Bindi, D., Parolai, S., Oth, A., Abdrakhmatov, K., Muraliev, A., Zschau, J. (2011): Intensity prediction Equations for Central Asia. - Geophysical Journal International, 187, 1, 327-337

DOI: 10.1111/j.1365-246X.2011.05142.x

# Intensity Prediction Equations for Central Asia

D. Bindi [1,2], S. Parolai[3], A. Oth[4], K Abdrakhmatov[5], A. Muraliev[5], J. Zschau[3]

[1] Center for Disaster Management and Risk Reduction Technology, Helmholtz Centre Postdam, GFZ German Research Centre for Geosciences, Potsdam, Germany
[2] Istituto Nazionale di Geofisica e Vulcanologia, Milano, Italy
[3] Earthquake Risk and Early Warning Section, Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Potsdam, Germany.
[4] European Center for Geodynamics and Seismology, Walferdange, Luxembourg
[5] Institute of Seismology, National Academy of Science, Bishkek, Kyrgyz Republic

**Summary**
In this study, new Intensity Prediction Equations (IPEs) are derived for Central Asia, considering about 6000 intensity data points from 66 earthquakes encompassing the surface-wave magnitude range 4.6 to 8.3. The suitability of the functional form used for constructing the model is assessed by comparing its predictions with those achieved through a non-parametric model. The parametric regressions are performed considering different measures of the source-to-site distance, namely the hypocentral, epicentral and the extended distance metrics. The latter is defined as the minimum distance from the site to a line crossing the epicenters, oriented along the strike of the earthquake and having a length estimated from the event's magnitude. Although the extended distance is introduced as a preliminary attempt to improve the prediction capability of the model by considering the finiteness of the fault extension, the standard deviation of the residual distribution obtained considering the extended distance ($\sigma$=0.734) does not show an improvement with respect to the results for the epicentral distance ($\sigma$=0.737). The similarity of the two models in term of average residuals is also confirmed by comparing the inter-event errors obtained for the two regressions, obtaining very similar values for all earthquakes but the 1911, M 8.2 Kemin event. In particular, different evidences suggest that the magnitude of this event could be overestimated by about half a magnitude unit. Regarding the variability of the residual distribution, all the three considered components (i.e. inter-event, inter-location and record-to-record variances) are not negligible, although the largest contribution is related to the record-to-record variability, suggesting that both source and propagation as well as site effects not captured by the considered model influence the spatial variability of the intensity values..

**Abbreviate title**: Intensity Prediction Equation for Central Asia
**Key words**: macroseismic intensity, prediction equation, Central Asia

## Introduction

In order to be able to assess the seismic hazard of any given region, the expected severity of the ground shaking generated by future earthquakes has to be estimated starting from the knowledge of the seismic history of the studied area. Although a characterization of ground shaking in terms of instrumental parameters (e.g. peak ground acceleration, peak ground velocity, spectral ordinates, etc.) is preferable for developing a quantitative approach, the seismic catalogs used for hazard assessment generally include earthquakes that occurred prior to the instrumental period. The size of the historical earthquakes is usually provided in terms of macroseismic intensity, assigned according to some scale like MMI, MSK-64 or EMS98 (Musson et al., 2010). Since the intensity levels are assigned according to the description of the effects generated by an earthquake, intensity values are qualitative in nature. On the other hand, being also related to effects induced on infrastructures, the intensity values are informative about the damage potential of an earthquake. Moreover, many computer codes used for seismic risk calculations are working on damage probability matrixes and in that case, they are intensity-based (e.g. Tyagunov et al, 2006; Colombi et al., 2010). Finally, in many countries the building codes account for the seismic excitations in terms of macroseismic intensity. From all the above considerations it clearly follows that macroseismic intensities are still widely used in seismic hazard assessment studies (e.g., Garcia-Mayordomo et al., 2004). Since in turn seismic hazard assessment requires attenuation models developed for the ground motion parameter of interest, several Intensity Prediction Equations (IPEs) have been developed worldwide (e.g., Allen and Wald, 2009; Cua et al., 2010).
In this study, we develop new IPEs for Central Asia, exploiting the macroseismic data set collected within the project Establishment of the Central Asia Seismic Risk Initiative (CASRI). Different approaches have been proposed for developing an IPE, such as the regression approach (among many others, Sørensen et al., 2009) or the fully probabilistic approach (e.g. Pasolini et al., 2008b). In this work, we follow a standard regression approach where the parameters of the considered model are determined by evaluating the best least squares fit to the set of observed intensities. Different models are calibrated, considering a non-parametric regression approach as well as parametric models constructed for different definitions of the source-to-site distance. In particular, we consider both epicentral and hypocentral distances as well as a simplified version of the Joyner-Boore distance metric. The residuals between observations and predictions are computed for all developed models and the variability is discussed in terms of inter-event and inter-location distribution of error.

