strong earthquakes of magnitude ≥ 6 have been observed in western Türkiye over the past ten

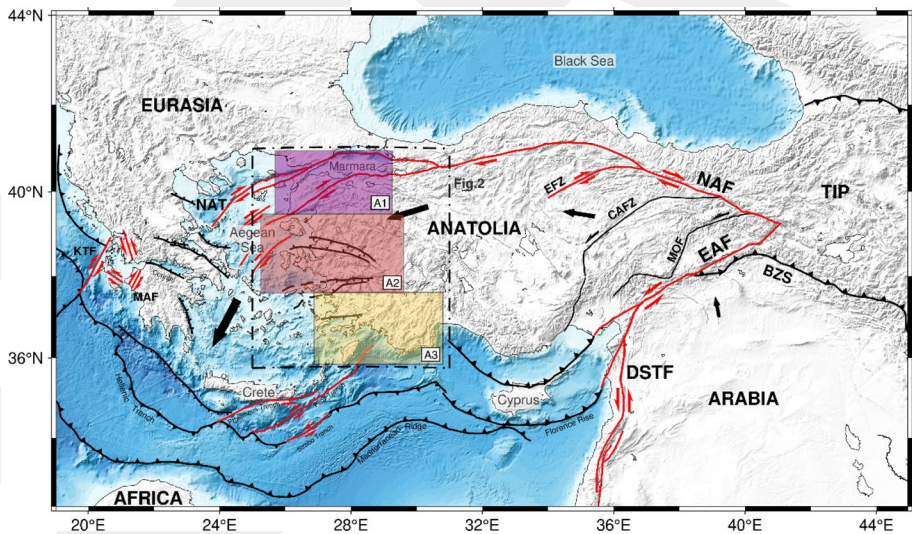


Fig. 1 Location map showing the main tectonic elements of the Aegean and Anatolian regions (adapted from Barbot and Weiss (2021)). The black arrows indicate the direction of plate movement relative to Eurasia. The east–to–west increase in the size of the arrows reflects the documented 20 mm/yr increase in surface velocity. The dashed black box denotes the location of the study area. The research region was divided into smaller sections as A1, A2, and A3 in accordance with the major seismotectonic domains identified by Duman et al. (2018). The major tectonic features on the map include *BZS* Bitlis–Zagros Suture, *CAFZ* Central Anatolian Fault Zone, *DSTF* Dead Sea Transform Fault, *EAF* East Anatolian Fault, *EFZ* Ezinepazari Fault Zone, *KTF* Kefalonia Transform Fault, *MAF* Movri–Amaliada Fault, *MOF* Malatya–Ovacik Fault, *NAF* North Anatolian Fault, *NAT* North Anatolian Trough, and *TIP* Turkish–Iranian Plateau

years, such as the 2014–M 6.9, Gökçeada; 2017–M 6.2 Karaburun; 2017–M 6.4 Kos; 2019–M 6.0, Denizli; and 2020–M 7.0 Samos earthquakes. Large-magnitude earthquakes have also occurred in the study area since 1900, e.g. 1919–Soma M 6.9, 1928–Torbalı M 6.3, 1933–Gökova M 6.8, 1956–Söke–Balat M 7.1, 1969–Alaşehir M 6.5, and 1970–Gediz M 7.2 earthquakes. Spatially distributed, relatively frequent, moderate magnitude–crustal earthquakes are concentrated in the western Anatolia graben systems, where normal faulting mechanisms are associated with extensional tectonics (Duman et al. 2018). The seismic pattern represents swarm-type activity with remarkable clusters (e.g. the Urla earthquakes swarm with M 5.0, 5.8, 5.5, and 5.9). Most of them were recorded as shallow and restricted to a seismogenic zone within the upper 30–40 km of the crust, except for the events that occurred in the southern Aegean area, which reached a depth of approximately 180 km (Bocchini et al. 2018). The bulk of the moderate-to-large magnitude events accumulated near the plate boundaries, which correspond to the northern and southern margins of the study area. Recent seismicity studies have identified that the dominant strike-slip stress regime is concentrated on the western branch of the NAFZ and its continuation in the Marmara Sea and the North Aegean Trough (Vamvakaris et al. 2016). This confirms the co-existence of active extension and strike-slip deformation along the coastal region of western Anatolia and the eastern Aegean Sea (Kiritzi et al. 2021). Moreover, the southern Aegean domain includes large thrust earthquakes on adjacent Hellenic and Cyprus Arc segments as well as a mixture of normal, strike-slip, and splay-thrust faulting events.

In this study, we separated the study area into smaller parts on the basis of criteria mainly related to the spatial distribution of earthquakes and similar types of fault zones identified in the region. Using the major seismotectonic domains proposed by Duman et al. (2018), we divided the study area into three smaller sections: (1) the western branch of the NAFZ as Area 1 (A1), (2) the western Anatolia graben systems as Area 2 (A2), and (3) the Aegean arc as Area 3 (A3). Thus, we only selected data with epicentral distances less than 200 km to avoid additional path complications.

3 Data

We compiled three-component strong-ground motion records obtained with 100 samples per second at accelerometer stations within the Turkish National Strong Ground Motion Observation Network operated by the Disaster and Emergency Management Presidency (AFAD). Raw versions of the data are available on the Turkish Accelerometric Database and Analysis System (TADAS, via <https://tadas.afad.gov.tr>). Records were selected on the basis of stations with known V_{S30} values, including Aydın, Denizli, İzmir, Kutahya, Manisa, and Muğla in the Aegean region and Balıkesir and Canakkale provinces in western Türkiye (Fig. 2). The station details for all sites are listed in Supporting Information Tables S1, S2, and S3, respectively. The multichannel analysis of surface waves (MASW) method was used to calculate station velocity profiles and mean V_{S30} values (Sandikkaya et al. 2010). Further details on the site characterization of these stations are available on the Strong Ground Motion Database of Türkiye (via <https://deprem.afad.gov.tr>). Because earthquake magnitudes have been reported on several scales, including the duration, local, and moment magnitude scales, we converted all of them to M_w , applying the conversion equation of Kadirioğlu and Kartal (2016) for consistency.

Figure 3a displays pairwise relationships among all selected independent variables, including epicentral distance (R_{epi}), depth, magnitude (M_w), and κ for whole data set. Our

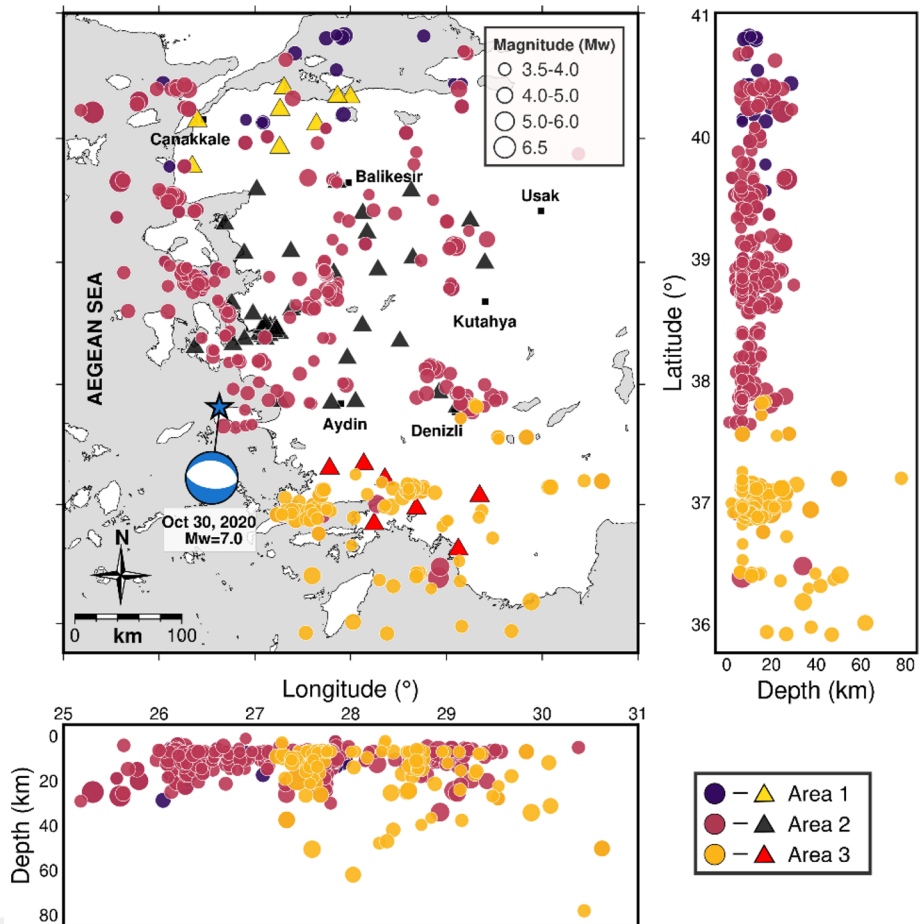


Fig. 2 Locations of the stations (triangles) and the distribution of earthquakes (circles) were used in this study. The yellow triangle–purple circle, black triangle–red circle, and red triangle–yellow circle pairs correspond to the strong–ground motion station and earthquake couples used in κ calculations for A1, A2, and A3. The blue star represents the epicenter of the M_w 7.0 Samos earthquake that occurred on October 30, 2020. The blue beachball displays the focal mechanism solution provided by USGS

dataset consists of 1398 records (4194 components) measured at 58 strong–ground motion stations from 408 earthquakes recorded between July 2003 and November 2018. The data cover earthquakes with M_w ranging from 3 to 6.5. Most of the records relate to events within M_w range of 3.5–4.5, while the remaining data are evenly distributed to larger earthquakes. Moreover, a few earthquakes with M_w over 5.5 are also included in the dataset. The epicentral distances range from 5.0 to 200 km, with most being between 10 and 100 km. The number of records decreases as R_{epi} increases. The strong–motion data cover various focal depths, from 1 to 78 km. Additional information regarding the earthquakes can be found in the Supporting Information (see Tables S4, S5, and S6 for each area, respectively).

Because the study area is divided into three smaller parts, all computations have been separately performed for these areas. Statistical distributions of recordings with respect to R_{epi} , M_w , focal depths, and κ can be seen in Fig. 3b, c, and d for the A1, A2, and A3

areas. These figures also include histograms showing the number of records within the independent variable bins. The size of the datasets is approximately the same for A1 and A3, whereas it is almost six times larger for A2. It can be easily seen that the distance distribution is reasonably homogeneous for the three regions. The selected data subsets indicate that there is no obvious trend with either R_{epi} , depth, or M_w , indicating that no bias is included during the κ calculations. The bulk of the data, which substantially originates from the central–western part of Türkiye, consists of small–to moderate–magnitude events (up to $M < 5.5$) and shallow depths within the range of 1–30 km, whereas significant events occurring at depths greater than 30 km originating from A3 are recorded.

We determine κ values using the original method proposed by Anderson & Hough (1984), where the high–frequency spectral decay is modeled as follows (e.g. Purvance and Anderson 2003; Castro et al. 2022; Lanzano et al. 2022):

$$A(f) = A_0 e^{-\pi\kappa f}, f > f_e \quad (1)$$

where the amplitude A_0 depends on the source and path properties and f_e is the frequency above which the spectrum is approximated with a linear decay on a log (Amplitude) versus linear frequency plot.

