CrossMark

Partially non-ergodic region specific GMPE for Europe and Middle-East

Sreeram Reddy Kotha[1] · Dino Bindi[1] · Fabrice Cotton[1]

Abstract The ergodic assumption considers the time sampling of ground shaking generated in a given region by successive earthquakes as equivalent to a spatial sampling of observed ground motion across different regions. In such cases the estimated aleatory variability in source, propagation, and site seismic processes in ground motion prediction equations (GMPEs) is usually larger than with a non-ergodic approach. With the recently published datasets such as RESORCE for Europe and Middle-East regions, and exploiting algorithms like the non-linear mixed effects regression it became possible to introduce statistically well-constrained regional adjustments to a GMPE, thus 'partially' mitigating the impact of the assumption on regional ergodicity. In this study, we quantify the regional differences in the apparent attenuation of high frequency ground motion with distance and in linear site amplification with V_{s30}, between Italy, Turkey, and rest of the Europe-Middle-East region. With respect to a GMPE without regional adjustments, we obtain up to 10 % reduction in the aleatory variability σ, primarily contributed by a 20 % reduction in the between-station variability. The reduced aleatory variability is translated into an epistemic uncertainty, i.e. a standard error on the regional adjustments which can be accounted for in the hazard assessment through logic-tree branches properly weighted. Furthermore, the between-event variability is reduced by up to 30 % by disregarding in regression the events with empirically estimated moment magnitude. Therefore, we conclude that a further refinement of the aleatory variability could be achieved by choosing a combination of proxies for the site response, and through the homogenization of the magnitude scales across regions.

✉ Sreeram Reddy Kotha
  sreeram@gfz-potsdam.de

[1] GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany

1 Introduction

Reliability of the ground motion predicted by empirical models mostly depends on the characteristics of the underlying calibration dataset. In the framework of seismic hazard assessment, the motivation behind compilation of a large strong motion dataset which includes recordings from different regions is twofold: first, to improve the magnitude-distance data distribution, and sampling different source characteristics and site conditions; second, to allow the calibration of models complex enough to describe the main physical processes contributing to the variability of the ground motion. The current practice for computing the seismic hazard is based on an ergodic assumption, where the aleatory variability, i.e. the standard deviation sigma (σ) of ground motion prediction equation (GMPE), includes the regional differences in ground motion. If on one hand the ergodic assumption allows to replace the time sampling of ground shaking generated in a given region by successive earthquakes with a spatial sampling of ground shaking observed across different regions, on the other hand it increases the aleatory variability associated with source, propagation, and site seismic processes. Allowing regional differences in the GMPE 'partially' removes this ergodicity by translating the aleatory variability into epistemic uncertainty which, in statistical sense, is the modelling uncertainty in region-specific adjustments.

The collection of data from different regions with similar tectonic features (e.g. shallow crustal active regions, stable continental regions, etc.) was performed in the past under the assumption that the trans-regional and between-country variability of the ground motion was either negligible or otherwise difficult to model due to the limitation in the sampling properties of the compiled datasets (e.g. Douglas 2004a, b). As an example, the NGA-West models (Abrahamson and Silva 2008) were derived from a dataset including recordings from multiple regions (mainly California, Taiwan, Japan) without modelling the regional effects. Later studies on the applicability of the NGA models to Europe (e.g. Stafford et al. 2008) highlighted the general agreement between predicted median values and the observations. The main difference was a faster distance attenuation observed in European data with respect to California; in agreement with previous findings (Douglas 2004a). Moreover a detailed comparison between the NGA models and strong motion data recorded in Italy (Scasserra et al. 2009) confirmed that it was possible to improve the predictive performance of NGA models for Europe by applying regional corrections to the attenuation with distance terms and to the overall scaling parameters (offset and pseudo-depth).

Extension of the NGA database into NGA-West2 (Ancheta et al. 2014) with introduction of several small magnitude events mainly from California, and moderate to large size earthquakes from other regions of the world, promoted the interest in evaluating regional effects in the ground motions. As a consequence, the most recent GMPEs developed from NGA-West2 include correction terms accounting for regional effects. Many authors (e.g. Boore et al. 2014; Chiou and Youngs 2014) introduced regional differences in the anelastic attenuation coefficient and the site term related to depth of basin. Regional differences in the V_{s30} scaling were also considered (e.g. Abrahamson et al. 2014), while information available in the dataset is not enough to constrain correction factors for other parameters.

RESORCE strong motion database (Akkar et al. 2014a) was compiled with recordings from different European and Middle-East countries, and was used to derive several GMPEs (Douglas et al. 2014). While these models do not account for regional differences in

ground-motion scaling, recent studies highlighted the presence of regional effects either between selected countries (e.g. between Turkey and Iran by Kale et al. 2015), or among different tectonic regions in Europe (Gianniotis et al. 2014). Ignoring the regional differences in ground motion scaling may result in an inflated residual standard deviation, and correction for regional bias in the median ground motion can be a first step towards 'partially non-ergodic' region-specific PSHA. With such a goal in mind, this study focuses on identification of systematic regional differences in ground motion scaling in Europe. Following the previous efforts of developing GMPE using RESORCE dataset (Douglas et al. 2014 and reference therein), we derive a new GMPE based on a relatively simple functional form which will still be able to capture the main features of ground motion-scaling (Bindi et al. 2014). However, unlike in previous studies, a non-linear mixed effect regression (NLMER by Bates et al. 2014) approach is applied where the regional differences are estimated as random effects applied to different model parameters. The advantages of using NLMER in place of the traditional random effect algorithm by Abrahamson and Youngs (1992) are discussed by Stafford (2014). For example, group specific adjustments can be estimated for any of the regression coefficients in a statistically correct way making NLMER much more extendable than traditional approaches. We identify the statistically significant random effects and the regional adjustments for relevant parameters are provided as final result.

2 Dataset and selection criteria

The most recent Pan-European GMPEs (Douglas et al. 2014) are based on the RESORCE strong motion dataset (http://www.resorce-portal.eu/). RESORCE extends the previous pan-European strong motion dataset (Ambraseys et al. 2004) with recently compiled Greek, Italian, Swiss and Turkish accelerometric archives (Akkar et al. 2014a). In this study, starting from the 2013 release of RESORCE, we performed a preliminary data selection to exclude the poor quality or unprocessed records, or those records lacking the three components of ground-motion; then, we applied the following criteria to select the input data for regression:

- Given the recent interest in considering small magnitude earthquakes for assessing the hazard in several regions of Europe (http://projet-sigma.com/ScientificObjectives.html), records from events with moment magnitudes larger than or equal to 4 are considered.
- Only focal depths shallower than 35 km, and distances (Joyner–Boore, R_{JB}, or epicentral R_{epi}) shorter than 300 km are selected. The epicentral distance, R_{epi} is used to approximate R_{JB} when the latter is unspecified, but only when $M \leq 5$ and $R_{epi} \geq 10$ km. For larger magnitudes and smaller epicentral distances, records without R_{JB} are disregarded.
- For each oscillator period T, only those recording filtered with high pass corner frequency (f_{hp}) smaller than or equal to $1/(1.25\ T)$, i.e. $f_{hp} \leq 0.8\ f_{oscillator}$ (Abrahamson and Silva 1997). For example, for T = 1 s ($f_{oscillator}$ = 1 Hz), we considered only recordings with f_{hp} smaller than or equal to 0.8 Hz; for T = 4 s ($f_{oscillator}$ = 0.25 Hz), we chose $f_{hp} \leq 0.2$ Hz. Single recorded earthquakes are not selected.
- We consider only recordings from sites with known or inferred V_{s30}.

In RESORCE, the moment magnitude is provided either as directly computed (e.g. from the moment tensor solutions), or converted from other magnitude scales (e.g. local

magnitude or surface wave magnitude) using country-based empirical regressions (see Akkar et al. 2014a for details). Earthquakes with M_w derived through empirical regressions are not considered in this study.