## Data

The catalogue used in this work to develop the Intensity Prediction Equations (IPE) for Central Asia (CA) is composed of 66 earthquakes encompassing the surface-wave magnitude range 4.6 to 8.3. The magnitude is measured according to the Moscow-Prague formula (Karnik 1962), and it is generally indicated as $M_{LH}$ in the Russian scientific literature. For a thorough discussion on the definition of this magnitude scale and its relationship with the surface-wave magnitude introduced by Gutenberg (1945), we refer the reader to Bormann (2002) and the references cited therein. In the former Soviet Union

countries, the size of earthquakes recorded by local and regional networks was routinely quantified by computing the energy-class K (Rautian 1960; Rautian et al., 2007), which is based on the sum of the maximum amplitudes of P- and S-waves. The energy-class K for the earthquakes considered in this study ranges from 12 to 18.5.

The locations of the 66 earthquakes are shown in **Figure 1**. The majority of the strongest events (e.g., 1887, M 7.3 Verny earthquake; 1889, M 8.3 Chilik earthquake; 1911, M 8.2 Kemin; 1938 M 6.9 Kemin-Chu earthquakes) occurred in the Northern Tian Shan seismic zone, along the Chon-Kemin-Chilik fault zone (e.g. Kalmetieva et al., 2009). Among the most recent events, the analyzed catalogue includes the felt report of the M 7.3 Suusamyr earthquake (e.g. Mellors et al., 1997), which occurred in 1992 north-east of the Talas-Fergan strike slip fault zone (e.g. Kalmetieva et al., 2009), in an area considered seismically safe before the occurrence of that earthquake.

From these earthquakes, about 6000 intensity points are available in the MSK-64 (Medvedev et al. 1964) intensity range 2-9.5. It is worth noting that in this work, the intensity values are processed as real numbers and the uncertain data (e.g. VII-VIII) are accounted for by introducing half-degree values (e.g., 7.5). Alternative approaches are those considering the intensity value as a real number but disregarding the ill-defined data (e.g. Sørensen et al., 2009) or those preserving the discrete nature of the intensity data (e.g. Pasolini et al 2008a) and describing the ill-defined data through a probability density function (e.g. Magri et al., 1994; Pasolini et al 2008b), following a fully probabilistic approach.

The magnitude and intensity versus epicentral distance scatter-plots are shown in **Figure 2**, whereas **Figure 3** shows the histograms of the magnitude, distance and intensity data distributions. The magnitudes are quite well distributed with distance over the range 1 to 600 km (Figure 2 left), with a decreasing number of observations for increasing distance (Figure 3, left). Most of the data points correspond to magnitudes in the range 5-7.5 (Figure 3, middle) and the best sampled intensity interval is ranging from 4 to 6 (Figure 3, right).

The un-sampled lower left triangle in the intensity versus distance distribution (Figure 2, right) is a consequence of the smallest Io value present in the catalogue, which is equal to 5.5, where Io is the maximum observed intensity present in the felt report of any given earthquake. This is due to the fact that for earthquakes causing small macroseismic intensities only, the intensity distribution is usually not mapped. Since such events are those associated to small intensity values at short distances, the lack of such events in the catalogue causes the aforementioned un-sampled triangle in the intensity versus distance distribution.

