



On the Effect of Synthetic and Real Data Properties on Seismic Intensity Prediction Equations

ROMAN N. VAKARCHUK,¹ PÄIVI MÄNTYNIEMI,²  and RUBEN E. TATEVOSSIAN¹

Abstract—The present investigation focuses on the effect of input data properties on the estimation of seismic intensity prediction equation (IPE) coefficients. Emphasis is placed on small-to-moderate magnitude earthquakes. Synthetic intensity data points (IDPs) are created using a given IPE, assuming independence of azimuth. Extensive simulations are performed for single earthquakes and a synthetic database. Tests of single earthquakes show that increasing the sample size narrows the range of obtained coefficients. The larger the difference between the shortest and longest distance of IDPs from the epicentre, the narrower is this range. A short radius of perceptibility is more rapidly saturated with new data points than a long one. The synthetic database is used to examine the effect of magnitude and depth errors. The performance of synthetic data gives a model with which the real data can be compared. The attenuation coefficient appears stable against magnitude errors of ± 0.2 units, but starts to be overestimated as magnitude errors increase. Assuming an erroneous regional depth easily leads to intensity differences of 1 degree. The mean coefficient values deviate from the correct ones and tend to increase with depth. The results resemble the synthetic ones, but imply larger uncertainties. The attenuation coefficient, v , appears to be the least sensitive coefficient to errors. Real data from seven post-1965 earthquakes in the magnitude range of 4.0–5.2 were retrieved from the intensity database of the United Kingdom.

Key words: Macroseismology, Intensity data point, Seismic intensity prediction equation, United Kingdom.

1. Introduction

The direct macroseismic observations also have value in the era of the specialized, highly technical science. There are abundant macroseismic observations for important earthquakes not recorded by modern instruments. Seismic-hazard analyses

calculated in terms of (macro)seismic intensity (from here on intensity) have been widespread (e.g. Mayer-Rosa and Schenk 1989; McGuire 1993), and compatible updates may be necessary. Site effects can be investigated using macroseismic data (Bossu et al. 2000; Mucciarelli et al. 2000). Empirical information on the impact of earthquakes on society provides a basis for the communication of seismic hazards, for example in ShakeMaps as part of early warning implementation (Worden et al. 2010).

Attenuation parameters can be used in the general regression scheme to obtain the location, magnitude and depth of historical earthquakes (e.g., Levret et al. 1994; Bakun and Wentworth 1997, 1999). Intensity models for this purpose have been developed, for example, by Gasperi and Ferrari (1997) for Italy, Bakun and Wentworth (1997) for coastal California, Hinzen and Oemisch (2001) for the northern and Middle Rhine Area in central Europe, Bakun et al. (2003) for eastern North America, Fäh et al. (2003) for Switzerland, Bakun (2006) for southern California and Bakun and Scotti (2006) for regions of France.

The attenuation structure of the crust and upper mantle can be determined by constructing an intensity prediction equation (IPE). IPEs based on isoseismals have been presented, for example, by Brazeo (1972) for the United States west of longitude 106°W, Gupta and Nuttli (1976) for the central U.S., Ambraseys (1985) for Northwest Europe, Lapajne (1987) for Slovenia, Levret et al. (1994) for France, Pantea (1994) for the Romanian territory with adjacent areas and Musson (2005) for the United Kingdom. These studies have made use of isoseismal maps for well-studied earthquakes.

Contouring of individual intensities is avoided when using intensity data points (IDPs). IPEs based on IDP data have been presented, among others, by Stromeyer and Grünthal (2009) for parts of Germany,

¹ Institute of Physics of the Earth, Russian Academy of Sciences, B. Gruzinskaya 10, Moscow 123242, Russia.

² Department of Geosciences and Geography, Institute of Seismology, University of Helsinki, P.O. Box 68, 00014 Helsinki, Finland. E-mail: paivi.mantyniemi@helsinki.fi

France, the Netherlands, and the Czech Republic in central Europe, Sørensen et al. (2009) for the Sea of Marmara region in Turkey, Bindi et al. (2011) for central Asia and Allen et al. (2012) for shallow active tectonic crust worldwide. Many IPEs have been presented for Italy, or parts of it, including those by Peruzza (1996), Gasperini (2001), Albarello and D'Amico (2004), Gómez (2006), Faccioli and Cauzzi (2006), Pasolini et al. (2008a, b) and Sørensen et al. (2010).

The present paper examines the effect of synthetic and real data properties on the IPE. It is explored how successfully its coefficients can be resolved using input data that contain errors. The motivation arises out from areas of low seismicity and/or poorly documented earthquake effects, where it is doubtful whether the available intensity data are sufficient for constructing a local IPE. The feasibility of obtaining the average attenuation trend using empirical intensities can be examined with the help of synthetic data. They have previously been used to study macroseismic location based on sparse intensity datasets by Mäntyniemi et al. (2017). This paper begins with a formulation of the inverse problem to be solved (Sect. 2) and a scheme for creating synthetic intensity data (Sect. 3). IDP samples of different sizes and effects of magnitude and depth errors on a synthetic database are investigated (Sect. 4). An IPE is constructed on the basis of intensity data from the database of the United Kingdom provided by the British Geological Survey (Sect. 5). Finally, the results are discussed (Sect. 6).

2. Regression Technique for Solving the Equation Coefficients

Much of the previous literature listed above has been concerned with estimating the coefficients of the Kövesligethy–Sponheuer equation using regional intensity data. The content of the equation and the coupling of its coefficients have been perused in detail, for example by Ambraseys (1985), Levret et al. (1994) and Stromeyer and Grünthal (2009). Different regression techniques are typically used to resolve the coefficients. The equation includes the epicentral intensity, which may be a source of bias

(e.g., Ambraseys 1985; Musson 2005; Pasolini et al. 2008a).

In the present analysis, the issue of epicentral intensity is avoided by using the equation by Kondorskaya and Shebalin (1982), which is as follows:

$$I_i = b \cdot M - v \cdot \lg \sqrt{R_i^2 + H^2} + c. \quad (1)$$

Above, I_i , $i = 1, \dots, n$, are the intensities at n localities (x_i, y_i) and R_i the corresponding epicentral distances (in km), M is the earthquake magnitude and H the focal depth (km). Equation (1) has been calibrated against the M_L scale. The M_L and M_S magnitudes were equal in the magnitude interval used. Intensities were given on the Medvedev–Sponheuer–Kárník (MSK-64) scale. The coefficients are the attenuation coefficient, v , as well as b and c . The logarithm is to the base 10. The square root term is the hypocentral distance R_{hyp} . Rupture dimensions are not considered in this study, because a point source is assumed. Equation (1) has been used to study attenuation by e.g., Shebalin et al. (1998). They proposed attenuation $v = 4.0$ for central and south-eastern Europe ($\varphi \leq 47^\circ\text{N}$) and $v = 3.5$ for its northern part ($\varphi > 47^\circ\text{N}$).

Depth estimates are not always given in the existing IDP databases. A single value may be taken to represent a regional average depth (e.g., Bindi et al. 2011), or it may be determined as an additional regression parameter (e.g., Sørensen et al. 2010). Depth errors are investigated in the later sections.

Assuming Eq. (1) and two different magnitudes M_1 and M_2 with the corresponding hypocentral distances R_{hyp1} and R_{hyp2} , respectively, we obtain

$$I_1 = b \cdot M_1 - v \cdot \lg(R_{hyp1}) + c, \quad (2)$$

and

$$I_2 = b \cdot M_2 - v \cdot \lg(R_{hyp2}) + c. \quad (3)$$

Extracting Eq. (3) from Eq. (2) gives

$$I_1 - I_2 = b \cdot (M_1 - M_2) - v \cdot \lg \left(\frac{R_{hyp1}}{R_{hyp2}} \right). \quad (4)$$

Dividing both sides of Eq. (4) by b gives

$$M_2 - M_1 = \left(\frac{v}{b} \right) \cdot \lg \left(\frac{R_{hyp2}}{R_{hyp1}} \right) + \frac{I_2 - I_1}{b}. \quad (5)$$

Substituting $y = M_1 - M_2$ and $x = \lg\left(\frac{R_{hyp2}}{R_{hyp1}}\right)$ leads to a linear equation

$$y = \left(\frac{v}{b}\right) \cdot x + \frac{(I_2 - I_1)}{b}. \quad (6)$$

The inverse problem is solved step by step as follows: First, magnitude differences and the ratio of hypocentral distances can be calculated according to Eq. (4) for the corresponding intensities. Each distance–intensity pair is assumed to carry information on the attenuation properties. Only positive intensity differences are used to avoid redundancy. Secondly, the procedure is repeated for all ΔI . A set of linear regressions (6) is obtained and can be used to solve v and b . Non-trivial solutions of (6) are found only for non-zero y . Coefficient c can then be determined using Eq. (1).