Considering the unbalanced composition of the dataset, we categorize the contributing regions into three groups: Italy, Turkey, and Others, where the latter collects data from all the countries contributing to RESORCE with less than 200 selected records. Although a regionalization based on the tectonic settings (e.g. Delavaud et al. 2012) could be more appropriate to explore regional differences in ground motion, we opt for a country-based categorization that reflects the structure followed for data compilation. The filtered dataset is composed of 1251 recordings, with 659 recordings from Turkey (TR), 378 from Italy (IT), 214 in Others group; primarily contributed to by Greece, Montenegro, Iran, and France.

In terms of magnitude range, distance range, and site characterization, the dataset is unbalanced among the regions. Douglas (2007) showed that the predicted median ground motions are not well-constrained away from the centroid of data, especially for sparse datasets. Figure 1 shows the magnitude—distance distribution of recordings in our dataset, categorized according to different regions and soil classes. For example, there are very few recordings from Turkey in site class A (rock with $V_{s30} > 800$ m/s), which means that when a GMPE is derived from the compendium dataset without regional distinction, the estimated site response for class A could be controlled by contributions from Italy and Other regions, even though the class A rock response in Turkey could be significantly different. Similarly, for distances larger than 100 km and empirical site response of class B (stiff soil
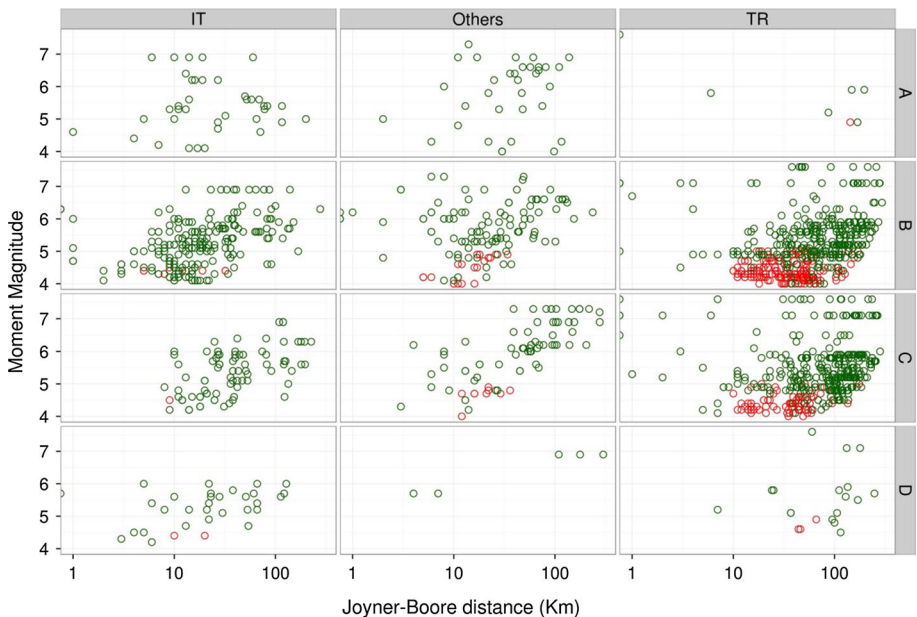
Fig. 1 *Scatter plot* showing the distribution of observed data in Magnitude - Distance ranges for different EC8 site classes for each region Italy (IT), Others, and Turkey (TR). The *red markers* correspond to events without a computed moment magnitude but only an empirically estimated/converted moment magnitude, and consequently excluded from the regression. *Green markers* show the final distribution of records that are used for PGA regression

with 800 m/s > $V_{s30} \geq$ 360 m/s) and Class C (soft soil with 360 m/s > $V_{s30} \geq$ 180 m/s) the predictions could be controlled by strong motion recordings from Turkey. Moreover, preliminary non-parametric analysis (here not shown) suggest that the average slope of distance scaling is different among the regions, hinting for possible regional differences in the distance scaling of high-frequency ground motions, which we could quantify as a regional variation during the GMPE regression. Based on these evidences, in the following we seek for ground motion regional variations related to the scaling with distance and to the site response.

3 Regression approach

Different models were derived from RESORCE dataset performing either a parametric regression (e.g. Akkar et al. 2014b; Bindi et al. 2014) or following non-parametric approaches (e.g. Derras et al. 2012; Hermkes et al. 2014). The parametric regression approaches were applied using the random effects methodology of Abrahamson and Youngs (1992), where the residuals are split into between-event (δB_e), and within-event (δW_{es}) residuals. The GMPE functional forms used were relatively simple with respect to those implemented within the NGA-West2 project (e.g. Abrahamson et al. 2014), reflecting the detail of information available in the RESORCE metadata. With the aim of investigating the presence of regional effects in ground motion variability, we also follow a parametric regression approach but using a non-linear mixed effect approach (NLMER, e.g. Bates et al. 2014). Following Bindi et al. (2014), we consider the following functional form:

$$\ln(GM) = e_1 + F_D(R,M) + F_M(M) + \delta B_e + \delta B_s + \varepsilon \qquad (1)$$

$$F_D(R,M) = \left[c_1 + c_2(M - M_{ref})\right]\ln\left(\frac{\sqrt{R^2 + h^2}}{R_{ref}}\right) + (c_3 + \Delta c_{3,r})\left(\sqrt{R^2 + h^2} - R_{ref}\right) \quad (2)$$

$$F_M(M) = \begin{cases} b_1(M - M_h) + b_2(M - M_h)^2 & \text{for } M < M_h, \text{ where } M_h = 6.75 \\ b_3(M - M_h) & \text{for } M \geq M_h \end{cases} \quad (3)$$

In Eq. (1), e_1 is the global off-set parameter; F_D, and F_M are the distance and magnitude scaling components as defined in Eqs. (2) and (3), respectively; δB_e and δB_s are random effects on e_1 describing the between-event and between-station variability, respectively (Stafford 2014; Al Atik et al. 2010); ε is the residual distribution accounting for the aleatory variability. In the following, the standard deviation of the between-event and residual distributions are indicated with the symbols τ and $\varnothing_0$, respectively. The hinge magnitude M_h is fixed at 6.75 and the parameter b_3, which controls the saturation with magnitude, is not constrained to be positive (i.e. the over-saturation at magnitudes >6.75 is allowed). As in Bindi et al. (2014), the reference moment magnitude M_{ref} and reference Joyner-Boore distance R_{ref} are set at M5.5 and 1 km, respectively.

The major contributor to 'Others' group in terms of recordings is Greece (137), followed by Montenegro (35), and Iran (20). We performed several preliminary regressions considering different number of geographical categories, including attempts of isolating the Greek recordings from Others. In order to get reliable regional adjustments for the anelastic attenuation, a minimum number of recording per category (i.e. per country) was needed. Since the adjustment factor for Greece, once isolated from Others, was not

significantly different from zero at 95 % CI, we kept the Greece recordings inside the Others category.

It is worth noting that we only introduced a regional adjustment factor for the apparent anelastic attenuation coefficient (i.e. c_3 in Eq. 1), but the magnitude scaling component (F_M in Eq. 1) is constrained by the data from all regions. When asked for a random-effect on a regression parameter (e.g. regional adjustment to c_3 in Eq. 1) for each level in the group (levels being Italy, Turkey, and Others), the NLME algorithm estimates scalar additive adjustments which follow a standard-normal distribution. Therefore, the GMPE regression-coefficient c_3 without any regional-adjustments (i.e. without adding $\Delta c_{3,r}$ to c_3), is a generic anelastic attenuation coefficient without a regional bias.