The scatter-plots in Figure 2 are color-coded according to either the energy-class (left) or the magnitude of the earthquake associated to the intensity point (right). The distribution of the K values against magnitude for the 66 considered earthquakes is shown in **Figure 4**. The linear correlation shown by M and K can be modeled as $K=a+bM$. Rautian et al. (2007) indicated that the accuracy of K-class values estimated at different stations is about 0.35 units, while the uncertainty on M can be assumed to range between 0.3 and 0.5 (e.g., Delvaux et al., 2001; Kondorskaya and Shebalin, 1982). Under these conditions, although the uncertainty affecting each single magnitude and K-class value is not known, it can be assumed that they are of the same order of magnitude. An orthogonal least-squares regression between K-class and M, performed assuming the same variance over

the two variables, led to the model K=4.42+1.70M (Figure 4, black line). The best-fit line is compared to the model used by Abdrakhmatov et al (2003) to convert K into M for their probabilistic assessment of the seismic hazard in Kyrgyzstan. The two models show a good agreement over the M and K ranges well constrained by observations. For the largest magnitude (M=8.3), the difference between K values predicted by the two models is about half K unit whereas, for the 1911 Kemin earthquake (M 8.2), either the magnitude appears to be overestimated or the K-class underestimated. However, it is worth noting that the magnitude of this earthquake is not well constrained and different references provide varying estimates. For example, Kanamori (1977) listed the values Mw=7.7 and Ms=8.4 while the value M=7.8 is found on the USGS webpage for "Today in earthquake history", taken from Kondorskaya and Shebalin (1982).

The considered intensity data are compared to the predictions obtained by applying the model proposed by Nazarov and Shebalin (1975) for Kazakhstan and Kirgizstan (see also Nurmagambetov et al., 1999):

$$I = 1.5M - 3.8\log\sqrt{R_{epi}^2 + h^2} + 3.6, \quad (1)$$

where $R_{epi}$ is the epicentral distance and $h$ the hypocentral depth. Hereinafter we refer to the model in equation (1) as NS75.

**Figure 5** shows the observed minus predicted residuals against distance (top left) and magnitude (bottom left). The average residual (bias) is -0.339, hinting at an average over-prediction of the model, with a standard deviation of the residual distribution equal to 0.818. When plotted against magnitude and distance, the overestimation is more evident for short distances and large magnitudes, whereas intensities for magnitudes smaller than 5.5 are underpredicted. The likelihood approach proposed by Scherbaum et al (2004) can be applied to rank the capability of the NS75 model in predicting the observations considered in this work. Following this approach, the normalized residuals $z_0$ are computed by normalizing each residual to the standard deviation of the prediction model. Then, the goodness of fit is measured by computing the probability LH(|z0|) for the absolute value of a random sample from the normalized distribution to fall between the modulus of a particular observed residual $z_0$ and $\infty$, considering both tails of the error distribution Erf(z) (equation (9) in Scherbaum et al., 2004), that is

$$LH(|z_0|) = Erf\left(\frac{|z_0|}{\sqrt{2}}, \infty\right) = \frac{2}{\sqrt{\pi}} \int_{|z_0|/\sqrt{2}}^{\infty} e^{-t^2} dt \quad (2)$$

Scherbaum et al. (2004) discussed some properties of the LH distribution that make it a good measure of the goodness of fit. Moreover, they used the median of the LH values and some central tendency parameters (mean and median) of the normalized residual distribution, together with the standard deviation, to define a rank ordering that can be adopted to quantify the suitability of a given attenuation relationship for prediction purposes in a given area. In particular, a uniform distribution of LH values between 0 and 1 is expected when the residuals follow a normal distribution with zero mean and unit variance. Considering the NS75 model, the distribution of LH values shown in Figure 5 (right) is obtained. The median of the LH distribution is 0.49, and considering that the absolute value of the mean and median of the normalized residuals are 0.42 and 0.37, respectively, the NS75 model is ranked in the intermediate prediction capability class (Scherbaum et al., 2004).