3. Generation of Synthetic Data for Stability Tests

The synthetic intensity, I_S , at locality i was computed using Eq. (1) with the coefficients $v = 3.5$, $b = 1.5$ and $c = 3$. Synthetic data were created for both single earthquakes and a database. Since intensity was taken to be independent of azimuth, their corresponding distances from the epicentre could be considered in one dimension.

An initial set of synthetic intensities was generated starting from $I = 1.51$ with an increment of 0.01 to a maximum intensity defined with Eq. (1) with an epicentral distance equal to zero. A value was selected randomly from this set, and the corresponding distance was calculated. Then it was rounded to the nearest integer. For example, intensity $I = 4$ results from one hundred possible initial values between 3.50 and 4.49. A uniform probability over distance means equality between different intensities. However, the maximum possible intensity is defined in such a way that the number of possible intensities is the same. For example, a magnitude of 4.7 and depth of 10 km yield an initial maximum intensity of 6.55, but there are only six possible initial values, from 6.50 to 6.55, which would be rounded to 7, so in this case the final maximum intensity is 6.

Samples with 5–90 IDPs, with increments of 5, were created. Eleven magnitudes from 4.5 to 6.5 with

increments of 0.2 units were used. One million samples were computed for each magnitude and number of IDPs, and 11 times 18 million samples were thus available.

As the next step, a database of synthetic intensities was created. It was taken to be composed of 18 earthquakes and 1110 IDPs. It included four different magnitude values, 4.5, 4.7, 5.1 and 5.7, with 10, 5, 2 and 1 earthquake(s), respectively (Table 1). The number of IDPs was assumed to increase with magnitude: there were 15 IDPs for each M4.5 earthquake, 40 for each M4.7 event, 200 for both the M5.1 events and 360 IDPs related to the M5.7 earthquake. Since at this point the magnitudes were assumed to be error-free, there was no difference between ten M4.5 earthquakes with 15 IDPs and one M4.5 event with 150 IDPs. Two sets of focal depths were tested (columns depth 1 and depth 2 in Table 1). The minimum of three different integer intensities was needed for each sample to be able to solve the coefficients. They were solved for each pair $I_2 - I_1$. The mean values of 10,000 inversions were taken to be the final coefficient values.

4. Results

The synthetic IDPs represent highly idealized intensity data, but nevertheless provide insight into the effect of the input data properties on the estimated IPE coefficients.

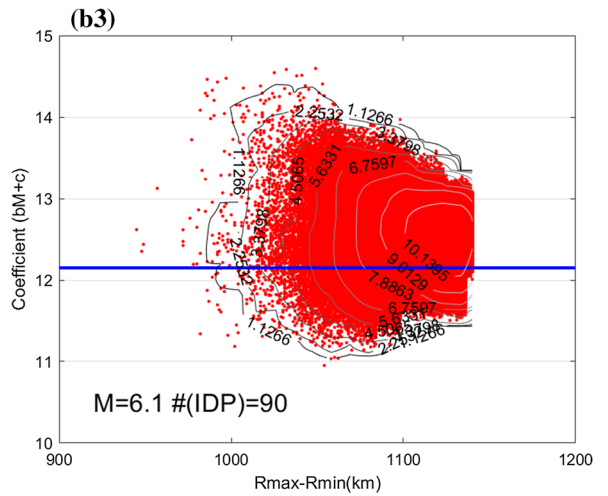
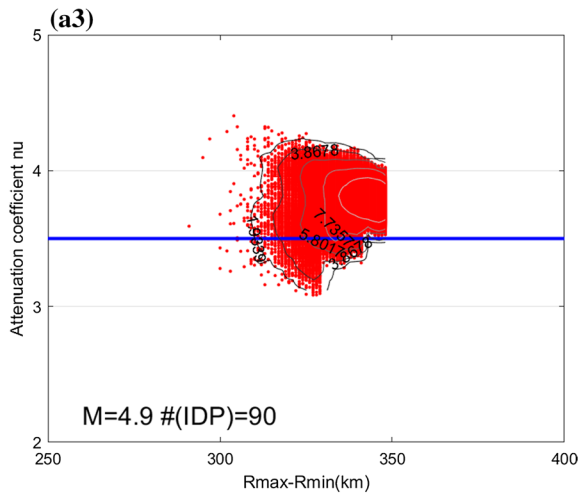
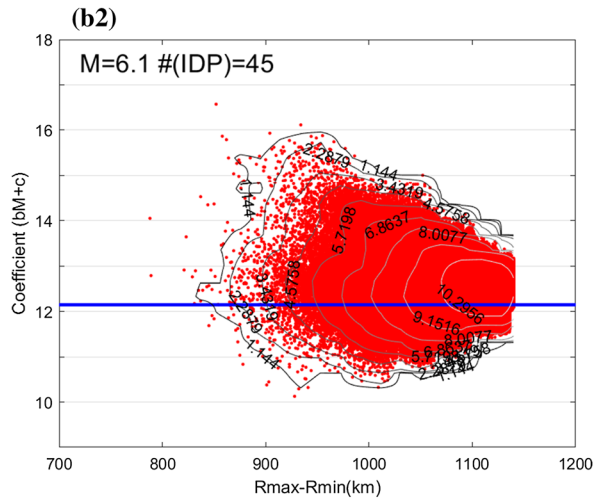
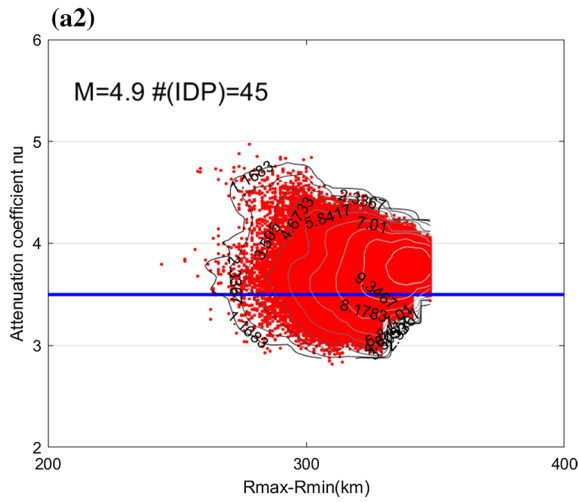
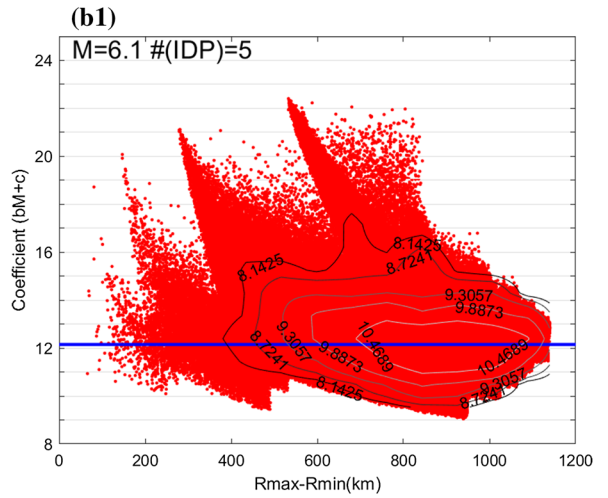
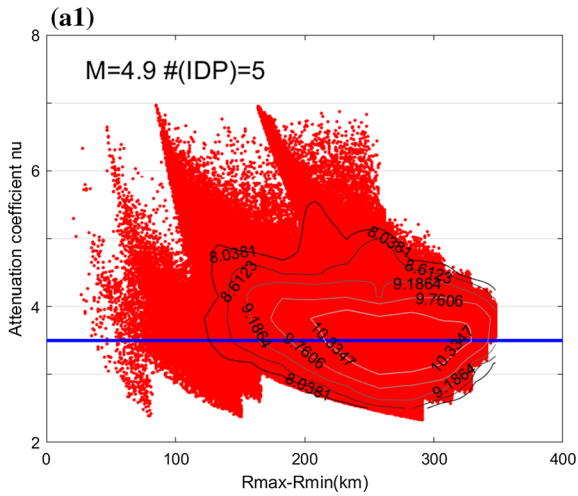
4.1. Single Earthquakes

Firstly, the initial set of synthetic intensities were used to test the selected inversion procedure.