3.1 Regional variability in apparent anelastic attenuation term

In Eq. (2), we introduce a country-based random effect $\Delta c_{3,r}$ on parameter c_3, where r represents the three selected regions, i.e. r = IT, Others, TR. Coefficients c_1, c_2 and c_3 in the scaling with distance F_D, correspond to the geometrical spreading, magnitude-dependent geometrical spreading, and apparent anelastic attenuation, respectively, although these names should be strictly used only for a model based on Fourier spectral amplitudes. Coefficient c_3 is constrained to being less than or equal to 0 for all spectral periods to disallow oversaturation at longer distances as in Bindi et al. (2014). Preliminary trials showed that for long periods (>1 s), c_3 is taking a positive value and has a large negative correlation with c_1, and a positive correlation with e_1. Since a Student's *t* test confirms that it is anyway losing its significance, c_3 and the associated regional variations are fixed at zero for periods longer than 1 s.
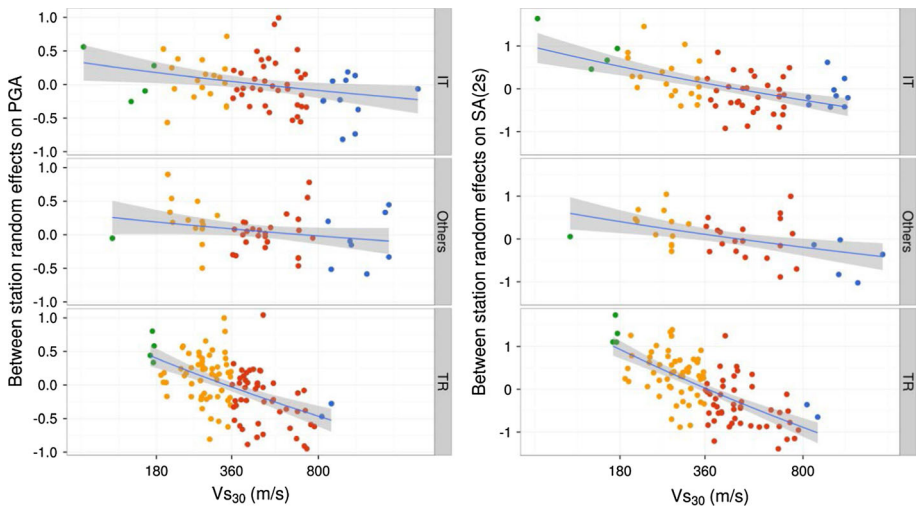
Fig. 2 Between station residuals (PGA in *left panel*, SA (2 s) in *right panel*) plotted against V_{s30} (m/s) with stations separated into regions. The *blue line* is a regression fit of residuals as a function $\ln(V_{s30})$. The *grey ribbon* shows the standard error on regression fit. Difference in slope of the regression fit shows regional difference in linear site-amplification (g_2), difference in x-intercept shows the regional difference in reference V_{s30} (V_{ref})

Table 1 Coefficient table for GMPE

t	e_1	b_1	b_2	b_3	c_1	c_2	c_3	h	τ	$\varnothing_0$	σ	$\Delta c_{3,IT}$	$\Delta c_{3,Others}$	$\Delta c_{3,TR}$	$SE(\Delta c_{3,IT})$	$SE (\Delta c_{3,Others})$	$SE (\Delta c_{3,TR})$
pgv	0.773	0.483	−0.101	−0.021	−1.198	0.229	−0.00008	5.845	0.349	0.496	0.683	−0.00189	0.00142	0.00050	0.00077	0.00074	0.00035
pga	2.982	−0.363	−0.195	−0.406	−1.231	0.272	−0.00395	6.390	0.350	0.451	0.657	−0.00326	0.00326	0.00000	0.00079	0.00076	0.00034
0.01	3.002	−0.366	−0.193	−0.412	−1.236	0.272	−0.00385	6.425	0.347	0.452	0.657	−0.00334	0.00341	−0.00007	0.00080	0.00076	0.00034
0.02	3.064	−0.368	−0.192	−0.425	−1.251	0.273	−0.00375	6.336	0.351	0.454	0.661	−0.00343	0.00349	−0.00006	0.00080	0.00076	0.00034
0.03	3.128	−0.378	−0.183	−0.440	−1.267	0.278	−0.00371	6.108	0.348	0.461	0.672	−0.00356	0.00364	−0.00008	0.00081	0.00077	0.00034
0.04	3.223	−0.414	−0.168	−0.487	−1.299	0.291	−0.00377	6.096	0.350	0.463	0.681	−0.00372	0.00371	0.00001	0.00082	0.00078	0.00034
0.05	3.304	−0.478	−0.165	−0.497	−1.321	0.301	−0.00388	6.086	0.352	0.469	0.704	−0.00374	0.00378	−0.00005	0.00083	0.00079	0.00035
0.10	3.757	−0.666	−0.232	−0.341	−1.342	0.295	−0.00522	7.658	0.375	0.459	0.702	−0.00330	0.00347	−0.00016	0.00084	0.00079	0.00035
0.15	3.877	−0.404	−0.226	−0.214	−1.212	0.243	−0.00693	7.468	0.362	0.463	0.693	−0.00371	0.00338	0.00033	0.00084	0.00080	0.00035
0.20	3.578	−0.217	−0.231	−0.122	−1.048	0.207	−0.00792	6.030	0.364	0.472	0.700	−0.00402	0.00348	0.00054	0.00084	0.00080	0.00035
0.30	3.482	0.107	−0.226	−0.042	−0.966	0.159	−0.00701	5.123	0.357	0.503	0.725	−0.00391	0.00308	0.00083	0.00085	0.00081	0.00036
0.40	3.340	0.243	−0.233	0.010	−0.947	0.142	−0.00539	4.750	0.366	0.540	0.737	−0.00366	0.00296	0.00070	0.00089	0.00085	0.00038
0.50	3.220	0.392	−0.191	−0.236	−0.946	0.163	−0.00497	4.580	0.382	0.528	0.767	−0.00343	0.00234	0.00109	0.00089	0.00085	0.00038
0.75	2.998	0.667	−0.169	−0.178	−0.972	0.144	−0.00197	4.685	0.382	0.541	0.769	−0.00229	0.00175	0.00052	0.00088	0.00084	0.00039
1.00	2.880	0.837	−0.176	−0.114	−0.990	0.128	−0.00094	5.392	0.369	0.523	0.786	−0.00226	0.00186	0.00039	0.00088	0.00086	0.00039
1.50	2.312	1.127	−0.127	−0.094	−0.948	0.139	0.00000	4.553	0.365	0.534	0.806	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
2.00	1.684	1.079	−0.159	−0.222	−0.911	0.162	0.00000	4.309	0.360	0.553	0.793	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
3.00	1.057	1.474	−0.039	0.052	−0.855	0.160	0.00000	4.365	0.433	0.519	0.774	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
4.00	0.755	1.775	0.035	0.302	−0.852	0.143	0.00000	4.990	0.429	0.507	0.683	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Units of estimated PGV is (m/s), PGA and SA (t) are estimated in (g)