Previous studies have mainly focused on the examination of horizontal κ values. However, it is critical to investigate the variability in κ values obtained from all three components of a single record for certain purposes. Stochastic ground motion simulations may require vertical κ values to adjust the empirical horizontal–to–vertical curves used for site amplifications (Motazedian 2006). When three–component stations are not available, κ values obtained from the vertical components can be used as an initial estimate for this parameter. However, it may be necessary to make slight adjustments to the κ values derived from the vertical component (Douglas et al. 2010). In this study, the North–South (NS), East–West (EW), and vertical (Z) components for each triaxial acceleration record are processed individually. We initially corrected the baseline for each time series by subtracting the linear trend. The lengths of the S–wave and noise windows, which vary between 3 and 40 s depending on the M_w and source–to–site distances, are manually decided for each record for precision and consistency. Following 5% cosine tapering at both edges of the time window, the Fourier amplitude spectrum (FAS) of the S–wave (and noise signal) is computed and plotted in log–linear space after smoothing between 0.3 and 35 Hz. Then, κ factor is computed from the FAS by fitting a straight line with least–squares linear regression where the spectral decay starts (f_c) and ends (f_x). We note that the lower and upper frequency bounds for the exponential decay are determined manually. While selecting these frequencies, the method requires satisfying the following conditions to avoid bias in κ estimates: the frequency range chosen (1) must lie above the corner frequency (f_c) due to the relation between seismic moment and f_c , (2) should supply adequate bandwidth to ensure robustness to the regressions and to resolve source, path, and site effects where the instrument’s response is considered flat (e.g. Ktenidou et al. 2017), and (3) should not exceed the high–frequency noise level (e.g. Douglas et al. 2010; Ktenidou et al. 2013).

On the basis of these criteria, we rejected events with M_w lower than 3.5 to avoid biased– κ estimations caused by the source effect on the measurements. We calculated empirical f_c values of M_w varying from 3.5 to 6.5 using the theoretical relationship among seismic moment, stress drop, and corner frequency described by Brune (1970). For M_w 3.5, we obtained typical estimates of f_c ranging from 1.36 to 6.32 Hz with an average shear–wave velocity in the crust of 3.5 km/s and variable stress drop values between 1 and 100 bars. Thus, f_c points greater than 6.5 Hz were selected for the events of M_w between

Fig. 3 Pairwise correlations among multiple variables including R_{epi} , depth, M_w , and κ in the following: **a** whole data set, **b** A1, **c** A2, and **d** A3. Each diagonal also includes histograms that demonstrate the number of records within the independent variable bins

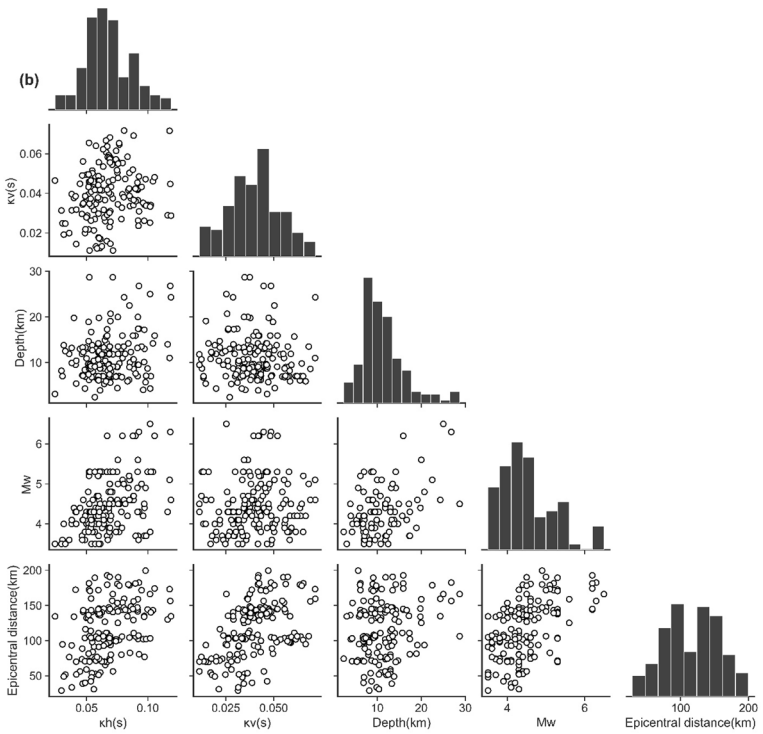
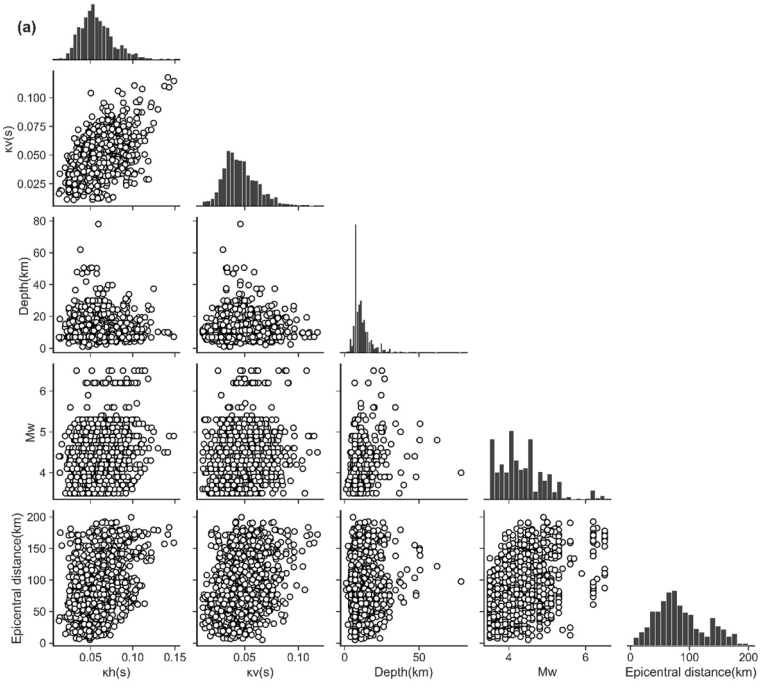
3.5 and 4 to limit the impact of f_c . Further details on the calculations can be found in Ktenidou et al. (2017). We also ensure that $f_c > f_c$ on the displacement amplitude spectra in the log–log scale. Thus, we ensure f_c exceeds f_c for all considered events. Several studies have shown that resonance peaks in site transfer functions may cause inaccurate selection of the initial frequency value or high–frequency site amplification, leading to significant overestimation or underestimation κ (Parolai and Bindi 2004; Kishida et al. 2014). Therefore, empirical transfer functions estimated through the horizontal–to–vertical spectral ratio for the stations considered here are studied to confirm that the fundamental frequencies are consistently below f_c , where the exponential decay starts. In selecting f_x , we analyzed the signal–to–noise ratios (SNR) and used only data with SNR above 3. A value of f_x equal to 35 Hz (70% of Nyquist frequency as a limit) was used to guarantee an entire frequency range in which the instrument has a flat response to ground acceleration. Furthermore, we only considered ground motions within a R_{epi} of 200 km to minimize the potential for capturing multiple ray paths because of the structural complexity of the region (Kurtulmus and Akyol 2015). Finally, we selected stations with at least 15 or more suitable ground motion records per station.

The average value of f_c is around 10 Hz for both horizontal components and 12 Hz for the vertical one, but it has a wide range between 3 and 25 Hz for all components, which is consistent with previous studies such as (Anderson and Hough 1984; Douglas et al. 2010). Here, we observe that f_c is smaller for large earthquakes than for moderate and small earthquakes, and this could be related to a seismic moment dependency as discussed by Tsurugi et al. (2020). The value of f_x was found to vary between 20 and 35 Hz. Figure 4 displays the sample κ calculation, including the S–wave and noise spectra. The κ values eventually determined from the NS and EW components are then averaged to yield a single horizontal κ value per event and station to reduce the influence of propagation heterogeneity. Askan et al. (2014) stated that such an averaging assumes that the direction of the incoming waves does not affect the κ value. To improve our analyses, we excluded individual κ calculations with a difference greater than 25% between the 2 horizontal components (e.g. van Houtte et al. 2011; Ktenidou et al. 2013).

4 Methodology

4.1 Multiple linear regression (MLR) model

We employed a regression model to determine the correlation between κ estimates and independent variables, such as M_w , R_{epi} , and V_{S30} . This approach is widely accepted and frequently used in data–driven mathematical analysis, although it has several variations. The MLR model is extensively preferred in many applications because of its well–established form and available computer packages. It is generally used to construct a relationship between selected independent (explanatory) variables and a dependent (response) variable. The standard MLR form (Neter et al. 1996), with N observations and p explanatory variables (x_1, x_2, \dots, x_p), is given as follows:



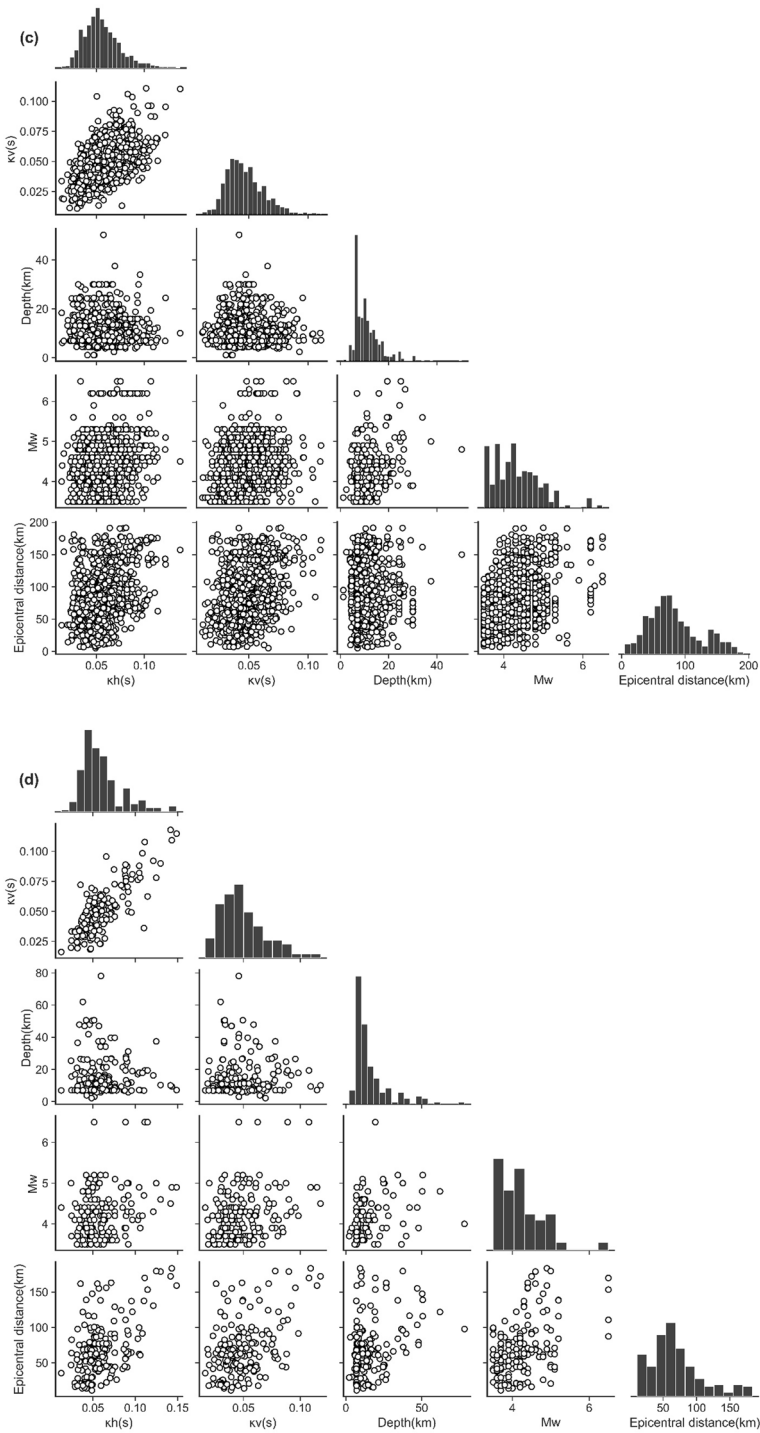


Fig. 3 (continued)