Table 1
The synthetic database

Magnitude	Depth 1 (km)	Depth 2 (km)	No. earthquakes	No. IDPs
4.5	20	5	10	15 × 10
4.7	20	10	5	40 × 5
5.1	20	15	2	200 × 2
5.7	20	25	1	360
			Σ 18	Σ 1100

IDP stands for intensity data point



◀Figure 1

Estimation of coefficients of the intensity prediction equation on the basis of one million randomly generated samples comprising 5, 45 and 90 intensity data points: The obtained values of **a** the attenuation coefficient, ν , for magnitude $M = 4.9$ and **b** ($b \cdot M + c$), where b and c are equation coefficients and $M = 6.1$. The thick horizontal lines indicate the correct coefficient values $\nu = 3.5$ and $b \cdot M + c = 1.5 \cdot M + 3 = 12.15$. It is not possible to resolve the coupling of b and c using a single magnitude. The numbers of solutions inside the contour lines are given as exponents of the base of the natural logarithm $e \approx 2.71828$

Coefficients ν (selected to be 3.5) and $b \cdot M + c$ ($= 1.5 \cdot M + 3$) were computed, because it is not possible to resolve the coupling together of b and c using a single magnitude.

Coefficient values of ν and $b \cdot M + c$ within a standard deviation of ± 0.2 of the correct value were regarded as good results. The full range of values was wide in the case of small samples. There could be unusual distributions of small numbers of IDPs, leading to anomalous coefficients. Increasing the sample size resulted in a narrower range of estimated values. This was observed for all magnitudes, of which M4.9 and M6.1 are illustrated in Fig. 1. Simultaneously with increasing sample size, the distance ranges of the samples varied less. This follows from the random generation of the IDPs: as the sample size increases, it becomes increasingly likely that the IDPs are spread over a wider range of distances. The distributions of the estimated coefficients were typically skewed: the coefficients were overestimated rather than underestimated (Fig. 1).

The proportions of solutions within ± 0.2 of the correct coefficient value were lower for $b \cdot M + c$ than for ν . This is expected, because, as explained in Sect. 2, the inaccuracy of ν is carried into the solutions of the other two coefficients. For example, in the case of 60 IDPs and coefficient ν , the proportions were 35%, 42% and 63% for magnitudes 4.5, 5.5 and 6.5, respectively. The corresponding proportions for $b \cdot M + c$ were 19.5%, 17.2% and 28%. In the case of the attenuation coefficient, ν , there was an overall trend of a larger proportion of good results for larger magnitudes and sample sizes, but this was not strictly linear with magnitude. When solving for $b \cdot M + c$, the pattern was more complex,

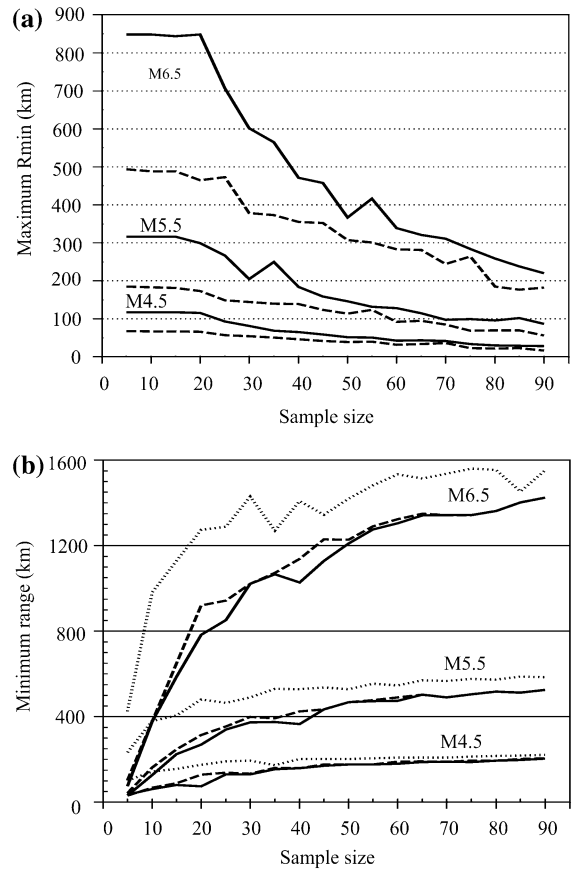


Figure 2

a The maximum shortest distance from the epicentre (R_{\min}) and **b** the minimum range of distances ($R_{\max} - R_{\min}$) of randomly generated samples as a function of sample size (number of intensity data points) for magnitudes of 4.5, 5.5 and 6.5. The solid lines correspond to one million samples and the dashed lines to a subset of samples that yielded an attenuation coefficient value in the range of $3.3 \leq \nu \leq 3.7$. In **b**, the dotted lines correspond to the second subset with the samples that yielded the correct value $\nu = 3.5$ exactly

although the largest proportions of acceptable coefficient values were obtained for the larger magnitudes.

Three sets of data were compared for each sample size using magnitudes 4.5, 5.5 and 6.5: all one million samples, a subset of samples that gave an attenuation coefficient in the range of $3.3 \leq \nu \leq 3.7$ and a second subset of samples leading to $\nu = 3.5$ exactly. The size of the second subset was typically of the order of fractions of a per cent. The longest distance from the epicentre was shorter in the samples of the first subset than in all one million samples (Fig. 2a). The benefit of increasing the sample size is largest for

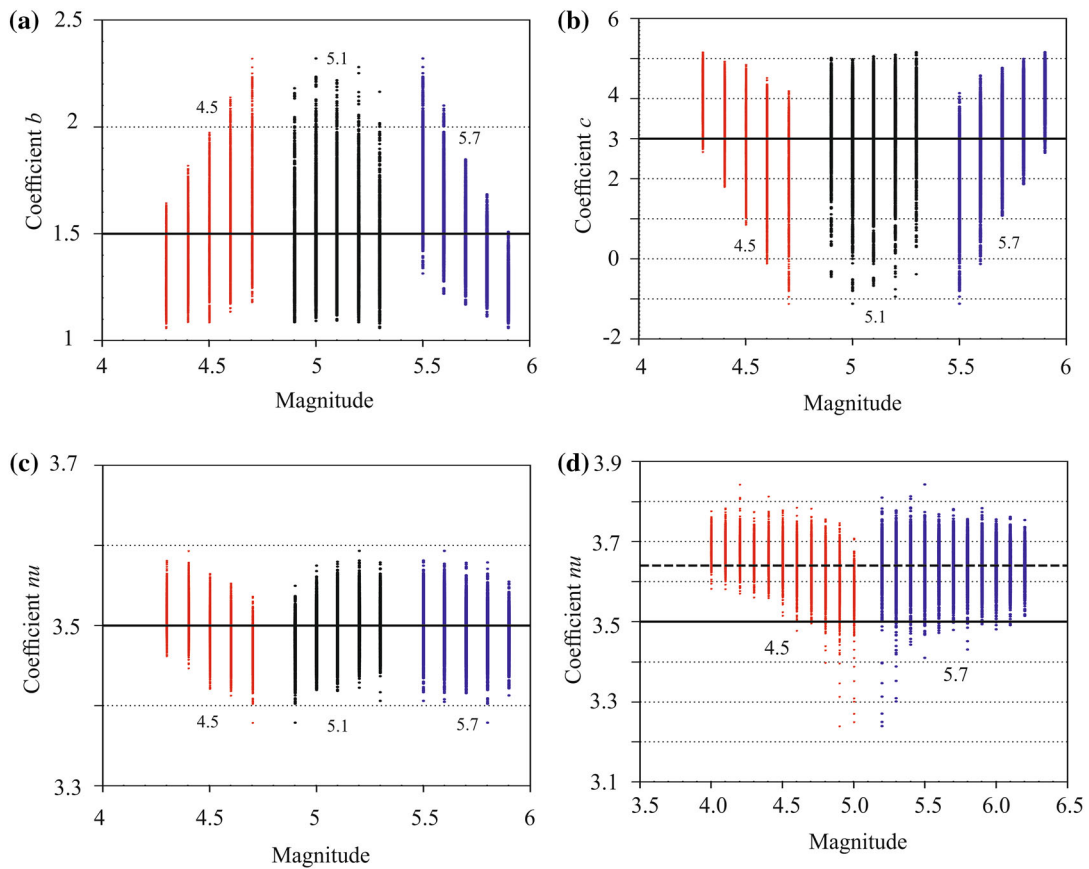


Figure 3

Effect of magnitude errors on the intensity prediction equation coefficients **a**, **b**, **c** and **ν** in the case of synthetic data and magnitudes of 4.5, 5.1 and 5.7. All combinations of magnitudes and their errors ± 0.1 and ± 0.2 magnitude units were tested. The thick horizontal lines indicate the correct coefficient values. Part **d** shows **ν** in the case of magnitudes 4.5 and 5.7 and their errors up to ± 0.5 units. The dashed line is the arithmetic mean of 10,000 inversions

magnitude 6.5: because the maximum distance range for magnitude 4.5 is not long, increasing the sample size brings no particular benefit, as the added IDPs are found to be close together. The samples leading to $\nu = 3.5$ exactly stand out in that the minimum distance range is much longer than in the other two cases (Fig. 2b). In all three cases, the minimum range of all samples approaches the maximum value as the sample size increases.