Table 2 Coefficient table for V_{s30} based site response

t	g_1	g_2	$\varnothing_{S2S}$	$\Delta g_{1,IT}$	$\Delta g_{1,Others}$	$\Delta g_{1,TR}$	$\Delta g_{2,IT}$	$\Delta g_{2,Others}$	$\Delta g_{2,TR}$	$SE(\Delta g_{1,IT})$	$SE(\Delta g_{1,Others})$	$SE(\Delta g_{1,TR})$	$SE(\Delta g_{2,IT})$	$SE(\Delta g_{2,Others})$	$SE(\Delta g_{2,TR})$
pgv	2.188	−0.364	0.314	−0.296	−1.082	1.378	0.051	0.186	−0.236	0.286	0.230	0.349	0.049	0.039	0.060
pga	1.407	−0.234	0.330	−0.360	−0.678	1.038	0.063	0.119	−0.182	0.258	0.212	0.314	0.045	0.037	0.055
0.01	1.399	−0.233	0.330	−0.351	−0.663	1.013	0.062	0.116	−0.178	0.256	0.210	0.310	0.045	0.037	0.054
0.02	1.382	−0.230	0.332	−0.379	−0.655	1.034	0.067	0.115	−0.182	0.253	0.208	0.307	0.045	0.037	0.054
0.03	1.312	−0.218	0.336	−0.376	−0.652	1.028	0.066	0.115	−0.182	0.252	0.207	0.304	0.044	0.037	0.054
0.04	1.244	−0.207	0.342	−0.409	−0.606	1.014	0.072	0.107	−0.179	0.255	0.210	0.307	0.045	0.037	0.054
0.05	1.163	−0.194	0.350	−0.439	−0.562	1.001	0.078	0.099	−0.177	0.259	0.213	0.311	0.046	0.038	0.055
0.10	0.962	−0.160	0.393	−0.344	−0.390	0.734	0.061	0.070	−0.131	0.261	0.220	0.304	0.047	0.039	0.054
0.15	1.066	−0.177	0.399	−0.072	−0.341	0.413	0.013	0.062	−0.075	0.212	0.184	0.233	0.039	0.034	0.043
0.20	1.207	−0.200	0.359	−0.094	−0.403	0.497	0.017	0.072	−0.089	0.224	0.192	0.256	0.040	0.034	0.046
0.30	1.462	−0.243	0.331	−0.109	−0.660	0.769	0.019	0.115	−0.134	0.264	0.217	0.316	0.046	0.038	0.055
0.40	1.779	−0.296	0.318	−0.206	−0.777	0.983	0.036	0.135	−0.171	0.261	0.213	0.318	0.045	0.037	0.055
0.50	2.236	−0.373	0.342	−0.294	−1.113	1.406	0.050	0.191	−0.242	0.308	0.247	0.375	0.053	0.043	0.065
0.75	2.931	−0.488	0.386	−0.329	−1.501	1.831	0.057	0.258	−0.315	0.351	0.281	0.430	0.060	0.048	0.074
1.00	3.348	−0.558	0.424	−0.603	−1.581	2.184	0.103	0.271	−0.374	0.392	0.321	0.479	0.067	0.055	0.082
1.50	3.395	−0.566	0.446	−0.720	−1.309	2.028	0.123	0.223	−0.346	0.429	0.384	0.505	0.073	0.065	0.086
2.00	3.337	−0.556	0.463	−0.952	−1.009	1.962	0.162	0.172	−0.334	0.449	0.419	0.520	0.076	0.071	0.089
3.00	2.964	−0.493	0.415	−0.394	−0.831	1.226	0.069	0.145	−0.214	0.354	0.390	0.417	0.062	0.068	0.073
4.00	2.707	−0.451	0.397	−0.341	−0.791	1.021	0.060	0.138	−0.178	0.332	0.386	0.387	0.058	0.067	0.068

3.2 Style of faulting terms

Dependence of the median ground motion on style of faulting (SoF) is generally accounted through a period-dependent SoF specific adjustment to the median. Trial regressions including SoF adjustment factors on the offset showed that the estimates were not well constrained and had large standard errors. In RESORCE the distribution of recorded focal mechanisms among different regions is strongly unbalanced since in Italy most of the events are normal and very few strike-slip events, unlike in Turkey. Moreover, reverse faulting events are very few in the dataset. Considering that the odd distribution of SoF among the regions could result in a trade-off with the regional random effects on the offset, and also based on a preliminary non-parametric analysis of the dataset that showed no clearly distinguishable differences among the distance scaling of ground motion between different SoF, we chose to drop the SoF term from the functional form.

3.3 Regional variability in site-response as a function of V_{s30}

In the model described by Eq. (1), site effects are captured by the between-station terms, which account for the systematic station-specific deviation in offset with respect to the generic prediction for the population. Figure 2 shows that δB_s scales with V_{s30} indicating that V_{s30} is a first order proxy for describing site response. Large scatter around the best fit model suggests that a combination with other proxies is needed to better capture the complexity of site response (e.g. Cadet et al. 2008; Luzi et al. 2011). Besides the clear region-dependent scaling with V_{s30}, Fig. 2 suggests that the distributions of velocities for three regions are compatible with the assumption of a region-dependent reference velocity (i.e. the value of V_{s30} corresponding to zero-crossing of δB_s). Hence, we perform a further mixed-effect regression considering the following model:

$$\delta B_s = \left(g_1 + \Delta g_{1,r}\right) + \left(g_2 + \Delta g_{2,r}\right)\ln(V_{s30}) + \delta_{S2S} \tag{4}$$

Regional effects on site term are captured by the random effects $\Delta g_{1,r}$ on offset g_1, and $\Delta g_{2,r}$ on the slope with V_{s30}. In Eq. (4), δ_{S2S} represents the systematic deviation of recordings for individual station with respect to the model accounting also for the scaling with V_{s30}. The standard deviation of δ_{S2S} is the between-station variability ($\varnothing_{S2S}$) in the

Fig. 3 Δc_3 for the three regions across different spectral time periods. Beyond spectral period of 1 s, c_3 in the regression is constrained to 0 with no regional variations. *Grey-ribbon* shows the 95 % CI about the median
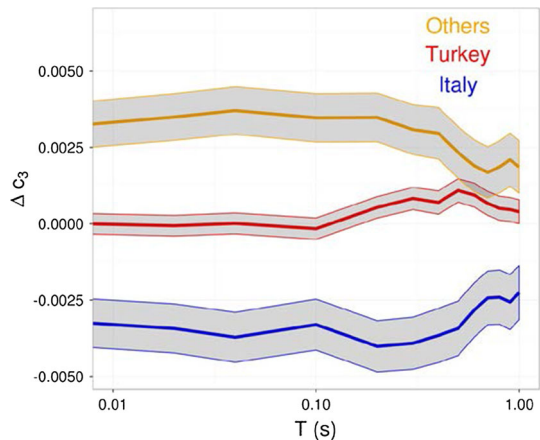
GMPE. It is worth noting that non-linear site amplification effects are not considered in the present study. Moreover, since the attenuation of high frequency ground motion can be a result of both anelastic attenuation and site effects, it is worth checking for a possible correlation (or a trade-off) between the parameters c_3 and g_2, as well as between the estimated regional variations $\Delta c_{3,r}$ and $\Delta g_{2,r}$. The results (here not shown) do not highlight any significant correlation among these parameters.

4 Results

The presence of ground motion regional variations in RESORCE dataset are modelled by allowing the site response component and the decay of ground motion with distance to be region specific. The fixed and random effects parameters relevant to regression (1) and (4) are listed in Tables 1 and 2. At each period, the mixed effect regression provides both the global c_3 value and the estimated deviation $\Delta c_{3,r}$ for each region (r), computed as random effect on c_3 in a region group. Figure 3 shows the random effects at different periods along with the associated 95 % CI, the standard error (grey ribbon). Regional variations in c_3 are shown only until spectral period of 1 s beyond which, along with c_3, they are constrained to zero. The apparent anelastic attenuation is higher for Italy than for Turkey or Others regions; a trend similar to that observed by Boore et al. (2014) in their within-event residuals which showed a faster distance-decay in Italy (and Japan) compared to Turkey (and China). The physical interpretation of the differences between the attenuation in Italy and Turkey is beyond the aim of our paper. A comparison of results available in literature for those physical properties that can influence the anelastic attenuation (e.g. velocity and attenuation topographic maps; heat flow distribution; etc.) is not straightforward because of the different implemented methodologies, the different investigated spatial scales, and the different data analyzed. In any case the standard errors on $\Delta c_{3,r}$ are small enough to indicate that the regional corrections at short periods are statistically significant. These standard errors represent the modelling uncertainty of regional adjustments to anelastic attenuation component and can be handled through ground motion logic trees. The $\Delta c_{3,r}$ random effects for different periods are listed in Table 1 along with their standard errors.