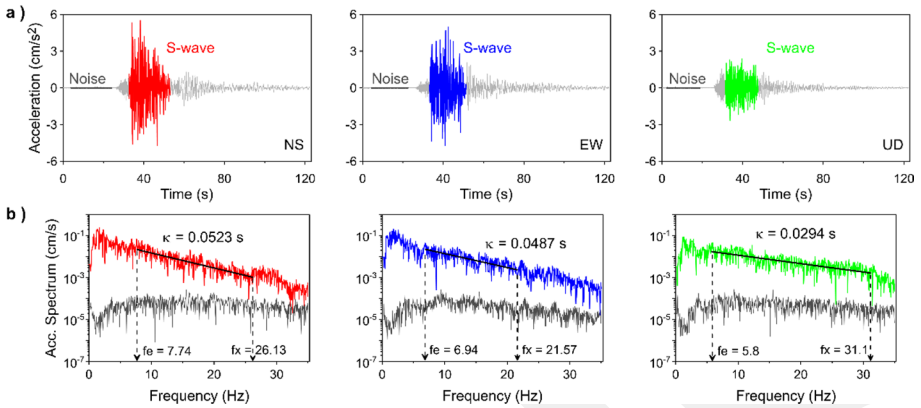


Fig. 4 Data from an M_w 5.1 earthquake on May 27, 2017, at 15:53:23. Acceleration time-series and their corresponding spectra from station 3509. **a** Three-component records of S-wave and noise signals are highlighted and labeled in the time-series plot. **b** Examples of the FAS of the S-window (the upper jagged line) and the noise signal in gray (lower jagged line) for each component. Vertical dashed lines represent the picking of f_c , the lower bound, and f_x , the upper bound on the FAS. The black line represents the fit to the spectrum over the frequency band for the κ measurement

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i, (i = 1, 2, \dots, N) \tag{2}$$

The matrix-vector notation of this model can also be expressed as follows:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N1} & \dots & x_{Np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{3}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N1} & \dots & x_{Np} \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \tag{4}$$

In short:

$$y = X\beta + \varepsilon \tag{5}$$

where y is an $(N \times 1)$ vector of the response variable, X is an $(N \times (p + 1))$ matrix of explanatory variables, β is a $((p + 1) \times 1)$ vector of unknown parameters, and ε is an $(N \times 1)$ vector of independent, identically distributed random errors. In this approach, to estimate unknown parameters, the least squares estimation method is used.

4.2 Multivariate adaptive regression splines (MARS) model

The commonly used MLR requires certain assumptions to be fulfilled. Therefore, its prediction power is restricted to the estimation of general functions of high-dimensional and sparse

data. In this study, we used a sophisticated nonparametric regression method called MARS, which makes no specific assumption about the underlying functional relationship between the response and input variables (Friedman 1991). The MARS model is represented by a linear combination of the basis functions (BFs) and the intercept as follows:

$$y = \theta_0 + \sum_{m=1}^M \theta_m B_m(x^m) + \varepsilon \tag{6}$$

where B_m ($m = 1, 2, \dots, M$) are BFs taken from a set of M linearly independent basis elements, and θ_m are the unknown coefficients for the m th BF ($m = 1, 2, \dots, M$). BFs can be in the form of the main or interaction of two or more spline functions. The following piecewise linear functions given in Eq. (7) that involve one independent variable with a knot value τ show main effects and are *reflected pairs* of each other:

$$c^+(x, \tau) = \max\{0, x - \tau\}, c^-(x, \tau) = \max\{0, \tau - x\} \tag{7}$$

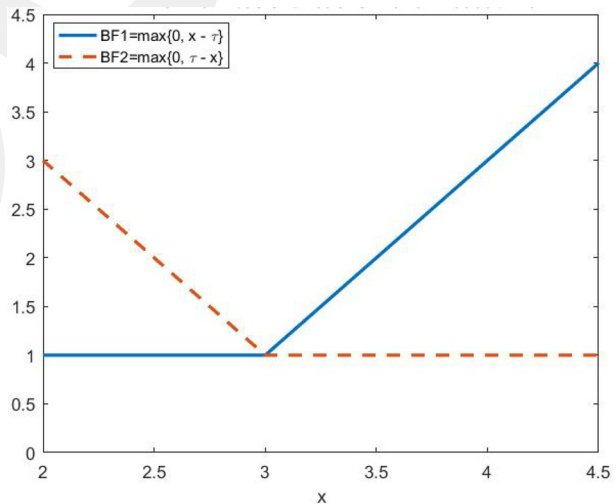
Figure 5 shows an illustration of these BFs with a knot value at $\tau=3$.

Because the MARS algorithm is an *adaptive* procedure, the selection of BFs is data-based and specific to the problem at hand. A special advantage of MARS is that it is also possible to estimate the response by the contributions of the BFs with the interaction effects of the independent variables. For a given data (\bar{x}_i, \bar{y}_i) ($i = 1, 2, \dots, N$), the form of the m th BF with the interaction effect is as follows:

$$B_m(x^m) := \prod_{j=1}^{K_m} \left[S_{\kappa_j^m} \cdot (x_{\kappa_j^m} - \tau_{\kappa_j^m}) \right]_+ \tag{8}$$

where $[q]_+ := \max\{0, q\}$, K_m is the number of truncated piecewise linear functions multiplied in the m th BF, $x_{\kappa_j^m}$ is the input variable corresponding to the j th truncated piecewise linear function in the m th BF, $\tau_{\kappa_j^m}$ is the knot value corresponding to the variable $x_{\kappa_j^m}$, and $s_{\kappa_j^m}$ is the selected sign +1 or -1. Figure 6 graphically represents an example of interaction

Fig. 5 Pair of main BFs with a knot at $\tau=3$



BF obtained based on two predictor variables (x_1 and x_2) and their knot values ($\tau_1 = 0.5$ and $\tau_2 = 0.1$) respectively.

The smoothness constraints in MARS are implemented using BFs to which a piecewise linear function is linked, as defined in Eq. (7). This function governs the contribution of a specific basis function over different parts of the input space. Smoothness constraints are inherent properties of these functions. The following factors affect the level of smoothness in the MARS-generated κ functions:

- Number of Basis Functions: increasing the number of basis functions allows the model to capture complex features and non-linear patterns in the data, but it may result in a less smooth overall fit.
- Number of Knots: an insufficient number of knots may result in an overly simplistic model, whereas an excessive number of knots might lead to a less smooth function.
- Interaction Terms: MARS allows the use of interaction terms, but it is important to note that multiple interactions may result in a less smooth response function.
- In the backward step of the MARS algorithm, the degree of pruning affects the overall smoothness of the response function.

The MARS algorithm consists of two parts (Friedman 1991): the forward stepwise algorithm searches for the BF, and at each step, the split that minimizes the residual sum of squares (RSS) criterion from all possible splits on each BF is chosen. The search for new BFs is restricted by the user-specified value M_{\max} and the maximum interaction order.

Two way interaction basis function with knot values at $\tau_1 = 0.5$ and $\tau_2 = 0.1$

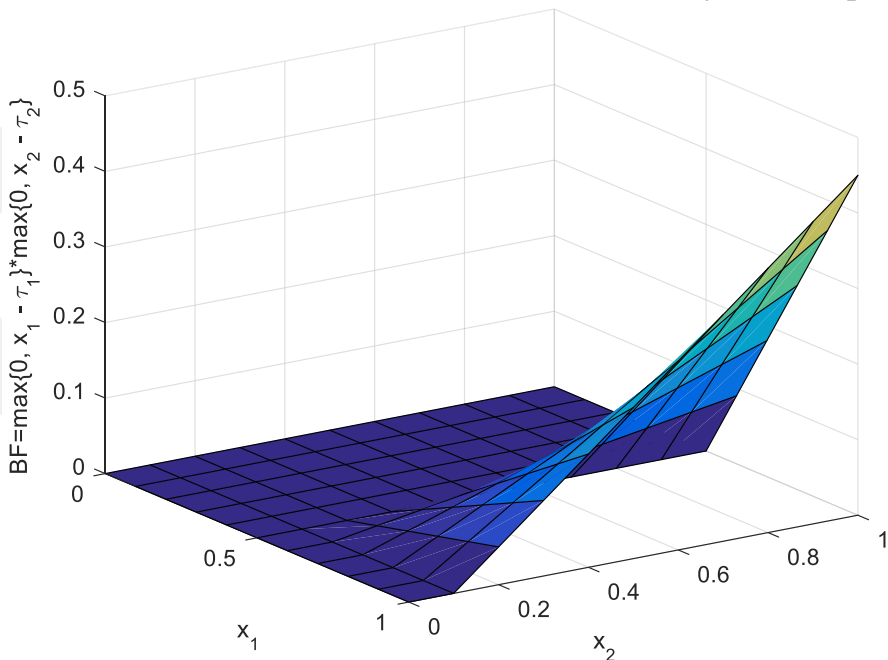


Fig. 6 Graphical representation of the interaction basis function based on x_1 and x_2 predictor variables

In the backward stepwise algorithm, on the other hand, the purpose is to prevent over-fitting by decreasing the complexity of the model without degrading the fit to the data by removing some BFs that contribute less to the model (i.e. the smallest increase in the RSS). MARS algorithm uses a generalized cross-validation (GCV) criterion to estimate an optimal number of parameters in the model. GCV is used in both parts of the MARS algorithm and is defined as:

$$GCV = \frac{1}{N} \frac{\sum_{i=1}^N (y_i - \hat{f}_M(\theta, x_i))^2}{(1 - Q(M)/N)^2} \quad (9)$$

where the numerator is the RSS and $Q(M)$ in the denominator represents the cost penalty measure of a model with M BFs. The final MARS model is obtained when the minimum value of the GCV is reached.

Summaries of the advantages of the MLR and MARS models used in this study are as follows: MLR is a statistical model used to analyze the relationship between a dependent variable and multiple independent variables. This model is simple and easy to understand. Compared with more complex models, linear regression models have better training and forecasting speeds. However, it is crucial to note that MLR may not be suitable for complex relationships or situations that involve significant non-linearities. Conversely, MARS excels at capturing non-linear correlations in the data. Unlike some complex machine learning models, it provides an interpretable and transparent representation of the relationship between the predictors and the response variable. The explicit functional form, which includes linear segments and breakpoints, allows for a precise understanding of the model's behavior. During model construction, the MARS algorithm identifies the most significant variables and their relationships while disregarding the less impactful ones. This can result in models that are more comfortable and easier to understand, and better generalizations to new data.