Briefly, the exercise suggests that the success of estimating the coefficient depends on the spread of intensities and the corresponding distances from the epicentre along the radius of perceptibility. A short radius of perceptibility is more rapidly saturated with new data points than a long one. These features also

indicate that the selected inversion procedure performs correctly and can be used to resolve the IPE coefficients.

4.2. The Database

The estimation of coefficients was also investigated using the synthetic database (Table 1). As expected, the presence of a magnitude range helps to resolve all the unknown coefficients, although here it extends slightly over one magnitude unit and is composed of only four different magnitudes. A minimum of three different integer intensities was needed for each sample to be able to solve the three coefficients.

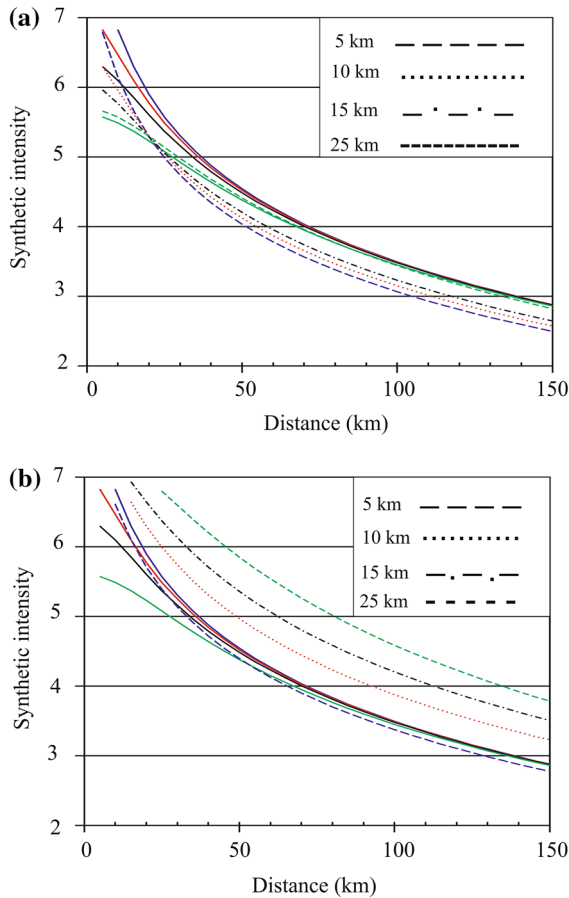


Figure 4

Effect of the assumed regional depths of 5, 10, 15 and 25 km on the intensity prediction equation (IPE) in the case of the synthetic database. The solid lines indicate the IPEs with the correct coefficients b , c and v . Depth 5 km is blue, 10 km red, 15 km black and 25 km green. The other IPEs were plotted using the mean coefficients of 10,000 inversions and the corresponding depths as given in Table 1 in the columns **a** depth 1 and **b** depth 2. All IPEs were plotted for a magnitude of 5

A test of 10,000 random selections of 5 IDPs per earthquake, totalling 90 IDPs, revealed that the coefficients could be successfully resolved. (Assuming error-free magnitudes, this is identical to one M4.5, M4.7, M5.1 and M5.7 earthquake with 50, 25, 10 and 5 IDPs, respectively). The mean of the attenuation coefficient v obtained was 3.50 ± 0.04 , and $v_{\min} = 3.32$ and $v_{\max} = 3.66$. The corresponding values of coefficient b were 1.50 ± 0.068 , $b_{\min} = 1.28$ and $b_{\max} = 1.78$, and those of coefficient c were 3.00 ± 0.30 , $c_{\min} = 1.68$ and $c_{\max} = 3.99$. Removal of the M5.7 earthquake from the synthetic database

Table 2

Effect of depth errors on the estimation of coefficients b ($= 1.5$), v (3.5) and c (3.0) using the synthetic database (Table 1)

Depth (km)	#(Eq)	%		b	v	c
5	10	13.5	Mean	1.02	3.41	5.10
			SD	0.07	0.10	0.30
			Min	0.78	3.07	3.68
			Max	1.37	3.90	6.16
10	5	18.0	Mean	1.00	3.71	6.33
			SD	0.08	0.13	0.31
			Min	0.72	3.22	4.81
			Max	1.48	4.34	7.72
15	2	36.0	Mean	0.98	4.02	7.36
			SD	0.08	0.17	0.39
			Min	0.68	3.44	5.97
			Max	1.44	5.05	9.17
25	1	32.5	Mean	0.99	4.76	9.20
			SD	0.10	0.30	0.69
			Min	0.51	3.88	6.85
			Max	1.53	7.23	12.59

The depth column shows the assumed regional depth. Column % shows the percentage of intensity data points that correspond to the depth assumed in the inversion

SD standard deviation

narrowed the magnitude range to only 0.6 units. The mean and standard deviation of coefficient v remained as above, but the standard deviation of b grew to 0.11 and that of coefficient c to 0.48. In all these calculations, there was a strong negative correlation, approximately -0.95 , between b and c , which is understandable from Eq. (1).

Then the coefficients were estimated using all possible combinations of magnitude errors ± 0.1 and ± 0.2 up to ± 0.5 units (Fig. 3). The obtained coefficients b were in the interval 1–2.4 (Fig. 3a). The biggest effect was observed on the c coefficient: It became negative, when the largest magnitude of 5.7 was replaced by 5.6 or 5.5 and simultaneously the smallest magnitude of 4.5 became either 4.6 or 4.7 (Fig. 3b). To compensate for this, the coefficient b increased above 2. Small errors up to ± 0.2 units had a minimal effect on the attenuation coefficient, which was almost resolved exclusively within the interval 3.4–3.6 (Fig. 3c). It can also be seen that the obtained coefficient values are not symmetrical around the correct values. When the magnitude errors increased, the range of attenuation coefficient values widened and the values were typically overestimated (Fig. 3d).

Table 3

The earthquakes selected from the intensity database of historical earthquakes in the United Kingdom provided by the British Geological Survey (<http://www.quakes.bgs.ac.uk/historical/>, last accessed November 2017)

Date	Origin time (UTC)	Longitude	Latitude (°N)	Magnitude	Depth (km)
1966 July 23	01:50	5.22°W	50.09	m_b 4.1	18
1972 Mar 7	06:52	2.20°W	53.70	m_b 4.0	14
1974 Aug 10	12:49	5.35°W	57.19	M_L 4.4	10
1979 Dec 26	03:57	2.97°W	55.01	M_L 5.2	11
1990 Apr 2	13:46	3.03°W	52.43	M_L 5.1	14
1994 Feb 15	10:15	0.91°E	52.56	M_L 4.0	9
2000 Sep 23	04:23	1.61°W	52.25	M_L 4.2	14

The depths are from the Bulletin of the International Seismological Centre

As the next step, the effect of the focal depth was examined using the concept of average regional depth. A total of 10,000 trials were carried out for each case. When all initial depths are equal to 20 km (column depth 1 of Table 1), the IPEs with the correct coefficients give higher intensities than those based on the assumed regional depths of 5, 10 and 15 km, whereas the IPE based on the depth of 25 km catches up with the correct ones as the hypocentral distance approaches the epicentral one (Fig. 4a). The curve corresponding to an assumed regional depth of 5 km deviates the most from the correct ones, which is reasonable. The differences due to the depth are not large except for epicentral distances below 30 km, where they can be 1 intensity degree, and below 10 km even 2 degrees. If the assumed regional depth is shallower than the correct one, the mean attenuation coefficient is underestimated and if it is deeper, v is overestimated.