Regarding the site response term, by allowing the offset g_1 to vary among regions, we can account for regional differences in the reference V_{s30} while with g_2 we quantify the regional differences in scaling with V_{s30}. Figure 4 shows the random effects $\Delta g_{1,r}$ and $\Delta g_{2,r}$
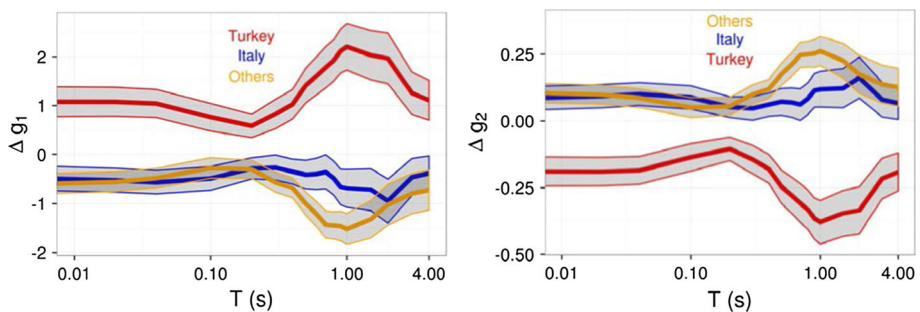
Fig. 4 Random effects on g_1 and g_2 along with their standard errors. Δg_1 and Δg_2 are estimated as a correlated-random effects. *Grey-ribbon* shows the 95 % CI about the median

for different periods, along with the estimation errors. A larger value of g_1 (and a smaller g_2) indicates a smaller reference V_{s30} for that region according to Eq. (4). Figure 1 showed that the largest fraction of recordings from Turkey come from EC8 soil class B and C stations compared to Italy and Others groups where the stations are more evenly distributed across soil classes. This means the 'centroid' V_{s30} (modal V_{s30} value) of the data is lower for Turkey as indicated by the higher positive Δg_1 value for Turkey in Fig. 4. Also seen in Fig. 2 is the stronger scaling with V_{s30} for Turkey indicated by a larger negative value for g_2 in Fig. 4. It is worth noting that by allowing regional variations in these two components of GMPE we move a fraction of the aleatory variability into epistemic uncertainty, quantified through the standard error on $\Delta c_{3,r}$, $\Delta g_{1,r}$ and $\Delta g_{2,r}$. These standard errors can be reduced by collecting more ground motion data from the regions.

5 Discussion

In the previous sections, we derived a GMPE from the European-Middle-East dataset (RESORCE), including regional (i.e. country-based) adjustments. Following recent studies, we introduced corrections for the ground motion decay and for the scaling with V_{s30} in terms of random effects.

5.1 Region dependent distance scaling and V_{s30} based site response

The regionalization of distance attenuation has been described by a region-dependent apparent anelastic attenuation model (Fig. 3). As also observed by Chiou (2012), the geometric spreading (term dependent on logarithm of distance) and of the anelastic (term dependent on distance) contribution to the attenuation show a high degree of correlation. Studies dealing with the parametrization of Fourier amplitude decay with distance in terms of geometrical spreading and anelastic attenuation shows that the trade-off between these two terms cannot be resolved using only the spectral amplitude information (e.g. Oth et al. 2011; McNamara et al. 2014). Although the model for Fourier spectral amplitude is not strictly applicable to response spectra (Bora et al. 2014), a similar situation arises with the GMPE, where the period-dependent terms controlling the linear decay with distance (i.e. c_1 and the magnitude correction c_2) are in trade-off with c_3, controlling the decay with the logarithm of distance. Since different wave types (body waves and surface waves) and phases (direct waves and reflect waves as SmS) contribute to the attenuation with distance over different distance and period ranges, the geometrical terms could be affected by regional bias related, for example, to differences in focal depths and crustal thickness (Cotton et al. 2006; Douglas 2007). Therefore, we tested a model including a correlated regional variation on the parameters controlling the distance scaling (c_1, c_2 and c_3), or considering combination of them (e.g. c_1 and c_3). Statistical tests using ANOVA (R Core Team 2013; Chambers and Hastie 1991) do not show appreciable improvements in prediction power of GMPE (e.g. comparing the Akaike Information Criterion values, performing significance tests, or analyzing the residual distributions). The estimated regional variations in anelastic attenuation ($\Delta c_{3,r}$) are similar to the ones in the simpler model discussed in previous section, and the random effects on c_2 ($\Delta c_{2,r}$) either have 0 values at high frequencies or large standard errors (encompassing 0) at low frequencies, which makes it not a well constrained regression parameter. We finally preferred not to include regional variations in c_1 and c_2 in our model.

By considering region specific reference V_{s30}, we observed remarkable differences in the site term scaling with V_{s30}. In particular, Fig. 2 shows that the slope of the between station random effects with V_{s30} is larger in Turkey than in the other two regions, both at short and long periods. Regional effects in the site term were already recognized in the NGA-West2 models. For example, Abrahamson et al. (2014) included regional corrections in the V_{s30} scaling for Taiwan, Japan and China with respect to California. As discussed in Boore et al. (2014), the observed regional variability of the site effects can be a consequence of using a simplified proxy (i.e. V_{s30}) to capture the site amplification which in fact depends on many other factors, such as the soil depth. Previous studies showed that regional differences in the depths of typical soil profiles lead amplification functions with peaks occurring at different period also for site sharing similar V_{s30} (e.g. Atkinson and Casey 2003; Ghofrani et al. 2013), as observed when comparing sites in Japan with those in California. Previous work (e.g. Boore et al. 2011) showed that the correlation of V_{s30} with the shear-wave velocity at different depths (either shallower or deeper than 30 m) is regional dependent. In particular, Boore et al. (2011) suggested that the differences in the correlation observed for Japan with respect to California, or Europe, could be ascribed to differences in the selection of the strong motion sites, since Japanese stations are mostly installed on stiff or rock material. Similar considerations could be applied also to discuss the differences observed for Italy and Turkey. Anyway, without any detailed analysis of the velocity profiles for the analysed stations, any conclusion would be speculative and we left this investigation for future studies.

Finally, in the NGA-West2 models the soil depth effect is considered through $\Delta Z_{1.0}$ (depth of basin to rock with V_{s30} of 1000 m/s), and a regionalization for this term is also
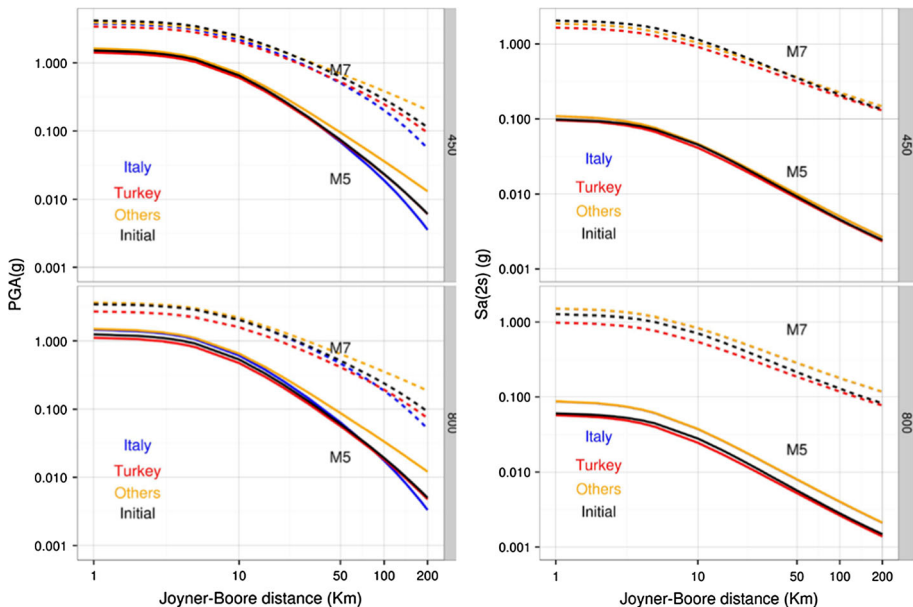
Fig. 5 Distance scaling for PGA (*left panel*) and SA (2 s) (*right panel*) at site with $V_{s30} = 450$ m/s (above panels), and $V_{s30} = 800$ m/s, for M5 and M7. Comparison of distance scaling with GMPE accounting regional variations in anelastic attenuation (*slope of the curves*) and V_{s30} scaling (*offset of the curves*), against the 'Initial' GMPE obtained from regression without accounting regional variations

considered. The site information included in RESORCE does not allow including soil depth in the model for site effects.