5 Application and results

The findings of the MLR and MARS models for predicting κ values in western Türkiye were analyzed to validate the study. The majority of research in the existing literature illustrates the correlation between κ and distance. In this area, where most earthquakes occur in the shallow crust, various earthquake distance measurements, such as epicentral and hypocentral distances, do not have a significant impact. We experimented with several statistical metrics and observed that the outcomes did not differ substantially. Hence, we have chosen to use R_{epi} as the main distance metric in the models. A limited number of studies argue that magnitude could influence κ as well, implying the existence of source-related effects on high-frequency spectral amplitudes (e.g. Purvance and Anderson 2003). To investigate the potential influence of magnitude, this study employs the MLR and MARS methodologies, where M_w is considered an independent variable in the κ model predictions. Moreover, the V_{S30} value, denoting the time-averaged shear wave velocity within the initial 30 m of depth, is the most common proxy for site classification globally. Therefore, in this study, R_{epi} , M_w , and V_{S30} are input variables included in the evaluation process. Tables 1, 2, 3 present the descriptive statistics for all variables related to each area. The tables show that the variables have a varied range of values. Hence, all input data are normalized so that each has a mean of zero

and a variance of one to overcome variable scale impacts on the findings while performing MLR and MARS modeling. Some well-known statistical performance indicators, such as the correlation coefficient (r), R -square, adjusted R -squared ($Adj. R^2$), mean absolute percentage error (MAPE), and mean squared error (MSE), were generated to assess the effectiveness and accuracy of the models. The following section compares the subsets based on the model metrics.

After the MLR and MARS prediction models were built for A1, some measures for the prediction performance of the models were computed, as listed in Table 4. The following results show that the MARS model performs quite well for the given dataset. Here, the smaller values of MSE and MAE and the higher values of R^2 , $Adj.R^2$, and r show better performance of the associated model. Moreover, scatter plots of the prediction models for both the MLR and MARS models are shown in Fig. 7 shows the performances by using the correlation between predicted and actual values. This figure indicates that the MARS model results for both the horizontal and vertical components have a better prediction of true values. The BFs in the MARS model use all predictor variables in the following order of importance: R_{epi} , V_{s30} , and M_w .

If we compare the MARS and MLR performance results for A2, the prediction capability of the MARS model is again stronger and the model captures the main structure

Table 1 Descriptive statistics of variables for A1

| Statistics | $\kappa_h(s)$ | $\kappa_v(s)$ | M_w | $R_{epi}(km)$ | $V_{s30}(m/s)$ |
|--------------|---------------|---------------|-------|---------------|----------------|
| Minimum | 0.0243 | 0.0111 | 3.5 | 29 | 192 |
| 1st Quartile | 0.0542 | 0.0314 | 4 | 89.3 | 321 |
| Median | 0.0655 | 0.0396 | 4.3 | 110.8 | 330 |
| Mean | 0.0678 | 0.0399 | 4.468 | 116.5 | 372 |
| 3rd Quartile | 0.0794 | 0.0481 | 4.875 | 144.3 | 403 |
| Maximum | 0.1189 | 0.0718 | 6.5 | 199.6 | 683 |

Table 2 Descriptive statistics of variables for A2

| Statistics | $\kappa_h(s)$ | $\kappa_v(s)$ | M_w | $R_{epi}(km)$ | $V_{s30}(m/s)$ |
|--------------|---------------|---------------|-------|---------------|----------------|
| Minimum | 0.0136 | 0.0109 | 3.5 | 5.2 | 131 |
| 1st Quartile | 0.0429 | 0.0355 | 3.9 | 55.5 | 270 |
| Median | 0.0529 | 0.0445 | 4.2 | 75.4 | 340 |
| Mean | 0.0561 | 0.0467 | 4.316 | 82.8 | 396.1 |
| 3rd Quartile | 0.0662 | 0.0558 | 4.6 | 105 | 460 |
| Maximum | 0.1375 | 0.1109 | 6.5 | 191.6 | 875 |

Table 3 Descriptive statistics of variables for A3

| Statistics | $\kappa_h(s)$ | $\kappa_v(s)$ | M_w | $R_{epi}(km)$ | $V_{s30}(m/s)$ |
|--------------|---------------|---------------|-------|---------------|----------------|
| Minimum | 0.0126 | 0.01634 | 3.5 | 11.1 | 248 |
| 1st Quartile | 0.0434 | 0.0346 | 3.7 | 43 | 338.8 |
| Median | 0.0539 | 0.046 | 4.2 | 62 | 372 |
| Mean | 0.0602 | 0.0498 | 4.22 | 69.3 | 398.4 |
| 3rd Quartile | 0.0693 | 0.0594 | 4.5 | 86.2 | 393 |
| Maximum | 0.1487 | 0.1179 | 6.5 | 183.6 | 696 |

Table 4 Comparison of performance measures for MLR and MARS in A1

| | | MSE | MAE | r | R ² | Adj.R ² |
|------------|------|--------|--------|--------|----------------|--------------------|
| Horizontal | MLR | 0.0292 | 0.1322 | 0.5194 | 0.2698 | 0.2563 |
| | MARS | 0.0167 | 0.1006 | 0.7634 | 0.5828 | 0.5501 |
| Vertical | MLR | 0.0354 | 0.1508 | 0.545 | 0.297 | 0.284 |
| | MARS | 0.0139 | 0.0922 | 0.8513 | 0.7247 | 0.7031 |

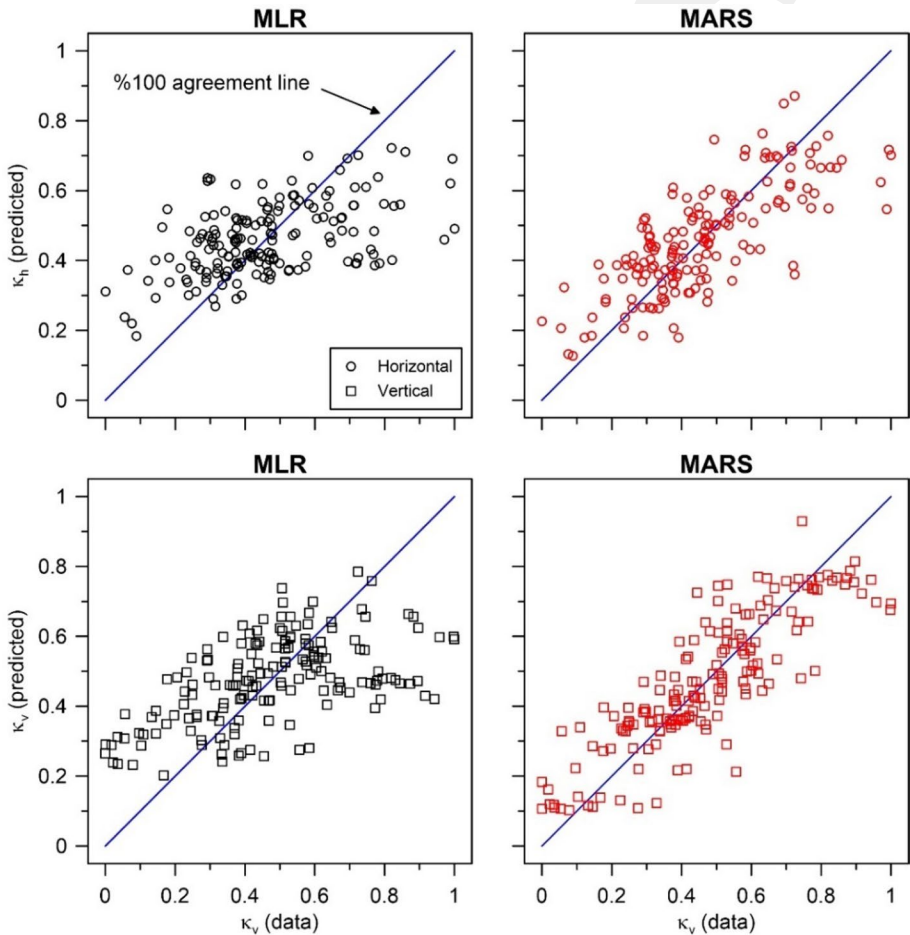


Fig. 7 Comparison of MLR and MARS predictions for A1. All plots show actual κ vs predicted values. Circles and squares correspond to the horizontal and vertical components, respectively. Blue lines represent the case when predicted κ equals to actual κ

of the given dataset well (see Table 5 and Fig. 8). The variable importance of the MARS model obtained according to the use of variables in the model is R_{epi} , V_{S30} , and M_w in order.

Table 5 Comparison of performance measures for MLR and MARS in A2

| | | MSE | MAE | r | R ² | Adj.R ² |
|------------|------|--------|--------|--------|----------------|--------------------|
| Horizontal | MLR | 0.0182 | 0.1053 | 0.4214 | 0.1776 | 0.1753 |
| | MARS | 0.0118 | 0.0845 | 0.6848 | 0.469 | 0.448 |
| Vertical | MLR | 0.0203 | 0.1108 | 0.4025 | 0.162 | 0.1596 |
| | MARS | 0.0134 | 0.0898 | 0.6682 | 0.4465 | 0.4269 |

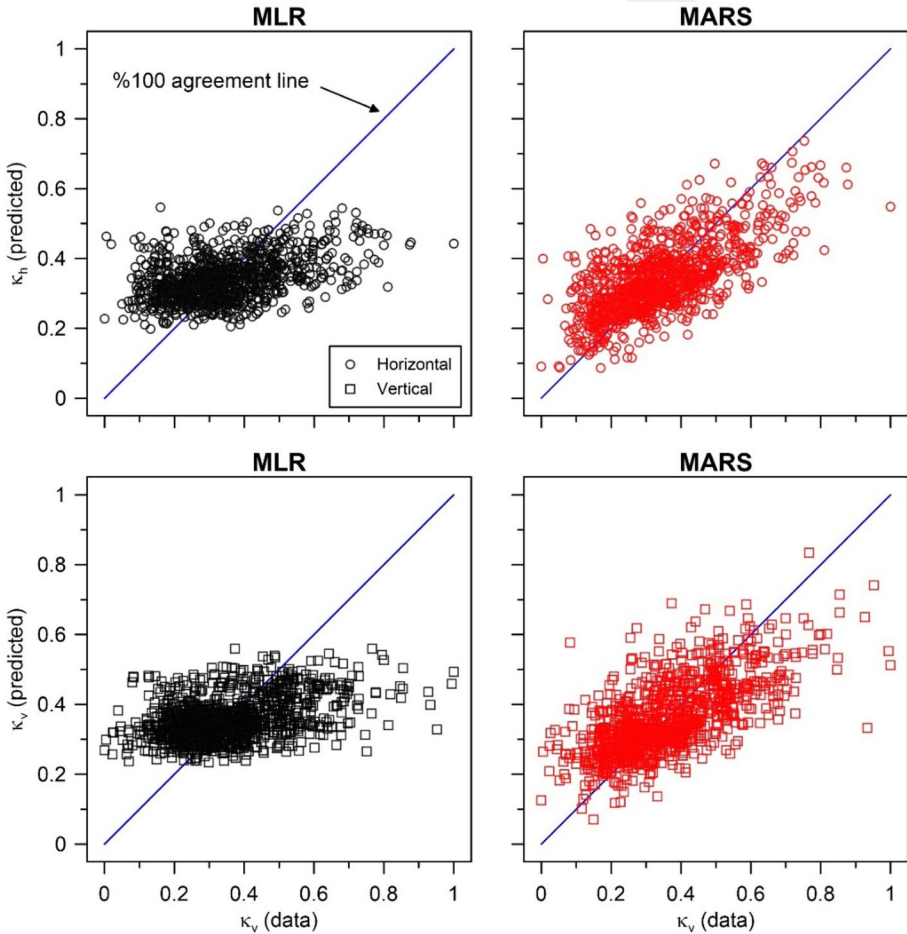


Fig. 8 Comparison of MLR and MARS predictions for A2. All plots show actual κ vs predicted model values. Circles and squares correspond to the horizontal and vertical components, respectively. Blue lines represent the case when predicted κ equals to actual κ