It is often observed that earthquake depth tends to increase with magnitude, but any possible effect of this is taken to be insignificant in the small magnitude span of the database. However, earthquake occurrences at different depths were also modelled (column depth 2 of Table 1). The obtained coefficients v and c increased with depth (Table 2). The mean of coefficient v was 3.41 and 3.71 at depths 5 and 10 km, respectively, and above 4 at depths 15 and 25 km. All coefficients c were too high, including the minimum values. Coefficient b was quite stable from one depth to another, but the estimated values were close to 1, which is incorrect. Figure 4b shows that the depth of 5 km gave the equation closest to the correct one. All M4.5 earthquakes,

corresponding to more than half of the events, but only to 13.5% of all IDPs, had correct depths. Typically the highest intensities follow from the depths of 15 and 25 km over the displayed distance range, and they are one intensity higher than the other curves. This runs counter to what is expected.

In conclusion, a rather small number of low-magnitude earthquakes appears sufficient for the successful solution of the IPE coefficients: the synthetic database only has four different magnitude in a narrow range. However, even small magnitude errors affect the values of b and c coefficients, and the attenuation coefficient is also affected by large magnitude errors. Assuming a regional depth that deviates from the correct one can be misleading, and the mean coefficient values may deviate from the correct ones.

5. Coefficient Estimation Using UK Data

The coefficients of the IPE Eq. (1) were also estimated using real data. Since this analysis focuses on small-to-moderate magnitude earthquakes, the data were retrieved from the intensity database of historical earthquakes in the United Kingdom provided by the British Geological Survey (BGS) (<http://www.quakes.bgs.ac.uk/historical/>, last accessed November 2017). The initial selection criteria were post-1965 earthquakes with $M \geq 4$. At the time of downloading, the database ended in 2001, and twelve earthquakes fulfilled the criteria. Five of them were omitted. Two earthquakes occurred in the same region within an hour and 57 min on 25 February

1974. The first earthquake had a magnitude M_L 3.9, the second 4.1. For the second earthquake, 90 intensities out of 97 are less than 4, so there is a possible mix-up between the two events. The earthquake of 19 July 1984 was located close to the shore, and the spatial distribution of intensities appears contaminated. The distance ranges of the IDPs associated with the earthquakes of 9 August 1970 and 4 March 1999 largely overlap, making the assessment of attenuation complicated.

A total of seven earthquakes were left in the dataset (Table 3, Fig. 5). The quality of their

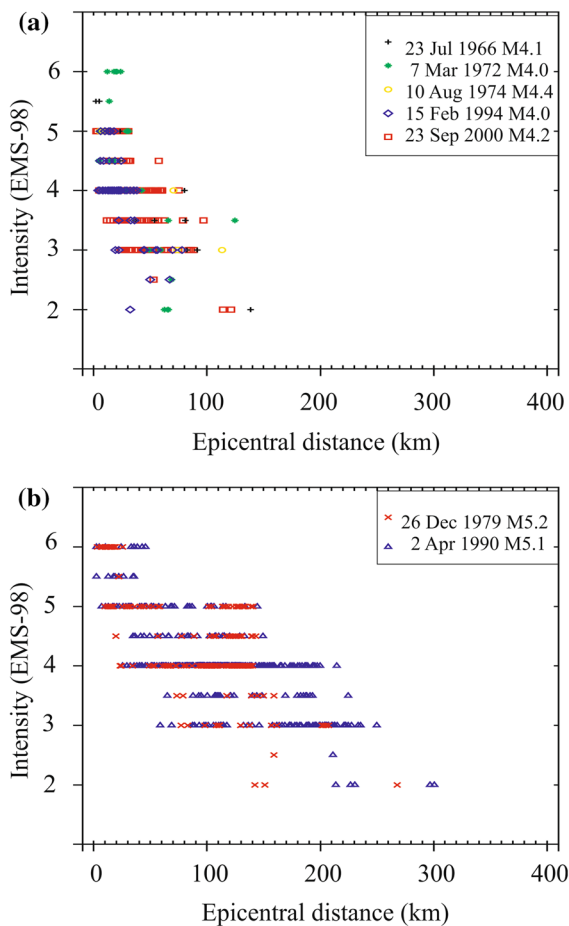


Figure 5

The set of intensity data points retrieved from the UK Historical Earthquake Database of the British Geological Survey (<http://www.quakes.bgs.ac.uk/historical/>, last accessed November 2017). Seven post-1965 earthquakes were used: **a** five with the given magnitude below 4.5 and **b** two with magnitude above 5. Magnitudes are on the M_L scale (m_b scale for 1966 and 1972). The uncertain intensities (17% of all) have been plotted between the integer intensity degrees

locations and magnitudes was examined. For example, the epicentre given for the earthquake of 7 March 1972 (m_b 4.0, at 06:52 UTC) is at least 15 km from the cluster of localities with the maximum intensity of 6. The intensity decreases as a function of distance when using the epicentre coordinates determined by Le Bureau Central Sismologique Français given in the Bulletin of the International Seismological Centre (ISC). The magnitude 4.7 given for the earthquake of 26 December 1979 leads to an anomalously large area of perceptibility. The anomaly disappeared when using the magnitude M_L 5.2 determined by the Institute of Geological Sciences in the UK.

The BGS database does not include focal depths, so they were taken from the ISC Bulletin. Priority was given to the depths estimated by agencies in the UK. Magnitude is given on the m_b instead of the M_L scale for the two oldest earthquakes. We assume that the difference in the magnitude scales is negligible in comparison with the accuracy of the magnitude evaluation. For example, magnitude estimates given by different agencies for the earthquake of 26 December 1979 vary by up to 0.5 units. Musson (1996) reported that recent magnitudes were of ± 0.2 M_L , which is taken to refer to the early 1990s. The accuracy is poorer for earlier events recorded at few stations.

The number of IDPs is 896, ranging between 5 and 426 IDPs per earthquake. Intensities assigned to monumental buildings or large territories, descriptions of “felt” and values of 1 (not felt) were excluded from the data. The reported maximum intensities are 5, 5–6 or 6 on the European Macroseismic Scale (EMS-98). The intensity degrees between EMS-98 and MSK-64 assumed by Eq. (1) are similar, except for the two highest intensities (Musson et al. 2010). An intensity $I = 2$ is infrequently reported (Fig. 5), so we have taken the minimum intensity to be 3. Many authors have pointed out the incompleteness of the IDP data in the far field, which may cause bias in statistical analysis (Albarello and D’Amico 2004; Allen et al. 2012, among others). A prevalent practice is to define 3 or 4 as the lower intensity limit in the final input data (e.g., Pasolini et al. 2008a; Stromeyer and Grünthal 2009; Sørensen et al. 2010).

Intensities may be uncertain, which is formally defined in the EMS-98 guidelines (Grünthal 1998). For example, an intensity given as 7–8 means that the intensity is either 7 or 8 at the locality in question. This implies epistemic uncertainty originating from inadequate or skewed documentation on earthquake effects, or poor survival of documents to the present time. The notation addresses a relevant issue, but does not provide practical instructions. It may easily be associated with “7 to 8”, and an intermediate degree appears appropriate, but it rather reads “7 or 8” (7/8).

So-called half-intensity values (7–8 is replaced by 7.5) are sometimes used when deriving IPEs (e.g., Stromeyer and Grünthal 2009; Sørensen et al. 2010; Bindi et al. 2011), although it is understood that this practice compromises the integer character of intensity. The new classes between integer values may be less dispersed but are less reliable (Peruzza 1996), and the practice suggests that the intensity scale has 23 degrees instead of twelve (Musson 1998). Statistical approaches have been used to account for the uncertainty (Magri et al. 1994; Peruzza 1996; Pasolini et al. 2008b), or the input data have been visually checked (Stromeyer and Grünthal 2009; Sørensen et al. 2010).

We propose to investigate the effect of uncertain values on the IPE. The present dataset includes 152 uncertain intensities, or about 17% of the IDPs. They were handled in three ways: they were omitted,

rounded down (for example, 4–5 was replaced by 4) and rounded up (4–5 was replaced by 5). Rounding down and up to the closest integer give the limits of the effect of uncertain intensities. The obtained coefficients were $b = 1.8$, $c = 1.9$ and $v = 3.5$ when the uncertain intensities were removed. When they were kept in the data and rounded down, the values were $b = 1.4$, $c = 4.6$ and $v = 3.9$, whereas rounding up gave $b = 1.4$, $c = 4.5$ and $v = 3.9$. Rounding up the uncertain values gives slightly lower values than rounding down, but the differences are not significant. Omitting the uncertain intensities has the largest impact on the coefficients in the present case.

A stability test was performed on the coefficient values (Table 4). The events of the database were removed one at a time, and the coefficients were estimated on the basis of the remaining six earthquakes. It can be seen, for example, that removing the five IDPs related to the Kintail earthquake of 10 August 1974 had a larger effect on the v coefficient than removing the Warwick earthquake of 23 September 2000 with its 157 IDPs. According to this test, the IPE coefficients can be resolved with an accuracy of 0.23 for b , 0.38 for v and 1.0 for c . However, no earthquake parameter errors were considered.