**Non-parametric versus parametric regressions**

The data set described in the previous section is now used for developing a new IPE for CA. First, we apply a non-parametric model to describe the source and attenuation characteristics, following a data-driven approach. Considering $N_{data}$ intensity values generated by $N_{eq}$ earthquakes and discretizing the hypocentral distance range into $N_{bin}$ intervals, the observed intensities are modeled as a linear combination of source and attenuation terms as follows:

$$I_k(R_H) = \alpha_i + w_j\beta_j + (1 - w_j)\beta_{j+1}, \quad (3)$$

where $k$=1,.., $N_{data}$ indicates the *k-th* observation, $i$=1,…, $N_{eq}$ indicates the *i-th* earthquake, $j$=1,…, $N_{bin}$+1 indicates the *j-th* distance value $r_j$ selected such that the hypocentral distance $R_H$ of the *k-th* data point is between $r_j \leq R_H \leq r_{j+1}$. The attenuation function is linearized between $r_j$ and $r_{j+1}$ using the weights $w$, computed as $w_j = (r_{j+1} - R_H)/(r_{j+1} - r_j)$ . In this application, the distance range 0 – 400 km in discretized into 40 intervals equal spaced over a logarithm distance scale. In equation (3), $\alpha$ and $\beta$ represent the source and attenuation terms, respectively. Model (3) assumes that the attenuation is independent on the earthquake size and it depends only on distance.

The parameters $\alpha_i$ and $\beta_j$ are determined by performing a least squares fit over the set of observations. In order to remove the trade-off between the source and attenuation terms, the attenuation function is constrained to assume a unit value at 10 km (e.g. Oth et al., 2008). Moreover, in order to assess the robustness of the results and to evaluate the uncertainties affecting the coefficients $\alpha_i$ and $\beta_j$, 50 bootstrap replications (Efron, 1979) of the original data set are considered. The results of the 50 regressions are in shown as gray circles in **Figure 6** for both the attenuation (top) and source (bottom) terms. For each coefficient $\alpha_i$ and $\beta_j$, the median (red circles) and the standard deviation ($\sigma_{boot}^{atte}$ and $\sigma_{boot}^{sorg}$ for attenuation and source, respectively) of the bootstrap distributions are computed. The histograms of the standard deviation distributions (i.e., for attenuation, the set of standard deviations of all distance bins as obtained from the bootstrap procedure, while for the sources, the set of standard deviations of all individual source terms) are shown in Figure 6 as insets. The coefficients $\beta_j$ describe an almost linear attenuation with log-distance, with a slight increase of the rate of attenuation over distances larger than 300 km. The bootstrap standard deviations are peaked at 0.16. In the bottom panel, the source coefficients $\alpha_i$ are shown against magnitude. The median values outline an almost linear trend with magnitude, with a median bootstrap standard deviation equal to 0.168 (the mode is 0.14). The overall standard deviation of the non-parametric distribution of residuals is 0.60.

Following the results of the non-parametric approach, a parametric regression is also performed, considering a model linear in magnitude and log-distance:

$$I = a_1 M + a_2 - a_3 \log_{10}(R_H/10) - a_4(R_H - 10) \quad (4)$$

The obtained model (Table 1), is depicted in Figure 6 as blue lines. The standard deviation of the residuals for model (4) is 0.710, larger than the one obtained for the non-parametric regression. A fairly good agreement between the non-parametric and parametric results is observed, with a slight tendency of the parametric model to produce slightly smaller source contributions combined with somewhat weaker attenuation. The good agreement between the non-parametric trend with magnitude and the parametric model for the source terms suggests that the magnitudes of the considered earthquakes are consistent. Regarding the M 8.2 Kemin earthquake, in agreement with the evidence shown in Figure 4, the coefficient $\alpha$ for this earthquake suggest an overestimation of the magnitude value.