Similar results were obtained for A3. Although the form and size of the datasets are changed, the MARS model performance is much better than that of MLR because it can produce a model suitable for the data structure (see Table 6 and Fig. 9). The BFs in

Table 6 Comparison of performance measures for MLR and MARS in A3

| | | MSE | MAE | r | R ² | Adj.R ² |
|------------|------|--------|--------|--------|----------------|--------------------|
| Horizontal | MLR | 0.0213 | 0.1139 | 0.5779 | 0.334 | 0.3224 |
| | MARS | 0.0147 | 0.0919 | 0.7345 | 0.5395 | 0.5145 |
| Vertical | MLR | 0.0273 | 0.1279 | 0.5377 | 0.2891 | 0.2767 |
| | MARS | 0.0221 | 0.1172 | 0.6514 | 0.4243 | 0.4074 |

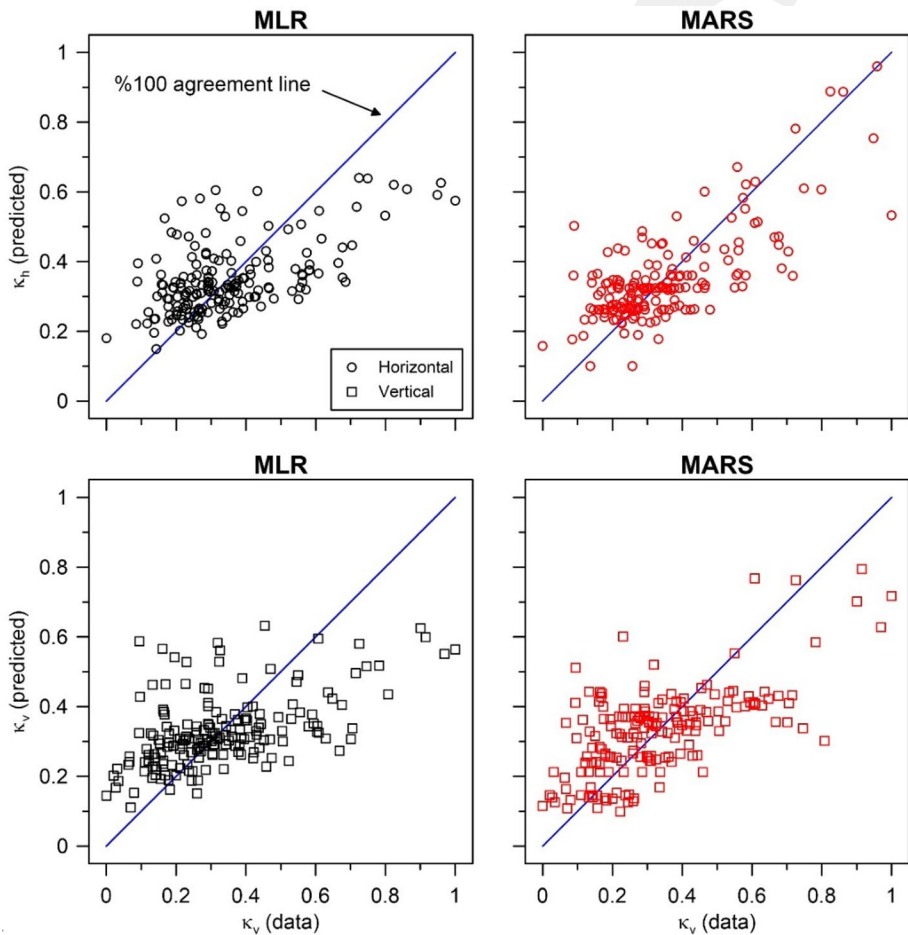


Fig. 9 Comparison of MLR and MARS predictions for A3. All plots show actual κ vs predicted model values. Circles and squares correspond to the horizontal and vertical components, respectively. Blue lines represent the case when predicted κ equals to actual κ

the MARS model use all predictor variables in the following order of importance: R_{epi} , Vs_{30} , and M_w .

Although MLR and MARS are both regression techniques used for prediction tasks in this study, the results show that they have distinct characteristics and prediction

performances. MARS can handle both linear and nonlinear relationships with the help of piecewise linear functions. It also adaptively models different segments of the data, making it more flexible in capturing complex nonlinear patterns. Moreover, MARS automatically captures interactions between variables by creating line segments that allow for varying relationships between different subsets of the data. Thus, MARS is likely to provide better prediction performance because of its ability to adaptively capture such relationships. For all these reasons, in the next part of the application, detailed modeling of the dataset is performed using the MARS approach.

In this part of the application, all of the areas are divided into 3 groups according to the M_w as follows: the first group includes M_w values between 3.5 and 4, the second group includes M_w values between 4 and 5, and the third group includes M_w values between 5 and 6.5. In each area, the number of observations and M_w values are different; thus, the corresponding number of observations differs according to each M_w bin. For example, for A1, the total number of observations is 166. The first group contains 38 observations, accounting for 22.89% of the total observations. The second group includes 89 observations, accounting for 53.61% of the total observations. On the other hand, 39 observations fall in the third group, accounting for 23.50% of the total observations. Table 7 presents the performance outcomes of the MARS models for A1. The findings demonstrate the highest performance values for group $M_w=4-5$, particularly for horizontal κ . However, they exhibit the highest performance values for vertical κ in the $M_w=3.5-4$ group. The data are more abundant within that particular range of M_w . Machine learning algorithms frequently exhibit outstanding performance when used on extensive datasets characterized by irregular structures, such as κ datasets, because of their ability to extract information directly from the data. Therefore, we attribute the algorithm’s achievement in this particular category to the greater volume of data. Overall, the findings suggest that MARS models may attain optimal performance independent of the data structure and M_w . In other words, MARS models tend to be more successful when used with M_w groups that have well-structured data dispersion.

For A2, the total number of observations is 1056, and the first bin contains 310 observations, accounting for 29.36% of the total observations. The second bin includes 614 observations, accounting for 58.14% of the total observations. On the other hand, 132 observations fall in the third bin, accounting for 12.50% of the total observations. In the area with the most data, the MARS model demonstrates superior outcomes for the largest-sized earthquakes when considering horizontal κ . Conversely, its effectiveness is more pronounced for the medium-sized earthquake group when considering vertical κ (see Table 8).

For A3, the total number of observations is 176, and the first bin contains 69 observations, accounting for 39.20% of the total observations. The second bin includes 88 observations that account for 50.00% of the total observations. On the other hand, 19 observations

Table 7 Performance results of the MARS models according to different M_w bins for A1

| Component | M_w range | MSE | MAE | r | R ² | Adj.R ² |
|------------|-------------|--------|--------|--------|----------------|--------------------|
| Horizontal | 3.5–3.9 | 0.0199 | 0.1117 | 0.7236 | 0.5236 | 0.4659 |
| | 4.0–4.9 | 0.0120 | 0.0844 | 0.7965 | 0.6344 | 0.5821 |
| | 5.0–6.5 | 0.0202 | 0.1074 | 0.6681 | 0.4464 | 0.3989 |
| Vertical | 3.5–3.9 | 0.0068 | 0.0683 | 0.9288 | 0.8627 | 0.8362 |
| | 4.0–4.9 | 0.0212 | 0.1161 | 0.7693 | 0.5919 | 0.5566 |
| | 5.0–6.5 | 0.0239 | 0.1137 | 0.6808 | 0.4635 | 0.4336 |

Table 8 Performance results of the MARS models according to different M_w bins for A2

| Component | M_w range | MSE | MAE | r | R^2 | Adj. R^2 |
|------------|-------------|--------|--------|--------|--------|------------|
| Horizontal | 3.5–3.9 | 0.0088 | 0.0744 | 0.6555 | 0.4296 | 0.3859 |
| | 4.0–4.9 | 0.0115 | 0.0832 | 0.6884 | 0.4740 | 0.4402 |
| | 5.0–6.5 | 0.0133 | 0.0941 | 0.7297 | 0.5324 | 0.4853 |
| Vertical | 3.5–3.9 | 0.0105 | 0.0798 | 0.6664 | 0.4442 | 0.4158 |
| | 4.0–4.9 | 0.0119 | 0.0830 | 0.7227 | 0.5223 | 0.4889 |
| | 5.0–6.5 | 0.0138 | 0.0961 | 0.6584 | 0.4335 | 0.3967 |

fall in the third bin, accounting for 10.80% of the total observations. Based on the results in Table 9, it can be observed that most model performances were more significant for the group $M_w=4-5$. However, it is noteworthy that the vertical κ model achieved the most favorable results in the $M_w=4-5$ category for all performance metrics.

Finally, the entire dataset, which includes all areas, is randomly split into train (80% of all data) and test (20% of all data) datasets, and the MARS model is obtained on the train dataset. The best MARS model obtained for the horizontal component is given in Table S7, and the highest degree of interactions between the variables and the maximum number of BFs are defined as 3 and 500, respectively. The best MARS model obtained for the vertical component is given in Table S8, and the highest degree of interactions between the variables and the maximum number of BFs are defined as 2 and 500, respectively. The results in Table 10 indicate that the prediction capability of the MARS models built both for horizontal and vertical κ are excellent for train and test datasets for most of the performance measures, and the train models capture the main structure of the train data well. It is important to keep in mind that the evaluation of model performance is not solely based on performance measures. For this purpose, to show the prediction ability of the MARS model, three different scatter plots (in Figs. 10, 11, 12, 13) that show MARS model performance using (1) the correlation between the predicted and actual response values, (2) the relationship between fitted values and standardized residuals, and (3) the plot of standardized residuals versus observation order are

Table 9 Performance results of the MARS models according to different M_w bins for A3

| Component | M_w range | MSE | MAE | r | R^2 | Adj. R^2 |
|------------|-------------|--------|--------|--------|--------|------------|
| Horizontal | 3.5–3.9 | 0.0122 | 0.0831 | 0.4003 | 0.1602 | 0.1348 |
| | 4.0–4.9 | 0.0176 | 0.0996 | 0.7614 | 0.5797 | 0.5429 |
| | 5.0–6.5 | 0.0223 | 0.1215 | 0.6636 | 0.4404 | 0.4075 |
| Vertical | 3.5–3.9 | 0.0183 | 0.1102 | 0.4878 | 0.2380 | 0.2028 |
| | 4.0–4.9 | 0.0240 | 0.1166 | 0.6991 | 0.4888 | 0.4576 |
| | 5.0–6.5 | 0.0265 | 0.1245 | 0.6229 | 0.3879 | 0.3519 |