Accounting for magnitude errors demonstrated that the outcome resembles that of the synthetic case (Fig. 3), but the ranges of the obtained coefficients

Table 4

Coefficients b , v and c when the earthquakes of the UK dataset were removed one at a time

Removed earthquake	b			v			c		
	Down	Up	Omitted	Down	Up	Omitted	Down	Up	Omitted
23 July 1966	1.39	1.61	1.40	3.70	3.85	3.65	4.13	3.43	4.08
7 Mar 1972	1.43	1.59	1.41	3.43	3.54	3.33	3.44	2.97	3.40
10 Aug 1974	1.51	1.55	1.46	4.59	4.37	4.46	5.24	4.78	5.34
23 Dec 1979	1.69	1.71	1.60	3.82	3.98	3.75	2.94	3.26	3.30
2 Apr 1990	1.02	1.11	0.91	3.65	3.80	3.53	5.38	5.40	5.73
15 Feb 1994	1.14	1.33	1.17	3.49	3.64	3.49	5.04	4.47	4.95
23 Sep 2000	1.18	1.40	1.18	3.32	3.51	3.32	4.55	3.88	4.56
Mean	1.34	1.47	1.30	3.71	3.81	3.65	4.39	4.03	4.48
SD	0.23	0.20	0.23	0.42	0.30	0.39	0.93	0.89	0.94
Median	1.39	1.55	1.40	3.65	3.80	3.53	4.55	3.88	4.56

Columns down, up and omitted refer to the uncertain intensities (17% of all intensity data points) that were rounded down and up to the closest integer or omitted from computations

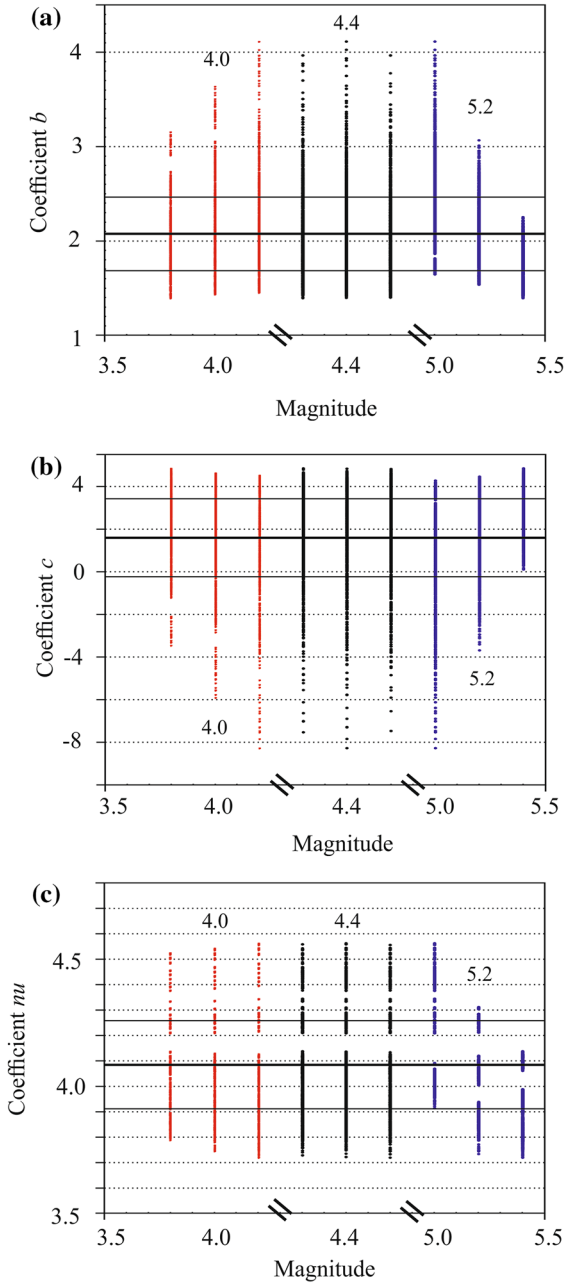


Figure 6

Effect of magnitude errors on the UK dataset. The coefficients of the intensity prediction equation **a** b , **b** c and **c** ν obtained for magnitudes M_L 4.0, 4.4 and 5.2 ± 0.2 units are shown. The thick horizontal lines are the arithmetic means of 10,000 inversions, and the thin lines are the corresponding standard deviations. The x-axis is discontinuous to avoid overlap

were larger and the range of coefficient ν was quite fragmentary (Fig. 6). This is attributed to inconsistencies in the real data. The M_L 4.0 earthquake

displayed in Fig. 6 is that of 15 February 1994. If the m_b 4.0 earthquake of 7 March 1972 is used instead, the range of obtained coefficient values increases. The arithmetic means and the corresponding standard deviations of 10,000 inversions are 2.1 ± 0.4 for b , 1.6 ± 1.8 for c and 4.1 ± 0.2 for ν . Here, there was also a strong negative correlation, about -0.98 , between b and c , which is understandable from Eq. (1).

The BGS intensities were compared to calculated ones in order to obtain an understanding of their variance. The coefficients of Eq. (1) were resolved for the three regional depths and the given depths (Table 3) in order to calculate the intensity at each available site. The uncertain intensities were downgraded, upgraded and omitted (Table 5). The variance was of the order of 0.5–0.6 in most cases, implying an uncertainty of 1 intensity degree.

Assuming regional depths of 5, 10 and 25 km for the UK dataset (Fig. 7) gave a different pattern from the synthetics (Fig. 4). The IPEs corresponding to the different depths were not parallel over the entire distance range, but intersected at intensity 4 at the distance of approximately 100 km from the epicentre. At the longer distances, the larger regional depths give smaller intensities.

In summary, the results obtained using real data resemble the pattern of magnitude errors of the synthetic results but imply larger existing inconsistencies. Assuming different regional depths affects the shape and level of the IPE. In all testing, coefficient c was the most and attenuation coefficient ν the least sensitive to errors. In the case of the synthetic and real databases, a strong negative correlation between coefficients b and c was observed. This is attributed to the form of Eq. (1), in which these coefficients try to accommodate to the data properties. In the inversion procedure (Sect. 2), coefficient c also accumulates some of the error of other two coefficients. It can be inferred that a value of c close to the value of 3 indicates less inconsistencies in the input data than the higher absolute values.

Table 5

The resolved coefficients of the intensity prediction equation using the depths given in the bulletin of the International Seismological Centre (Table 3) and three assumed regional depths

Depth	b_{down}	b_{up}	b_{omitted}	v_{down}	v_{up}	v_{omitted}	c_{down}	c_{up}	c_{omitted}
ISC	1.34	1.47	1.3	3.72	3.81	3.65	4.39	4.03	4.48
5 km	1.32	1.52	1.29	3.01	3.15	3.05	3.13	2.51	3.38
10 km	1.33	1.53	1.33	3.37	3.51	3.37	3.76	3.2	3.87
25 km	1.42	1.61	1.43	4.88	5.14	4.78	6.34	6.08	6.21

Subscripts down, up and omitted refer to the uncertain intensities (17% of all intensity data points in the UK dataset) that were rounded down and up to the closest integer or omitted from computations

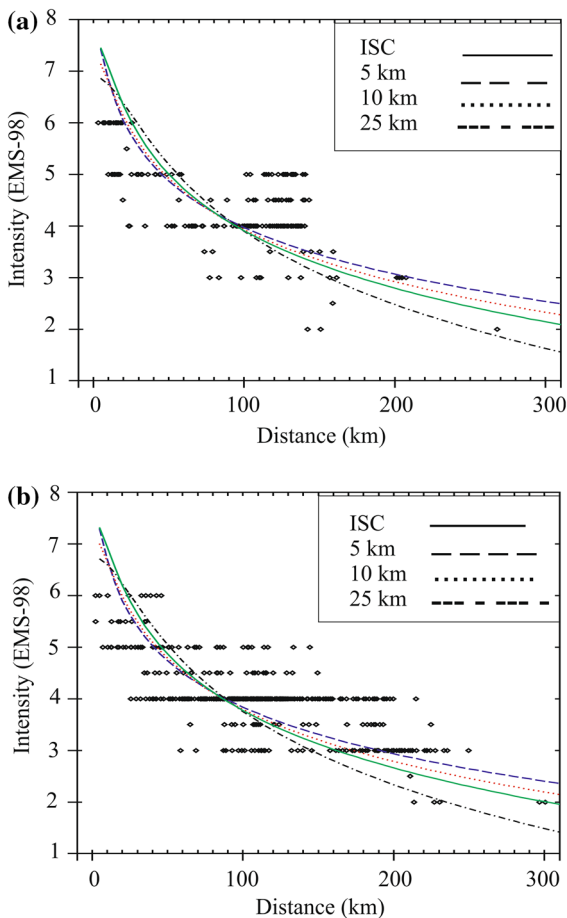


Figure 7

Effect of assumed regional depths of 5, 10 and 25 km on the intensity prediction equation in the case of the UK earthquakes of **a** 26 Dec 1979 (M_L 5.2, H = 11 km) and **b** 2 Apr 1990 (M_L 5.1, H = 14 km). The corresponding coefficients are given in Table 5 (column 'down'). The open diamonds are the intensity data points available for the two earthquakes