5.2 Impact of the regionalization on the median predictions

The impact of regional adjustments on distance scaling (Fig. 5), and magnitude scaling (Fig. 6) obtained with and without allowing the regional corrections in the regressions are compared. Included in these figures is an 'Initial' model which is a GMPE without any regional variations with functional form as in Eq. (5). Note that in Eq. (1) the regional variability in site-response is left to be examined using Eq. (4), while for the 'Initial' GMPE without regional variability a generic site response term $g * \ln(V_{s30})$ is included in the median (Eq. 5).

In the left panel of Fig. 5, regional differences in high frequency ground motion are observed as difference among the offset of the curves, and in slope of the curve at distances greater than 50 km, which are a combination of V_{s30} scaling and anelastic attenuation effects. In the right panel however, which is for lower frequency ground motion, the differences are solely due to variations in V_{s30} scaling, more pronounced for rock sites (800 m/s). Similarly, in magnitude scaling (Fig. 6) the differences in offset of the curves are a combination of regional variations in anelastic attenuation and V_{s30} scaling.

$$\ln(GM) = e_1 + F_D(R, M) + F_M(M) + g * \ln(V_{s30}) + \delta B_e + \delta B_{S2S} + \varepsilon \qquad (5)$$
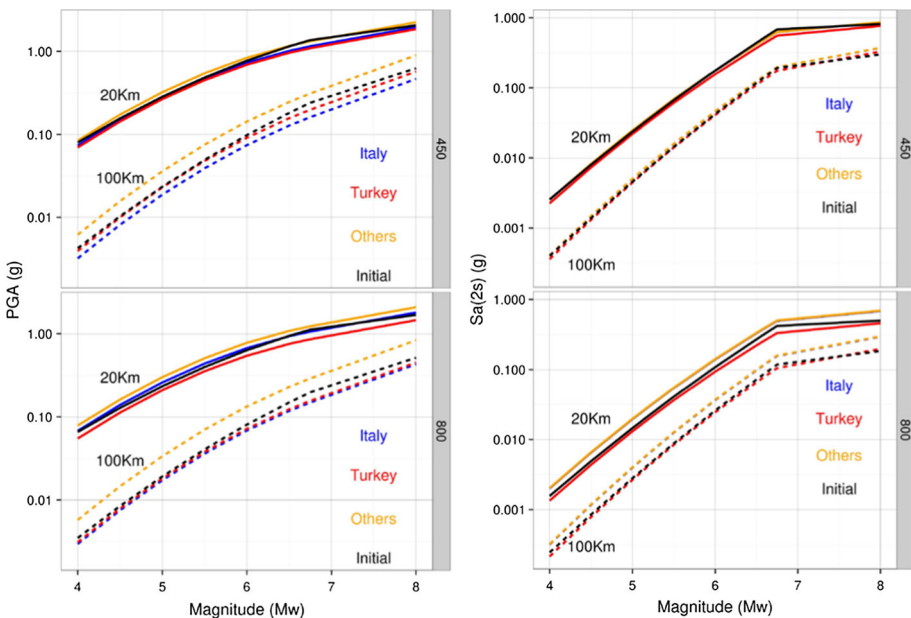
Fig. 6 Magnitude scaling for PGA (*left panel*) and SA (2) (*right panel*) at site with $V_{s30} = 450$ m/s (above panels), and $V_{s30} = 800$ m/s, for Joyner – Boore distances 20 km, and 100 km. Comparison of magnitude scaling with GMPE accounting regional variations in anelastic attenuation and V_{s30} scaling (offset of the curves), against the 'Initial' GMPE obtained from regression without accounting regional variations

Cumulative effect of all regional adjustments across spectral periods is shown in the response spectra (Fig. 7). For a site with V_{s30} of 450 m/s at distance 10 km, regional variations in anelastic attenuation and site response are negligible at all spectral periods. On the other end is a site with V_{s30} of 800 m/s located 100 km from the seismic source; in this case both anelastic attenuation and site response terms are significantly different across the regions. At the same site, for spectral periods larger than 1 s the regional differences are solely contributed to by differences in site response. The two intermediate scenarios, V_{s30} 450 m/s at distance 100 km, and V_{s30} 800 m/s at distance 10 km show effect of regional differences in anelastic attenuation, and site response scaling with V_{s30} respectively. For example, at a rock site (800 m/s) located 25 km from a rupture of magnitude M6.5 the predicted ground motion at spectral frequency of 3 Hz is 1.51 g in Italy, 1.47 g in Turkey, and 1.96 g in Others region. The differences in predicted ground motion are significant across regions after correcting the GMPE median for regional bias.
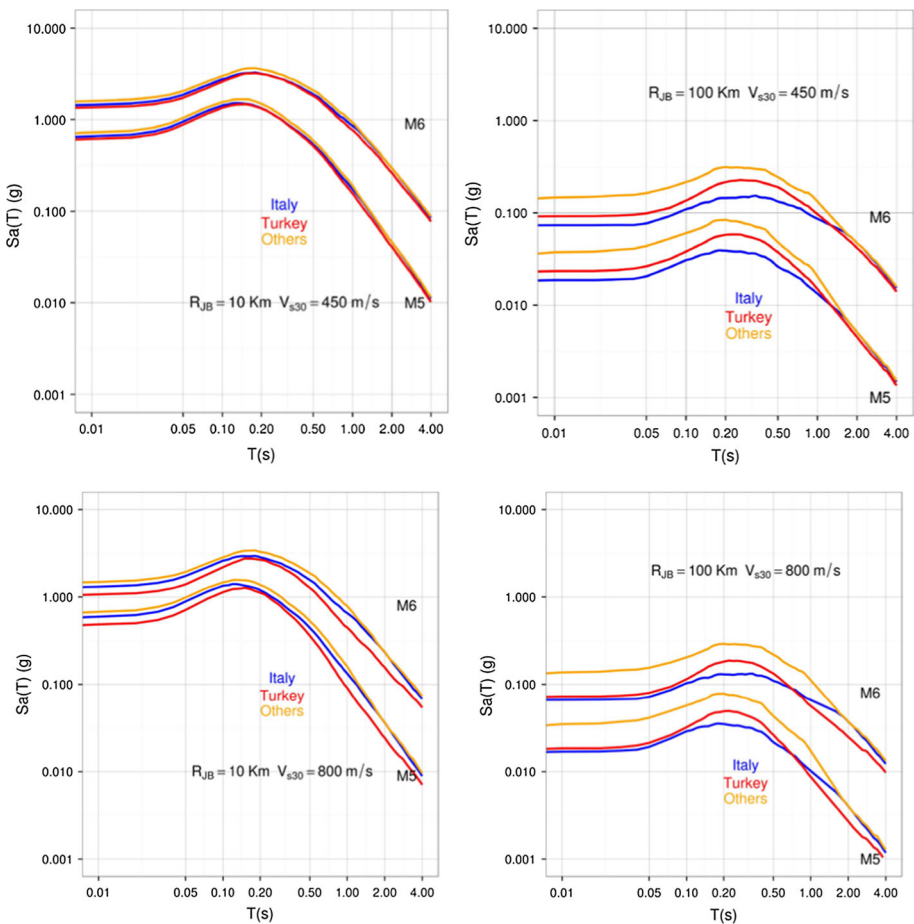
Fig. 7 Response spectra showing the cumulative effect of regional adjustments to the GMPE. Most significant differences are observed for rock sites (V_{s30} = 800 m/s) at distances larger than 50 km

5.3 Impact of regionalization on the model uncertainty

Introducing regional differences reduced the aleatory variability at the cost of an increased epistemic uncertainty in GMPE. The increase in modelling uncertainty is captured by standard errors on regional adjustments, while the reduction of variability is captured by decrease in standard deviation of GMPE given by Eq. (6)

$$\sigma = \sqrt{\tau^2 + \varnothing_{s2s}^2 + \varnothing_0^2} \tag{6}$$

Figure 8 shows the comparison of standard deviations between the model with and without regional variations 'Initial'. There is a 5–10 % reduction in the total standard deviation (σ) by introducing regional variations, primarily from the reduction of between-station variability ($\varnothing_{s2s}$) by 13–20 %. Reduction in residual ($\varnothing_0$) standard deviation is small (<2 %). There is no noticeable change in between-event standard deviation (τ).