Table 10 Train and test data performance results of the MARS models for horizontal and vertical κ responses

| Component | Patterns | MSE | MAE | r | R^2 | Adj. R^2 |
|------------|----------|--------|--------|--------|--------|------------|
| Horizontal | Train | 0.0105 | 0.0779 | 0.7134 | 0.5091 | 0.4850 |
| | Test | 0.0130 | 0.0840 | 0.6247 | 0.6972 | 0.6284 |
| Vertical | Train | 0.0130 | 0.0869 | 0.6654 | 0.4428 | 0.4161 |
| | Test | 0.0173 | 0.0980 | 0.4949 | 0.6313 | 0.5495 |

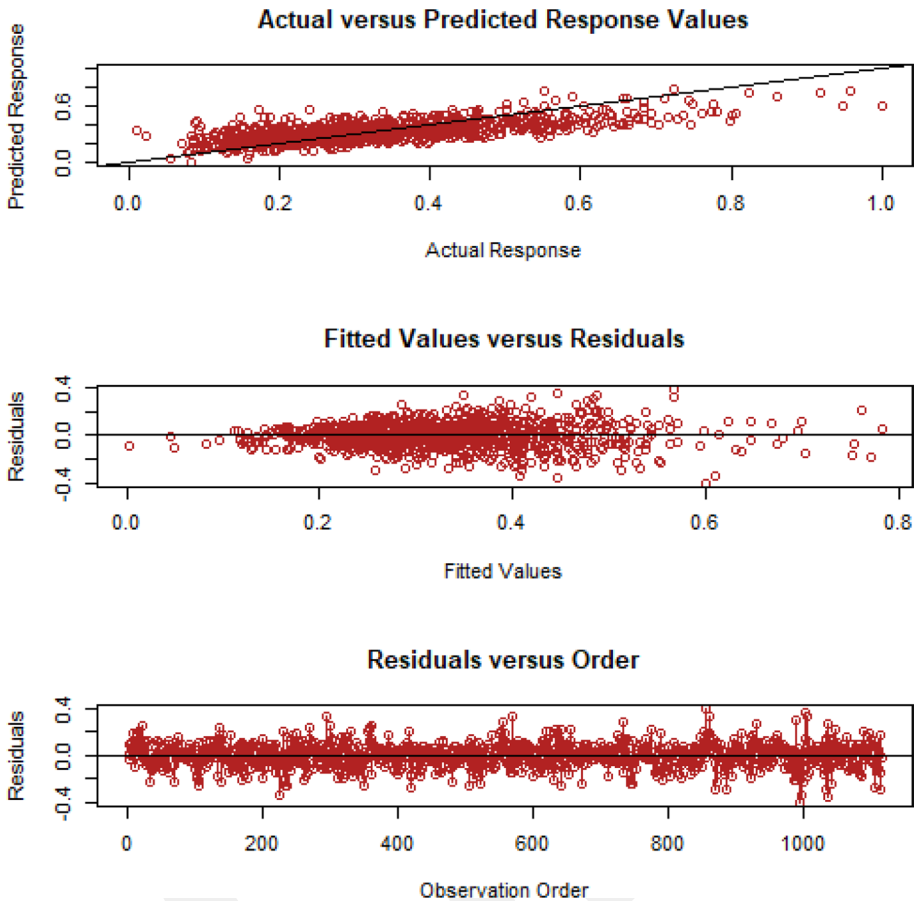


Fig. 10 Scatter plots of the prediction MARS model for the horizontal κ response on the train dataset

obtained both for the train and test datasets. In a scatter plot representing the correlation between predicted and actual response values, the best result is typically achieved when the predicted values align closely with the actual values along a diagonal line. This diagonal line represents a perfect correlation, where the predicted values are exactly equal to the actual values. In the case of modeling horizontal κ , if the points on the scatter plot around the diagonal line are examined, the MARS model's predictions are close to the true values, and there is a good relationship between the predicted and actual responses. Furthermore, the best result in the scatterplot occurs when the standardized residuals are randomly scattered around the horizontal line at 0 (the x-axis). In Figs. 10, 11, 12, 13, the second subplot indicates that the residuals have no systematic patterns or biases. In other words, the model effectively captures the underlying patterns in the data for both the training and testing datasets, and the errors are evenly distributed across the range of fitted values. Finally, a more efficient result is obtained for the scatter plot for standardized residuals versus observation order because the residuals are randomly scattered around the horizontal zero line, indicating a well-behaved and unbiased model that meets the assumptions of error independence. Considering the performance results

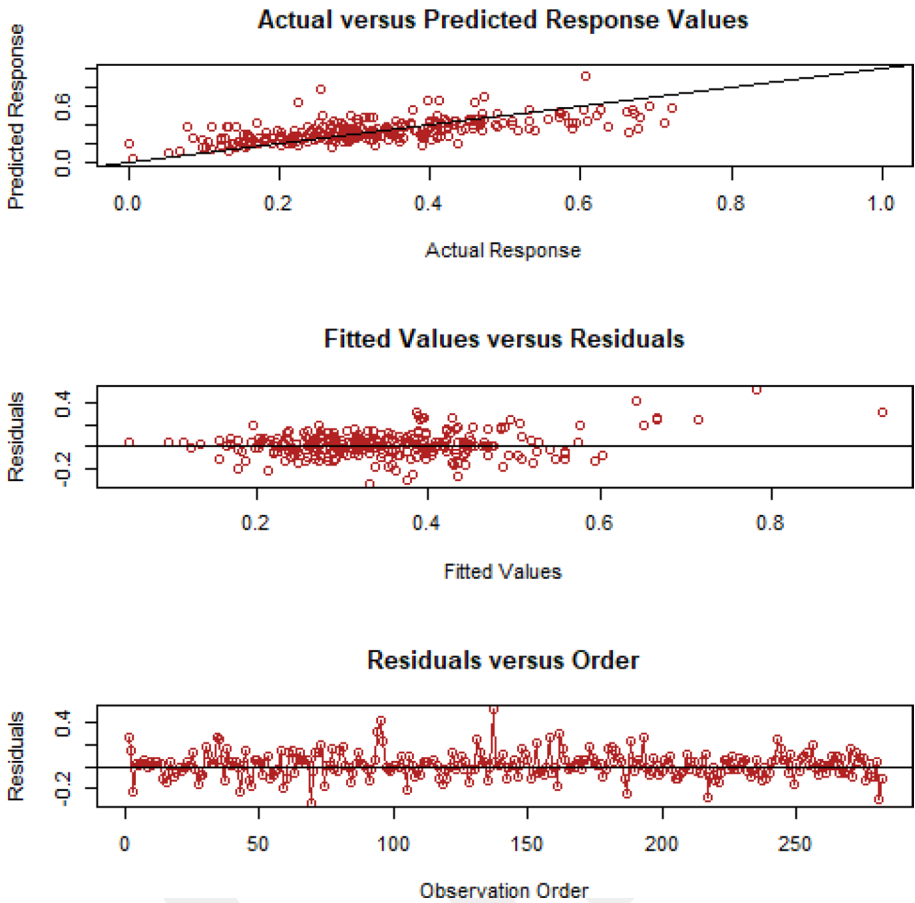


Fig. 11 Scatter plots of the prediction MARS model for the horizontal κ response on the test dataset

of the MARS model obtained for vertical κ , the results obtained from the train data are generally good, but R^2 and Adj. R^2 values obtained for the test data are higher than the train data results. This shows that the MARS model obtained in vertical κ can be used for future prediction.

6 Conclusions

In this study, classic MLR models and MARS-based predictions are developed for the high-frequency spectral attenuation parameter κ . Horizontal and vertical κ models were created for three areas within the Aegean Sea region of Türkiye. The selection of areas is contingent on the tectonic and geological characteristics present within the given area. The predictive capabilities of the methods were tested in terms of selected, well-known performance measures. Based on the numerical results of this study, the following conclusions are drawn:

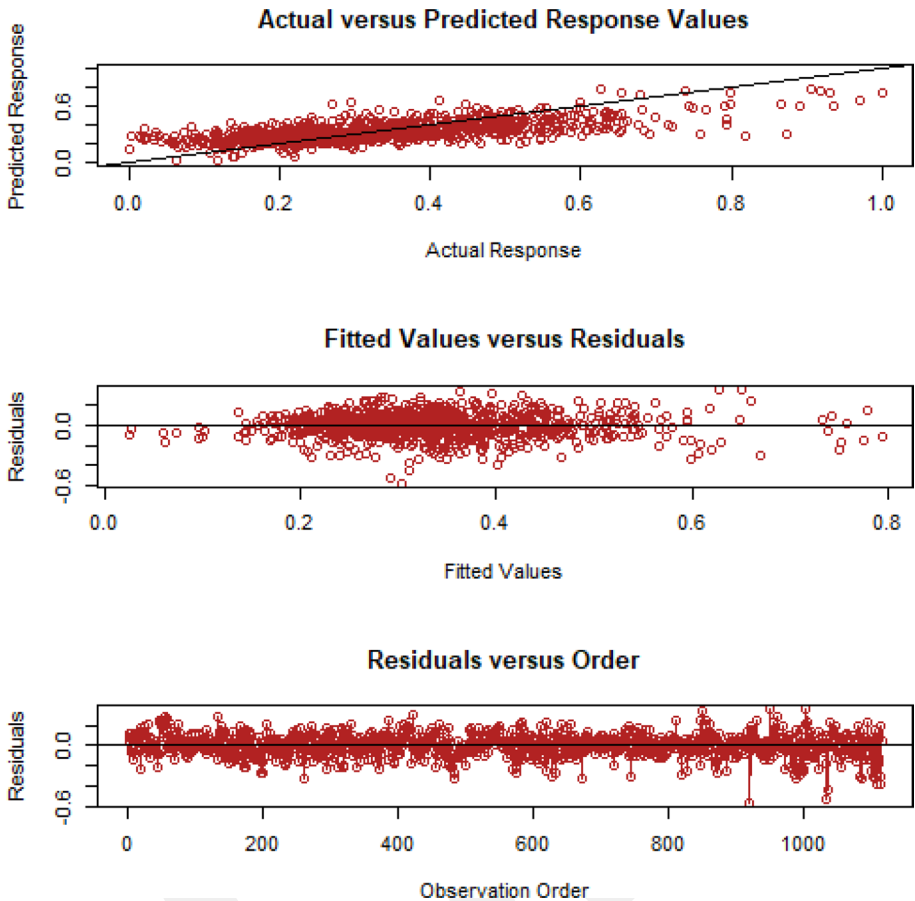


Fig. 12 Scatter plots of the prediction MARS model for the vertical κ response on the train dataset

- In the MLR models, in terms of M_w , R_{epi} , and V_{S30} , the performance measures indicate poor predictive capability regardless of the area.
- The MARS model, which employs piecewise continuous basis functions, demonstrates significantly enhanced predictive performance compared with the MLR model. This improvement can be attributed to the MARS model's capacity to dynamically capture and represent the unique properties inherent in the dataset.
- The performance of the MARS model is shown to be nearly equivalent across all areas, regardless of whether it is applied to horizontal or vertical components.
- The findings indicate that optimal MARS performance may be achieved irrespective of the data structure and earthquake magnitude bins.
- The MARS-based prediction equations developed in this study may be used to perform more precise and reliable κ estimations in the Aegean Sea region. The κ models developed here can be used in future studies concerned with stochastic ground motion simulations, site characterization, and ground motion models in the Aegean Sea region of Türkiye.