6. Discussion and Conclusions

In the available literature, variation in intensity close to the epicentre is regarded as a potential source of bias in the analysis (e.g., Bakun and Scotti 2006). The shortest distances from the epicentre, some 100 km, are of principal interest in applications such as ShakeMaps. Davis et al. (2000) demonstrated that an anomalous intensity was caused by geological focusing of seismic waves at the distance of 21 km from the epicentre. Hinzen and Oemisch (2001) attributed the observed irregularities in near-focal areas to source radiation and varying ground amplification conditions. Using the Kövesligethy–Sponheuer equation, Levret et al. (1994) concluded that, particularly in the near field, intensity attenuation is much more dependent on the depth of focus than on absorption by the soil.

Given such challenges, alternative approaches to the modelling of intensity attenuation have been proposed. For example, Magri et al. (1994) used a logistic model to estimate the probability that the attenuation exceeds a threshold value at a given distance from the epicentre. When intensity attenuation, ΔI , is regarded as a random variable, the probability of the site intensity becomes the convolution of the probability distribution of epicentral intensity, I_0 , and that of the intensity attenuation (e.g., Tsapanos et al. 2002). Rotondi and Zonno (2004) took intensity attenuation to be a random variable that follows the binomial distribution with parameters (I_0 , p), where p depends on the distance from the epicentre and is a Beta random variable according to the Bayesian paradigm. The probabilistic approaches avoid the

construction of a local IPE and are often targeted at seismic-hazard analyses.

In the large majority of available literature, a deterministic function and local intensity data are employed to investigate attenuation. One reason behind the IPE is the need for the regional correlation between magnitude and area of perceptibility to estimate the size of historical earthquakes. A prevalent approach is to take attenuation to be a function of I_0 and the distance of the site from the epicentre and use an exemplary earthquake of the target region. Alternatively, only the distance from the epicentre and all the sites in the database are considered.

This investigation belongs to the latter type, which may be attractive in regions with small intensity databases. The basic elements of input are intensities that are equal to, or larger or smaller than, the other available values, the configuration of localities and the distances between them (e.g., Mäntyniemi et al. 2014). In the present paper, independence of azimuth was assumed. In real target regions, population centres remain fixed for many years, and randomness of the distance to them is created by different earthquake locations. The synthetic samples suggest that widespread intensities along the radius of perceptibility are a property of good data. In real cases, for example an offshore earthquake, part of the distance and intensity range is unavailable to the analysis. The synthetic single earthquakes clearly showed that an IDP not far from the epicentre implies good chances of solution (Fig. 2). The distribution of settlements can influence the data (e.g., Musson 2005). For example, the IDPs of the Carlisle earthquake of 26 December 1979 have a gap in the distance range because of mountainous territory (Figs. 5b, 7a). In extreme cases, such features can affect the median distance. The synthetic and real database of the present study revealed a clear effect of errors of magnitudes and focal depths on the IPE coefficients (Figs. 3, 4, 6, 7). They indicate that bias does not necessarily only follow from intensity assessments. Instrumental parameter determination and the collection of macroseismic data are parallel activities, and their combined results can provide outliers. Magnitude and location errors can affect the results and a wrong choice for the regional depth may affect the IPE coefficients. This is different from the

findings of Sørensen et al. (2010), who concluded that the uncertainties in earthquake source parameters are negligible in comparison to the spread in the intensity data. Their investigation included also historical earthquakes in the magnitude range from M_w 6.3 to 7.0.

Figure 8 compares the present analysis with the IPE provided by Musson (2005). It was based on 727 isoseismals from 326 British earthquakes. These data also included historical earthquakes and covered the magnitude range of M_L 2.0–6.1. The present evaluation is based on a much narrower magnitude range 4.0–5.2 with a gap range of 4.5–5.0. The values used are from rounding up of the uncertain intensities of the UK data as such ($b = 1.4$, $c = 4.5$, $v = 3.9$; Sect. 5), the data shaking experiment (Table 4) with the uncertain values upgraded (corresponding means $b = 1.47$, $c = 4.03$ and $v = 3.81$) and the means from the inclusion of magnitude errors (Fig. 6; 2.1 for b , 1.6 for c and 4.1 for v). In the case of a magnitude of 4, all curves give a similar intensity until 35 km, after which the curve related to the magnitude errors attenuates the fastest. In the case of a magnitude of 5, the differences between the present tests are small in comparison with the Musson (2005) equation, which gives systematically larger values beyond 50 km. According to the Musson (2005) equation, the attenuation of small intensities is very slow. There are

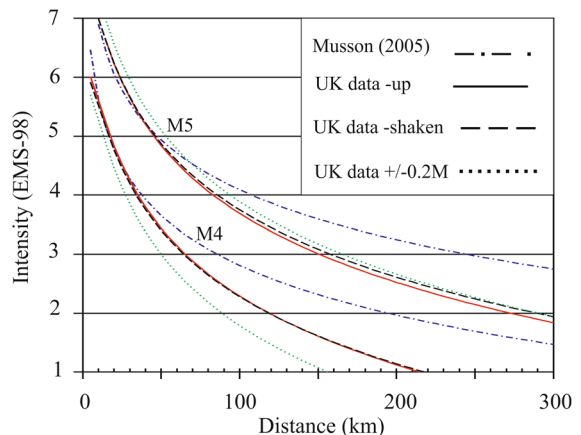


Figure 8

Comparison of the Musson (2005) intensity prediction equation with the present tests for magnitudes M_L 4 and 5 and a depth of 10 km. The uncertain intensities of the UK data were rounded up, the UK earthquakes were removed from the data one at a time (Table 4) and magnitude errors of ± 0.2 were included

no data beyond 300 km in the current dataset (Fig. 5), which advocates faster attenuation.

Synthetic data help to analyse IPEs, although they are idealized. The synthetic tests suggest that even modest numbers of IDPs give correct coefficients: the synthetic database included only four different magnitudes in a narrow range. There is no principal reason for not using small-to-moderate magnitude earthquakes to construct IPEs in the absence of large earthquakes, but the errors in real data complicate the sound evaluation of coefficients.

To conclude, this investigation calls attention to basic data properties. Sophistication is no attribute of intensity, so it is proposed to investigate the effect of uncertain intensities in each dataset instead of using decimals. The synthetic data suggest that small-to-moderate earthquakes can be used in constructing IPEs. The performance of synthetic data gives a model with which the real data can be compared. The attenuation coefficient is insensitive to small magnitude errors, so its large variation may tell of the presence of large magnitude errors. An erroneously assumed regional depth may lead to unusual patterns of intensities as a function of depth. Intensity data should not be downloaded from the available databases without critical revision.