Improvement in median prediction of the GMPE by correcting regional bias with regional adjustments is quantified in terms of Akaike Information Criterion (AIC), which is a measure of the relative quality of a statistical model for a given set of data penalized by the number of model parameters. Introducing the regional variations in this case increases the number of regression parameters by 3, yet a smaller AIC value of the model with regional variations justifies its increased complexity.

5.4 Potential regional differences in magnitude scale

The between-event residual δB_e can be used to evaluate the impact of considering earthquakes with converted moment magnitude from other magnitude scales (local magnitude, surface-wave magnitude, and body-wave magnitude). In Fig. 1, the recordings relevant to these earthquakes are shown in red and mainly correspond to magnitude smaller than 5 in Turkey. Figure 9 is a box-plot of δB_e at Sa(1 s) for each country in the regressed dataset. The scatter in δB_e from considering events with both computed (from moment tensor solutions) and empirically estimated M_w is larger than that when considering only those with computed M_w (refer to Akkar et al. 2014a for details on empirical estimation of M_w). In Fig. 9 this reduction in scatter can be seen as a shift of the country-wise median towards 0, from left to right panel. Within-country scatter shown as the height of the box-plot has also reduced, especially in case of Turkey. Filtering out events with empirically

Fig. 8 Comparison of individual components of aleatory variability in GMPE (standard deviations) between the model with regional variations (*solid line*) and without regional variations 'Initial' (*dashed lines*)
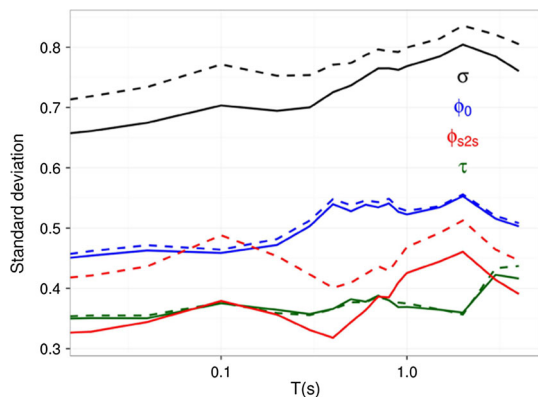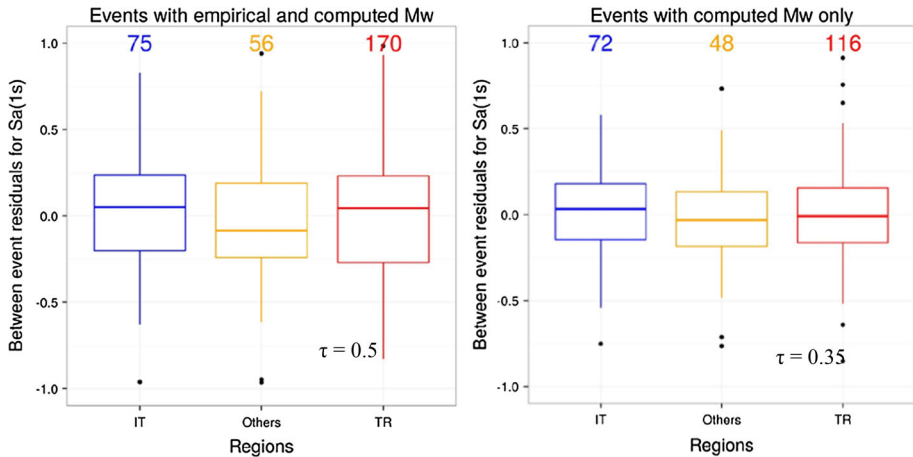
Fig. 9 Regional variation of between event residuals at SA (1 s). *Box-plot* the median (50th) and the quartiles (5th, 25th, 75th and 95th). The *left panel* shows residuals from all the events whose M_w is either computed (as calculated from moment tensor solutions) or empirically estimated (for details refer to Akkar et al. 2014a). *Right panel* shows the residuals from only the events whose M_w is computed, and not empirically estimated/converted. The decrease in height of the box plots reflects a decrease in between event variability within and across regions

estimated M_w reduced the between-event standard deviation of the GMPE (τ) by an average of 10 % (and a maximum of 30 %) across the periods, without losing constrain on other regression parameters (i.e. increase in standard error of estimate of coefficients). We note that this filter primarily removes small magnitude events from Turkey (less than M5), which could also be the reason for decrease in τ. A further study could be focused on examining the regional differences in moment tensor solutions based computed M_w which, once homogenized, may allow analyzing other regional differences in source physical parameters.

6 Conclusions

RESORCE database and the Non-linear mixed effects regression tools allowed analyzing and quantifying regional variations in ground motion data for Europe–Middle-East regions. The GMPE is developed specifically for active crustal earthquakes in Europe–Middle-East regions, and we do not recommend using it elsewhere without a prior compatibility check. The dataset is strongly unbalanced across the contributing regions in terms of magnitude, distance and recording station site classification. If separate GMPEs were to be developed for each of the regions, then the applicability of each GMPE would be strongly limited in magnitude, distance, and site V_{s30} range. By allowing regional variability only on specific terms (anelastic attenuation and site response), and estimating all the regression coefficients (magnitude scaling, geometric spreading) using the entire dataset we overcome this limitation. In its current form, the GMPE is recommended to be used for following scenarios:

- Active crustal earthquakes magnitude range from 4 to 7.6: since the magnitude distribution is symmetric around the median magnitude of M5.5, neither the small nor the large events are likely to bias the prediction.
- Sites with V_{s30} from 180 to 1000 m/s : Even though the range of V_{s30} used in regression is 90 m/s–2000 m/s, the bulk of data is within 200–600 m/s. We suggest using the GMPE in a range narrower than its underlying dataset, and especially not to extrapolate beyond the suggested V_{s30} limits.
- Joyner–Boore (R_{JB}) distances up to 200 km: the GMPE is calibrated with data up to 300 km with the bulk of data from within 150 km.
- Partially non-ergodic region specific seismic hazard assessment by adjusting the GMPE median (Table 1) and linear site-amplification model (Table 2) with the provided regional adjustments. The reported standard errors are estimated as square-root of conditional variances estimated by the Markov Chain Monte Carlo bootstrap method available in LME4.0 package in R (Bates et al. 2014). These values can also be used as epistemic uncertainty on the regional adjustments. Since the underlying distribution is not known, the epistemic uncertainty can be assumed to be normally distributed and modeled using a three-point distribution that maintains the mean and the standard deviation of the original distribution. Under such an assumption the upper and lower limits on regional adjustments can be set as ±1.6 times the standard error, with logic tree weights 0.2, 0.6 and 0.2 for upper, middle and lower branches respectively.