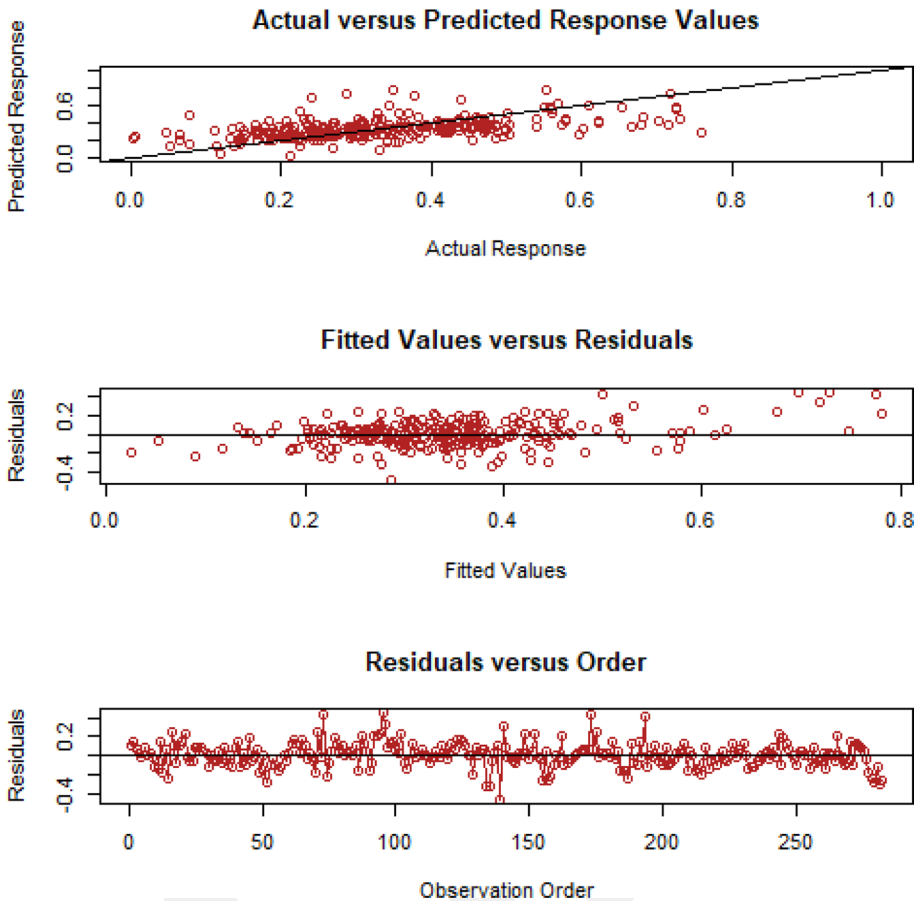


Fig. 13 Scatter plots of the prediction MARS model for the vertical κ response on the test dataset

- Finally, the findings of this study indicate that MARS-based models are promising for capturing the complex characteristics of κ datasets globally.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11069-024-06535-y>.

Acknowledgments We are grateful to Dr. Talip Güngör for providing support on section 2 of this manuscript. The strong-motion data for this study were provided by the Turkish National Strong Ground Motion Observation Network operated by the Disaster and Emergency Management Presidency (AFAD) via <https://tadas.afad.gov.tr>. The multiple linear regression and multivariate adaptive regression splines frameworks have been developed in R programming environment and are not available for sharing. Figs. 1 and 2 were produced using Generic Mapping Tools (GMT) and PyGMT (Wessel et al. 2019; Uieda et al. 2022). Matplotlib (Hunter 2007) and Jupyter notebook programming environment (Kluyver et al. 2016) are used to generate the remaining figures in this manuscript.

Author contributions TÖK, FYÖ, and AA contributed to the study conception and design. Material preparation and data collection were performed by TÖK. Numerical analyses were performed by TÖK and FYÖ.

The first draft of the manuscript was written by TÖK, FYÖ, and AA and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akyol N, Kurtuluş TÖ, Çamyıldız M, Güngör T (2013) Spectral ratio estimates for site effects on the horst-graben system in west Turkey. *Pure Appl Geophys* 170:2107–2125. <https://doi.org/10.1007/s00024-013-0661-2>
- Al Atik AL, Kottke A, Abrahamson N, Hollenback J (2014) Kappa (κ) scaling of ground-motion prediction equations using an inverse random vibration theory approach. *Bull Seismol Soc Am* 104:336–346. <https://doi.org/10.1785/0120120200>
- Altındal A, Askan A (2022) Predictive kappa (κ) models for Turkey: regional effects and uncertainty analysis. *Earthq Spectra* 38:2479–2499. <https://doi.org/10.1177/87552930221116651>
- Anderson JG (1991) A preliminary descriptive model for the distance dependence of the spectral decay parameter in southern California. *Bull Seismol Soc Am* 81:2186–2193
- Anderson JG, Hough SE (1984) A model for the shape of the fourier amplitude spectrum of acceleration at high frequencies. *Bull Seismol Soc Am* 74:1969–1993. <https://doi.org/10.1785/BSSA0740051969>
- Asare EN, Affam M, Ziggah YY (2023) A hybrid intelligent prediction model of autoencoder neural network and multivariate adaptive regression spline for uniaxial compressive strength of rocks. *Model Earth Syst Environ* 9:3579–3595. <https://doi.org/10.1007/s40808-023-01717-2>
- Askan A, Sisman FN, Pekcan O (2014) A regional near-surface high frequency spectral attenuation (kappa) model for northwestern Turkey. *Soil Dyn Earthq Eng* 65:113–125. <https://doi.org/10.1016/j.soildyn.2014.06.007>
- Barbot S, Weiss JR (2021) Connecting subduction, extension and shear localization across the Aegean Sea and Anatolia. *Geophys J Int* 226:422–445. <https://doi.org/10.1093/gji/ggab078>
- Biro Y, Siyahi B, Akbas B (2022) Spectral decay parameter (κ) analysis for a single fault source and an automatic selection procedure for spectral flattening frequency (f_x). *Soil Dyn Earthq Eng* 153:107122. <https://doi.org/10.1016/j.soildyn.2021.107122>
- Bocchini GM, Brüstle A, Becker D et al (2018) Tearing, segmentation, and backstepping of subduction in the Aegean: New insights from seismicity. *Tectonophysics* 734–735:96–118. <https://doi.org/10.1016/j.tecto.2018.04.002>
- Bora SS, Cotton F, Scherbaum F et al (2017) Stochastic source, path and site attenuation parameters and associated variabilities for shallow crustal European earthquakes. *Bull Earthq Eng* 15:4531–4561. <https://doi.org/10.1007/s10518-017-0167-x>
- Bozkurt E (2001) Neotectonics of turkey—a synthesis. *Geodin Acta* 14:3–30. <https://doi.org/10.1080/09853111.2001.11432432>
- Bozkurt E, Mittwede SK (2005) Introduction: evolution of continental extensional tectonics of western Turkey. *Geodin Acta* 18:153–165. <https://doi.org/10.3166/ga.18.153-165>
- Brun JP, Sokoutis D (2007) Kinematics of the southern Rhodope core complex (North Greece). *Int J Earth Sci (geol Rundsch)* 96:1079–1099. <https://doi.org/10.1007/s00531-007-0174-2>
- Brune JN (1970) Tectonic stress and the spectra of seismic shear waves from earthquakes. *J Geophys Res* 75:4997–5009

- Cabas A, Rodriguez-Marek A, Bonilla LF (2017) Estimation of site-specific Kappa (κ)-consistent damping Values at KiK-Net sites to assess the discrepancy between laboratory-based damping models and observed attenuation (of seismic waves) in the field. *Bull Seismol Soc Am* 107:2258–2271. <https://doi.org/10.1785/0120160370>
- Castro RR, Colavitti L, Vidales-Basurto CA et al (2022) Near-source attenuation and spatial variability of the spectral decay parameter kappa in central Italy. *Seismol Res Lett* 93:2299–2310. <https://doi.org/10.1785/0220210276>
- Çiftçi NB, Temel RÖ, Iztan YH (2010) Hydrocarbon occurrences in the western Anatolian (Aegean) grabens, Turkey: is there a working petroleum system? *Am Assoc Pet Geol Bull* 94:1827–1857. <https://doi.org/10.1306/06301009172>
- Conoscenti C, Ciaccio M, Caraballo-Arias NA et al (2015) Assessment of susceptibility to earth-flow landslide using logistic regression and multivariate adaptive regression splines: A case of the Belice River basin (western Sicily, Italy). *Geomorphology* 242:49–64. <https://doi.org/10.1016/j.geomorph.2014.09.020>
- Deng ZP, Pan M, Niu JT et al (2021) Slope reliability analysis in spatially variable soils using sliced inverse regression-based multivariate adaptive regression spline. *Bull Eng Geol Environ* 80:7213–7226. <https://doi.org/10.1007/s10064-021-02353-9>
- Dewey JF, Hempton MR, Kidd WSF et al (1986) Shortening of continental lithosphere: the neotectonics of Eastern Anatolia — a young collision zone. *Geol Soc Spec Publ* 19:1–36. <https://doi.org/10.1144/GSL.SP.1986.019.01.01>
- Douglas J, Gehl P, Bonilla LF, Gelis C (2010) A κ model for mainland France. *Pure Appl Geophys* 167:1303–1315. <https://doi.org/10.1007/s00024-010-0146-5>
- Drouet S, Cotton F, Guéguen P (2010) VS30, κ , regional attenuation and Mw from accelerograms: application to magnitude 3–5 French earthquakes. *Geophys J Int* 182:880–898. <https://doi.org/10.1111/j.1365-246X.2010.04626.x>
- Duman TY, Çan T, Emre Ö et al (2018) Seismotectonic database of Turkey. *Bull Earthq Eng* 16:3277–3316. <https://doi.org/10.1007/s10518-016-9965-9>
- Feliciísimo AM, Cuartero A, Remondo J, Quirós E (2013) Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study. *Landslides* 10:175–189. <https://doi.org/10.1007/s10346-012-0320-1>
- Fernández AI, Castro RR, Huerta CI (2010) The spectral decay parameter kappa in northeastern Sonora, Mexico. *Bull Seismol Soc Am* 100:196–206. <https://doi.org/10.1785/0120090049>
- Friedman JH (1991) Multivariate adaptive regression splines. *Ann Stat* 19:1–67
- Gičev V, Trifunac MD (2022) High-frequency decay of Fourier spectra of strong motion acceleration and nonlinear site response. *Earthq Eng Resil* 1:302–316. <https://doi.org/10.1002/eer.2.28>
- Hai T, Theruvil Sayed B, Majdi A et al (2023) An integrated GIS-based multivariate adaptive regression splines-cat swarm optimization for improving the accuracy of wildfire susceptibility mapping. *Geocarto Int*. <https://doi.org/10.1080/10106049.2023.2167005>
- Hong L, Wang X, Zhang W et al (2024) System reliability-based robust design of deep foundation pit considering multiple failure modes. *Geosci Front* 15:101761. <https://doi.org/10.1016/j.gsf.2023.101761>
- Hunter JD (2007) Matplotlib: a 2D graphics environment. *Comput Sci Eng* 9:90–95
- Iwakiri K, Hoshiba M (2012) High-frequency (10 Hz) content of the initial fifty seconds of waveforms from the 2011 off the Pacific coast of Tohoku earthquake. *Bull Seismol Soc Am* 102:2232–2238. <https://doi.org/10.1785/0120110241>
- Jolivet L, Faccenna C, Huet B et al (2013) Aegean tectonics: strain localisation, slab tearing and trench retreat. *Tectonophysics* 597–598:1–33. <https://doi.org/10.1016/j.tecto.2012.06.011>
- Kadıroğlu FT, Kartal RF (2016) The new empirical magnitude conversion relations using an improved earthquake catalogue for Turkey and its near vicinity (1900–2012). *Turkish J Earth Sci* 25:300–310. <https://doi.org/10.3906/yer-1511-7>
- Kiratzi A, Papazachos C, Özacar A et al (2021) Characteristics of the 2020 Samos earthquake (Aegean Sea) using seismic data. *Bull Earthq Eng* 20:7713–7735. <https://doi.org/10.1007/s10518-021-01239-1>
- Kishida T, Kayen RE, Ktenidou O-J, et al (2014) PEER Arizona strong-motion database and GMPEs evaluation
- Kluyver T, Ragan-Kelley B, Pérez F, et al (2016) Jupyter Notebooks—a publishing format for reproducible computational workflows. In: *Position Power Acad Publ Play Agents Agendas - Proc 20th Int Conf Electron Publ ELPUB 2016*, pp 87–90. <https://doi.org/10.3233/978-1-61499-649-1-87>
- Ktenidou OJ, Gélis C, Bonilla LF (2013) A study on the variability of Kappa (κ) in a borehole: implications of the computation process. *Bull Seismol Soc Am* 103:1048–1068. <https://doi.org/10.1785/0120120093>