Acknowledgements

Open access funding provided by University of Helsinki including Helsinki University Central Hospital. Authors R.N.V. and R.E.T. acknowledge the financial support under the State Task 0144-2014-0100. Author P.M. is thankful for the international mobility grants of the Academy of Finland related to this bilateral cooperation. The two reviewers are thanked for their thorough work which helped to improve the manuscript considerably.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

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

REFERENCES

- Albarelo, D., & D'Amico, V. (2004). Attenuation relationship of macroseismic intensity in Italy for probabilistic seismic hazard assessment. *Bollettino di Geofisica Teorica ed Applicata*, 45(4), 271–284.
- Allen, T. I., Wald, D. J., & Worden, B. C. (2012). Intensity attenuation for active crustal regions. *Journal of Seismology*, 16, 409–433.
- Ambraseys, N. N. (1985). Intensity-attenuation and magnitude-intensity relationships for Northwest European earthquakes. *Earthquake Engineering and Structural Dynamics*, 13, 733–778.
- Bakun, W. H. (2006). Estimating locations and magnitudes of earthquakes in southern California from Modified Mercalli Intensities. *Bulletin of the Seismological Society of America*, 96(4A), 1278–1295.
- Bakun, W. H., Johnston, A. C., & Hopper, M. G. (2003). Estimating locations and magnitudes of earthquakes in Eastern North America from Modified Mercalli Intensities. *Bulletin of the Seismological Society of America*, 93(1), 190–202.
- Bakun, W. H., & Scotti, O. (2006). Regional intensity attenuation models for France and the estimation of magnitude and location of historical earthquakes. *Geophysical Journal International*, 164, 596–610.
- Bakun, W. H., & Wentworth, C. M. (1997). Estimating earthquake location and magnitude from seismic intensity data. *Bulletin of the Seismological Society of America*, 87(6), 1502–1521.
- Bakun, W. H., & Wentworth, C. M. (1999). Erratum to estimating earthquake location and magnitude from seismic intensity data. *Bulletin of the Seismological Society of America*, 89, 557.
- Bindi, D., Parolai, S., Oth, A., Abdrakhmatov, K., Muraliev, A., & Zschau, J. (2011). Intensity prediction equations for Central Asia. *Geophysical Journal International*, 187, 327–337.
- Bossu, R., Scotti, O., Cotton, F., Cushing, M., & Levret, A. (2000). Determination of geomechanical site effects in France from macroseismic intensities and reliability of macroseismic magnitude of historical events. *Tectonophysics*, 324, 81–110.
- Braze, R. J. (1972). Attenuation of Modified Mercalli intensities with distance for the United States east of 106°W. *Earthquake Notes*, 43(1), 41–52.
- Davis, P. M., Rubinstein, J. L., Liu, K. H., Gao, S. S., & Knopoff, L. (2000). Northridge earthquake damage caused by geologic focusing of seismic waves. *Science*, 289(5485), 1746–1750.
- Faccioli, E., & Cauzzi, C. (2006). Macroseismic intensities for seismic scenarios, estimated from instrumentally based correlations. In Proceedings of the First European Conference on Earthquake Engineering and Seismology, Geneva, Switzerland, 3–8 September, Paper no. 569.
- Fäh, D., Giardini, D., Bay, F., Bernardi, F., Braunmiller, J., et al. (2003). Earthquake catalogue of Switzerland (ECOS) and the related macroseismic database. *Eclogae Geologicae Helvetiae*, 96, 219–236.

- Gasperi, P., & Ferrari, G. (1997). Stima dei parametrici sintetici. In E. Boschi, E. Guidoboni, G. Ferrari, G. Valensise, & P. Gasperini (Eds.), *Catologo dei forti terremoti in Italia dal 461 a.C. al 1990* (pp. 56–64). Rome: ING-SGA Publishing.
- Gasperini, P. (2001). The attenuation of seismic intensity in Italy: A bilinear shape indicates the dominance of deep phases at epicentral distances longer than 45 km. *Bulletin of the Seismological Society of America*, 91(4), 826–841.
- Gómez, A. A. (2006). Seismic hazard map for the Italian territory using macroseismic data. *Earth Sciences Research Journal*, 10(2), 67–90.
- Grünthal, G. (Ed.) (1998). European Macroseismic Scale 1998. Cahiers du Centre Européen de Géodynamique et de Séismologie 15, Luxembourg.
- Gupta, I. N., & Nuttli, O. W. (1976). Spatial attenuation of intensities for central U.S. earthquakes. *Bulletin of the Seismological Society of America*, 66(3), 743–751.
- Hinzen, K.-G., & Oemisch, M. (2001). Location and magnitude from seismic intensity data of recent and historic earthquakes in the northern Rhine area, central Europe. *Bulletin of the Seismological Society of America*, 91(1), 40–56.
- Kondorskaya, N. V., & Shebalin, N. V. (Eds.). (1982). *New catalog of strong earthquakes in the territory of the Soviet Union from ancient times till 1977*. Boulder: National Oceanic and Atmospheric Administration.
- Lapajne, J. (1987). A simple macroseismic attenuation model. *Geologija*, 30, 391–409.
- Levret, A., Backe, J. C., & Cushing, M. (1994). Atlas of macroseismic maps for French earthquakes with their principal characteristics. *Natural Hazards*, 10, 19–46.
- Magri, L., Mucciarelli, M., & Albarello, D. (1994). Estimates of site seismicity rates using ill-defined macroseismic data. *Pure and Applied Geophysics*, 143(4), 617–632.
- Mäntyniemi, P., Tatevossian, R. E., & Tatevossian, T. N. (2014). Uncertain historical earthquakes and seismic hazard: Theoretical and practical considerations. *Geomatics, Natural Hazards and Risk*, 5(1), 1–6. <https://doi.org/10.1080/19475705.2012.751633>.
- Mäntyniemi, P., Tatevossian, R. E., & Vakarchuk, R. N. (2017). On the epicenter determination of historical earthquakes attested to by sparse intensity data sets. *Bulletin of the Seismological Society of America*, 107(3), 1136–1146.
- Mayer-Rosa, D., & Schenk, V. (Eds.) (1989). Test regions for evaluation of algorithms for seismic hazard assessment in Europe (TERESA). *Natural Hazards*, 2(3&4), 215–404 (Special Issue).
- McGuire, R. K. (Ed.) (1993). *The practice of earthquake hazard assessment*. Denver: International Association of Seismology and Physics of the Earth's Interior and European Seismological Commission.
- Mucciarelli, M., Peruzza, L., & Caroli, P. (2000). Tuning of seismic hazard estimates by means of observed site intensities. *Journal of Earthquake Engineering*, 4(2), 141–159.
- Musson, R. M. W. (1996). Determination of parameters for historical British earthquakes. *Annali di Geofisica*, 39(5), 1041–1047.
- Musson, R. M. W. (1998). Intensity assignments from historical earthquake data: Issues of certainty and quality. *Annali di Geofisica*, 41(1), 79–91.
- Musson, R. M. W. (2005). Intensity attenuation in the U.K. *Journal of Seismology*, 9, 73–86.
- Musson, R. M. W., Grünthal, G., & Stucchi, M. (2010). The comparison of macroseismic intensity scales. *Journal of Seismology*, 14, 413–428.
- Pantea, A. (1994). Macroseismic intensity attenuation for crustal sources on Romanian territory and adjacent areas. *Natural Hazards*, 10, 65–72.
- Pasolini, C., Gasperini, P., Albarello, D., Lolli, B., & D'Amico, V. (2008a). The attenuation of seismic intensity in Italy, Part I: theoretical and empirical backgrounds. *Bulletin of the Seismological Society of America*, 98(2), 682–691.
- Pasolini, C., Albarello, D., Gasperini, P., D'Amico, V., & Lolli, B. (2008b). The attenuation of seismic intensity in Italy, Part II: Modeling and validation. *Bulletin of the Seismological Society of America*, 98(2), 692–708.
- Peruzza, L. (1996). Attenuating intensities. *Annali di Geofisica*, 39(5), 1079–1093.
- Rotondi, R., & Zonno, G. (2004). Bayesian analysis of a probability distribution for local intensity attenuation. *Annals of Geophysics*, 47, 1521–1540.
- Shebalin, N. V., Leydecker, G., Mokrushina, N. G., Tatevossian, R. E., Erteleva, O. O., & Vassiliev, V. J. (1998). Earthquake catalogue for central and southwestern Europe (342 B.C.–1990 A.D.) European Commission, Rep. No. ETNU CT 93-0087, Brussels.
- Sørensen, M. B., Stromeyer, D., & Grünthal, G. (2009). Attenuation of macroseismic intensity: A new relation for the Marmara Sea Region, Northwest Turkey. *Bulletin of the Seismological Society of America*, 99(2A), 538–553.
- Sørensen, M. B., Stromeyer, D., & Grünthal, G. (2010). Intensity attenuation in the Campania region, Southern Italy. *Journal of Seismology*, 14, 209–223.
- Stromeyer, D., & Grünthal, G. (2009). Attenuation relationship of macroseismic intensities in central Europe. *Bulletin of the Seismological Society of America*, 99(2A), 554–565.
- Tsapanos, T. M., Galanis, O. Ch., Koravos, G. Ch., & Musson, R. M. W. (2002). A method for Bayesian estimation of the probability of local intensity for some cities in Japan. *Annals of Geophysics*, 45, 657–671.
- Worden, C. B., Wald, D. J., Allen, T. I., Lin, K., Garcia, D., & Cua, G. (2010). A revised ground-motion and intensity interpolation scheme for ShakeMap. *Bulletin of the Seismological Society of America*, 100(6), 3083–3096.