At the moment statistically significant regional variations in apparent anelastic attenuation, and V_{s30} scaled linear site response could be captured and accounted in the new GMPE; thereby correcting median for regional bias and deflating the total variability by 5–10 % depending on the spectral period. Regional differences in distance scaling found in this study are in agreement with recently published studies. The largest reduction in GMPE standard deviation comes from allowing regional variations in the site response component. This variability could be further reduced by using a combination of site-response proxies, instead of V_{s30} alone. Another large reduction in standard deviation comes from using only the events with moment tensor solutions based moment magnitude in regression, at the cost of losing many small magnitude events. It is desirable to plug such data losses by homogenizing the magnitude scale across regions. In summary, a decrease in aleatory variability of ground motion prediction as demonstrated in this study is accompanied by a new epistemic uncertainty on estimated regional adjustments, which in turn may only be reduced by improving the underlying datasets.

References

Abrahamson NA, Silva WJ (1997) Empirical response spectral attenuation relations for shallow crustal earthquakes. Seismol Res Lett 68(1):94–127

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

Abrahamson NA, Youngs RR (1992) A stable algorithm for regression analysis using the random effects model. Bull Seismol Soc Am 82(1):505–510

Abrahamson NA, Silva WJ, Kamai R (2014) Summary of the ASK14 ground motion relation for active crustal regions. Earthq Spectra 30(3):1025–1055

Akkar S, Sandıkkaya MA, Şenyurt M, Sisi AA, Ay BÖ, Traversa P, Douglas J, Cotton F, Luzi L, Hernandez B, Godey S (2014a) Reference database for seismic ground-motion in Europe (RESORCE). Bull Earthq Eng 12(1):311–339

Akkar S, Sandıkkaya MA, Bommer JJ (2014b) Empirical ground-motion models for point-and extended-source crustal earthquake scenarios in Europe and the Middle East. Bull Earthq Eng 12(1):359–387

Al Atik L, Abrahamson N, Bommer JJ, Scherbaum F, Cotton F, Kuehn N (2010) The variability of ground-motion prediction models and its components. Seismol Res Lett 81(5):794–801

Ambraseys N, Smit P, Douglas J, Margaris B, Sigbjörnsson R, Olafsson S, Suhadolc P, Costa G (2004) Internet site for European strong-motion data. Bollettino di Geofisica Teorica ed Applicata 45(3):113–129

Ancheta TD, Darragh RB, Stewart JP, Seyhan E, Silva WJ, Chiou BSJ, Wooddell KE, Graves RW, Kottke AR, Boore DM, Kishida T (2014) NGA-West2 database. Earthq Spectra 30(3):989–1005

Atkinson GM, Casey R (2003) A comparison of ground motions from the 2001 M 6.8 in-slab earthquakes in Cascadia and Japan. Bull Seismol Soc Am 93(4):1823–1831

Bates D, Mächler M, Bolker B, Walker S (2014) Fitting linear mixed-effects models using lme4. arXiv preprint arXiv:1406.5823

Bindi D, Massa M, Luzi L, Ameri G, Pacor F, Puglia R, Augliera P (2014) Pan-European ground-motion prediction equations for the average horizontal component of PGA, PGV, and 5%-damped PSA at spectral periods up to 3.0 s using the RESORCE dataset. Bull Earthq Eng 12(1):391–430

Boore DM, Thompson EM, Cadet H (2011) Regional correlations of V_{S30} and velocities averaged over depths less than and greater than 30 meters. Bull Seismol Soc Am 101(6):3046–3059

Boore DM, Stewart JP, Seyhan E, Atkinson GM (2014) NGA-West2 equations for predicting PGA, PGV, and 5 % damped PSA for shallow crustal earthquakes. Earthq Spectra 30(3):1057–1085

Bora SS, Scherbaum F, Kuehn N, Stafford P (2014) Fourier spectral-and duration models for the generation of response spectra adjustable to different source-, propagation-, and site conditions. Bull Earthq Eng 12(1):467–493

Cadet H, Bard PY, Duval AM (2008) A new proposal for site classification based on ambient vibration measurements and the Kiknet strong motion data set. In: Proceedings of the 14th World Conference on Earthquake Engineering, Beijing, pp 12–17

Chambers JM, Hastie TJ (1991) Statistical models in S. CRC Press, Inc, Boca Raton

Chiou BSJ (2012) Updating the Chiou and Youngs NGA model: regionalization of anelastic attenuation. In: Proceedings, 15th World Conference on Earthquake Engineering

Chiou BSJ, Youngs RR (2014) Update of the Chiou and Youngs NGA model for the average horizontal component of peak ground motion and response spectra. Earthq Spectra 30(3):1117–1153

Cotton F, Scherbaum F, Bommer JJ, Bungum H (2006) Criteria for selecting and adjusting ground-motion models for specific target regions: application to central Europe and rock sites. J Seismol 10(2):137–156

Delavaud E, Cotton F, Akkar S, Scherbaum F, Danciu L, Beauval C, Drouet S, Douglas J, Basili R, Sandikkaya MA, Segou M (2012) Toward a ground-motion logic tree for probabilistic seismic hazard assessment in Europe. J Seismol 16(3):451–473

Derras B, Bard PY, Cotton F, Bekkouche A (2012) Adapting the neural network approach to PGA prediction: an example based on the KiK-net data. Bull Seismol Soc Am 102(4):1446–1461

Douglas J (2004a) Use of analysis of variance for the investigation of regional dependence of strong ground motions. In: Proceedings of Thirteenth World Conference on Earthquake Engineering

Douglas J (2004b) An investigation of analysis of variance as a tool for exploring regional differences in strong ground motions. J Seismol 8(4):485–496

Douglas J (2007) On the regional dependence of earthquake response spectra. ISET J Earthq Technol 44(1):71–99

Douglas J, Akkar S, Ameri G, Bard PY, Bindi D, Bommer JJ et al (2014) Comparisons among the five ground-motion models developed using RESORCE for the prediction of response spectral accelerations due to earthquakes in Europe and the Middle East. Bull Earthq Eng 12(1):341–358

Ghofrani H, Atkinson GM, Goda K (2013) Implications of the 2011 M9. 0 Tohoku Japan earthquake for the treatment of site effects in large earthquakes. Bull Earthq Eng 11(1):171–203

Gianniotis N, Kuehn N, Scherbaum F (2014) Manifold aligned ground motion prediction equations for regional datasets. Comput Geosci 69:72–77

Hermkes M, Kuehn NM, Riggelsen C (2014) Simultaneous quantification of epistemic and aleatory uncertainty in GMPEs using Gaussian process regression. Bull Earthq Eng 12(1):449–466

Kale Ö, Akkar S, Ansari A, Hamzehloo H (2015) A ground-motion predictive model for Iran and Turkey for horizontal PGA, PGV, and 5 % damped response spectrum: investigation of possible regional effects. Bull Seismol Soc Am 105(2A):963–980

Luzi L, Puglia R, Pacor F, Gallipoli MR, Bindi D, Mucciarelli M (2011) Proposal for a soil classification based on parameters alternative or complementary to Vs, 30. Bull Earthq Eng 9(6):1877–1898

McNamara DE, Gee L, Benz HM, Chapman M (2014) Frequency-dependent seismic attenuation in the eastern United States as observed from the 2011 central Virginia earthquake and aftershock sequence. Bull Seismol Soc Am 104:55–72. doi:10.1785/0120130045

Oth A, Bindi D, Parolai S, Di Giacomo D (2011) Spectral analysis of K-NET and KiK-net data in Japan, Part II: on attenuation characteristics, source spectra, and site response of borehole and surface stations. Bull Seismol Soc Am 101(2):667–687

R Core Team (2013) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. http://www.R-project.org/

Scasserra G, Stewart JP, Kayen RE, Lanzo G (2009) Database for earthquake strong motion studies in Italy. J Earthq Eng 13(6):852–881

Stafford PJ (2014) Crossed and nested mixed-effects approaches for enhanced model development and removal of the ergodic assumption in empirical ground-motion models. Bull Seismol Soc Am 104(2):702–719

Stafford PJ, Strasser FO, Bommer JJ (2008) An evaluation of the applicability of the NGA models to ground-motion prediction in the Euro-Mediterranean region. Bull Earthq Eng 6(2):149–177