- Ktenidou O-J, Cotton F, Abrahamson NA, Anderson JG (2014) Taxonomy of κ : a review of definitions and estimation approaches targeted to applications. *Seismol Res Lett* 85:135–146. <https://doi.org/10.1785/0220130027>
- Ktenidou OJ, Abrahamson NA, Drouet S, Cotton F (2015) Understanding the physics of kappa (κ): Insights from a downhole array. *Geophys J Int* 203:678–691. <https://doi.org/10.1093/gji/ggv315>
- Ktenidou OJ, Silva WJ, Darragh RB et al (2017) Squeezing kappa (κ) out of the transportable array: A strategy for using bandlimited data in regions of sparse seismicity. *Bull Seismol Soc Am* 107:256–275. <https://doi.org/10.1785/0120150301>
- Kurtulmus TO, Akyol N (2015) Separation of source, site and near-surface attenuation effects in western Turkey. *Nat Hazards* 77:1515–1532. <https://doi.org/10.1007/s11069-015-1660-7>
- Lanzano G, Felicetta C, Paor F et al (2022) Generic-to-reference rock scaling factors for seismic ground motion in Italy. *Bull Seismol Soc Am* 112:1583–1606. <https://doi.org/10.1785/0120210063>
- Le Pichon X, Kreemer C (2010) The miocene-to-present kinematic evolution of the eastern mediterranean and middle east and its implications for dynamics. *Annu Rev Earth Planet Sci* 38:323–351. <https://doi.org/10.1146/annurev-earth-040809-152419>
- Motazedian D (2006) Region-specific key seismic parameters for earthquakes in northern Iran. *Bull Seismol Soc Am* 96:1383–1395. <https://doi.org/10.1785/0120050162>
- Motazedian D, Atkinson GM (2005) Stochastic finite-fault modeling based on a dynamic corner frequency. *Bull Seismol Soc Am* 95:995–1010. <https://doi.org/10.1785/0120030207>
- Neter J, Kutner MH, Nachtsheim CJ, Wasserman W (1996) *Applied linear statistical models*, 4th Editio. WCB McGraw-Hill, Boston
- Özkaptan M, Gülyüz E, Uzel B et al (2021) Deformation in SW Anatolia (Turkey) documented by anisotropy of magnetic susceptibility data. *Tectonics* 40:1–26. <https://doi.org/10.1029/2021TC006882>
- Pamuk E, Özdağ ÖC, Tunçel A et al (2018) Local site effects evaluation for Aliğağ/Izmir using HVSR (Nakamura technique) and MASW methods. *Nat Hazards* 90:887–899. <https://doi.org/10.1007/s11069-017-3077-y>
- Pamuk E, Özdağ ÖC, Akgün M (2019) Soil characterization of Bornova Plain (Izmir, Turkey) and its surroundings using a combined survey of MASW and ReMi methods and Nakamura's (HVSR) technique. *Bull Eng Geol Environ* 78:3023–3035. <https://doi.org/10.1007/s10064-018-1293-7>
- Parolai S, Bindl D (2004) Influence of Soil Properties on κ evaluation. *Bull Seismol Soc Am* 94:349–356. <https://doi.org/10.1785/0120030022>
- Petukhin A, Irikura K (2000) A method for the separation of source and site effects and the apparent Q structure from strong motion data. *Geophys Res Lett* 27:3429–3432
- Purvance MD, Anderson JG (2003) A comprehensive study of the observed spectral decay in strong-motion accelerations recorded in Guerrero, Mexico. *Bull Seismol Soc Am* 93:600–611. <https://doi.org/10.1785/0120020065>
- Samui P (2013) Multivariate adaptive regression spline (Mars) for prediction of elastic modulus of jointed rock mass. *Geotech Geol Eng* 31:249–253. <https://doi.org/10.1007/s10706-012-9584-4>
- Sandikkaya MA, Yilmaz MT, Bakir BS, Yilmaz Ö (2010) Site classification of Turkish national strong-motion stations. *J Seismol* 14:543–563. <https://doi.org/10.1007/s10950-009-9182-y>
- Şengör AMC, Görür N, Şaroğlu F (1985) Strike-Slip Faulting and Related Basin Formation in Zones of Tectonic Escape: Turkey as a Case Study. In: Biddle KT, Christie-Blick N (eds) *Strike-Slip Deformation, Basin Formation, and Sedimentation*. SEPM Society for Sedimentary Geology, pp 227–264
- Şengör AMC, Yazıcı M (2020) The aetiology of the neotectonic evolution of Turkey. *Med Geosc Rev* 2:327–339. <https://doi.org/10.1007/s42990-020-00039-0>
- Sertcelik F, Akçay D, Livaoglu H, Gerdan S (2022) The spectral decay parameter kappa in marmara region, Turkey. *Arab J Geosci*. <https://doi.org/10.1007/s12517-021-09308-0>
- Sotiriadis D, Margaris B, Klimis N, Sextos A (2021) Implications of high-frequency decay parameter, “ κ -kappa”, in the estimation of kinematic soil-structure interaction effects. *Soil Dyn Earthq Eng* 144:106665. <https://doi.org/10.1016/j.soildyn.2021.106665>
- Tanırcaan G, Dikmen SÜ (2018) Variation of high frequency spectral attenuation (Kappa) in vertical arrays. *Soil Dyn Earthq Eng* 113:406–414. <https://doi.org/10.1016/j.soildyn.2018.06.016>
- Tsai CCP, Chen KC (2000) A model for the high-cut process of strong-motion accelerations in terms of distance, magnitude, and site condition: an example from the Smart 1 array, Lotung. *Taiwan Bull Seismol Soc Am* 90:1535–1542. <https://doi.org/10.1785/0120000010>
- Tsurugi M, Tanaka R, Kagawa T, Irikura K (2020) High-frequency spectral decay characteristics of seismic records of inland crustal earthquakes in Japan : evaluation of the f max and κ models. *Bull Seismol Soc Am* 110:452–470. <https://doi.org/10.1785/0120180342>
- Uieda L, Tian D, Leong WJ, et al (2022) PyGMT: A Python interface for the Generic Mapping Tools. <https://doi.org/10.5281/ZENODO.6426493>

- Vaheddoost B, Safari MJS, Yilmaz MU (2023) Rainfall-runoff simulation in ungauged tributary streams using drainage area ratio-based multivariate adaptive regression spline and random forest hybrid models. *Pure Appl Geophys* 180:365–382. <https://doi.org/10.1007/s00024-022-03209-3>
- Vamvakaris DA, Papazachos CB, Papaioannou CA et al (2016) A detailed seismic zonation model for shallow earthquakes in the broader Aegean area. *Nat Hazards Earth Syst Sci* 16:55–84. <https://doi.org/10.5194/nhess-16-55-2016>
- van Houtte C, Drouet S, Cotton F (2011) Analysis of the origins of κ (kappa) to compute hard rock to rock adjustment factors for GMPEs. *Bull Seismol Soc Am* 101:2926–2941. <https://doi.org/10.1785/01201100345>
- van Houtte C, Ktenidou OJ, Larkin T, Holden C (2014) Hard-site κ (kappa) calculations for Christchurch, New Zealand, and comparison with local ground-motion prediction models. *Bull Seismol Soc Am* 104:1899–1913. <https://doi.org/10.1785/0120130271>
- Wessel P, Luis JF, Uieda L et al (2019) The generic mapping tools version 6. *Geochem Geophys Geosyst* 20:5556–5564. <https://doi.org/10.1029/2019GC008515>
- Zhang W, Goh ATC, Zhang Y et al (2015) Assessment of soil liquefaction based on capacity energy concept and multivariate adaptive regression splines. *Eng Geol* 188:29–37. <https://doi.org/10.1016/j.enggeo.2015.01.009>
- Zhang W, Zhang Y, Goh ATC (2017) Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation. *Tunn Undergr Space Technol* 64:24–33. <https://doi.org/10.1016/j.tust.2017.01.009>
- Zhang WG, Li HR, Wu CZ et al (2021) Soft computing approach for prediction of surface settlement induced by earth pressure balance shield tunneling. *Undergr Sp* 6:353–363. <https://doi.org/10.1016/j.undsp.2019.12.003>
- Zhang W, Gu X, Tang L et al (2022) Application of machine learning, deep learning and optimization algorithms in geoenvironment and geoscience: comprehensive review and future challenge. *Gondwana Res* 109:1–17. <https://doi.org/10.1016/j.gr.2022.03.015>
- Zhang W (2020) MARS applications in geotechnical engineering systems: multi-dimension with big data. Springer, Berlin
- Zheng G, Zhang W, Zhou H, Yang P (2020) Multivariate adaptive regression splines model for prediction of the liquefaction-induced settlement of shallow foundations. *Soil Dyn Earthq Eng* 132:106097. <https://doi.org/10.1016/j.soildyn.2020.106097>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Tevfik Özgür Kurtulmuş¹  · Fatma Yerlikaya-Özkurt²  · Aysegul Askan³ 

✉ Tevfik Özgür Kurtulmuş
ozgur.kurtulmus@deu.edu.tr

Fatma Yerlikaya-Özkurt
fatmayerlikaya@gmail.com

Aysegul Askan
aaskan@metu.edu.tr

¹ Department of Geophysical Engineering, Dokuz Eylül University, İzmir 35390, Türkiye

² Department of Industrial Engineering, Atılım University, 06830 Ankara, Türkiye

³ Department of Civil Engineering, Middle East Technical University, 06800 Ankara, Türkiye