**Parametric models for epicentral distance**

Most of the IPDs developed worldwide consider either the epicentral distance $R_{epi}$ , the minimum distance from the rupture $R_{rup}$ or ithe Joyner-Boore $R_{JB}$ measure of the source-to-site distance (e. g.Allen and Wald, 2009; Cua et al, 2010). The calculation of $R_{rup}$ and $R_{JB}$ distance metrics requires knowledge of the fault geometry, which is rarely available, especially for historical earthquakes. For recent earthquakes, the fault extension is indirectly estimated from other information but without accounting for the large uncertainties affecting the estimation of the fault parameters when the finite-fault measures of distance are evaluated. Moreover, both $R_{epi}$ and $R_{rup}$ (and $R_{JB}$) suffer from the limitation of considering, for a given site, only a point on the source plane, which is constant for all sites when $R_{epi}$ is evaluated and variable from site to site when $R_{rup}$ or $R_{JB}$ are estimated. The main advantage of $R_{rup}$ and $R_{JB}$ is that they introduce a link to the finite extension of the fault, which in turn modifies the isotropic decay of intensity with distance. On the other hand, there is the possibility of strong over-prediction of intensity values for sites close to the fault (or fault projection on the surface), but far away from the patches on the fault where most of the seismic energy is released.
In this section, we perform a further regression, starting from equation (4) but considering the epicentral distance, that is (e.g. Sorensen et al, 2009):

$$I = a_1 M + a_2 - a_3 \log_{10}\left(\sqrt{\frac{R_{epi}^2 + h^2}{h^2}}\right) - a_4\left(\sqrt{R_{epi}^2 + h^2} - h\right), \qquad (5)$$

where the estimate of the hypocentral depths $h$ is used for each earthquake.
The distributions of the model parameters obtained by performing 100 bootstrap inversions are shown in **Figure 7**. The median values are ($a_1$, $a_2$ ,$a_3$ ,$a_4$)= (0.898, 1.215, 1.809, 3.447 $10^{-3}$) (Table 1) and the variability is mainly controlled by the unavoidable negative trade-off internal to the source (i.e., between a1 and a2) and attenuation (i.e., between a3 and a4) part of the model, as shown in the bottom panel of Figure 7 (and by the negative values of the off-diagonal entries of the covariance matrix, here not shown). The residual distribution obtained considering the median parameters is shown in **Figure 8** (left panels). The predictions are unbiased, without any clear trend with magnitude or distance. Only when computed over different distance bins (vertical bars in Figure 8), a

slightly positive residual is obtained for short distances, in particular for distances smaller than 10 km (average residual 0.4). The standard deviation of the whole distribution of residuals is $\sigma$=0.737, slightly larger than the one obtained using hypocentral distance.
In order to investigate systematic errors in the assignment of the magnitude or related to the locations where the intensity values were surveyed, the median model is used to estimate the average residual for each event (inter-event error, $\varepsilon_{eve}$) and the average residual for each location (inter-location error, $\varepsilon_{loc}$). In the case of inter-location error, the residual distribution has been corrected for the inter-event error before calculating the inter-location error. **Figure 9** shows the distribution of the inter-event and inter-location errors, as well as their spatial variability.
The standard deviation $\sigma_{eve}$ of the inter-event distribution is 0.42, while the inter-location sigma is $\sigma_{loc}$=0.37, which reduces to 0.25 if only locations characterized by at least 5 observations are considered (red circles). Since the total sigma is 0.737, and assuming that the inter-event and inter-location distributions of error are independent, the standard deviation of the record-to-record distribution of error is 0.48. Figure 9 also shows the inter-event distribution of error obtained considering the K-class as explanatory variable instead of magnitude. A similar value for the standard deviation of the inter-event distribution is obtained (0.723), although small differences between errors obtained for some earthquakes exist. For the 1911, M=8.2 Kemin earthquake, the negative error equal to -0.535 obtained considering the magnitude as explanatory variable is a further indication about a possible overestimation of the magnitude for this event. When K is used as explanatory variable, the error for this earthquake is still negative (-0.262), but significantly reduced.
The right panels of Figure 9 show the spatial distribution of $\varepsilon_{eve}$ and $\varepsilon_{loc}$. The error distributions do not show clear spatial patterns, with positive and negative values located in immediate vicinity to each other.

**Parametric models for extended rupture distance**

In order to evaluate the impact of the point source assumption when using epicentral of hypocentral distance, regression (5) is repeated considering an approximation of $R_{JB}$ as distance measure. We approximate the surface projection of the fault with a segment centered on the epicenter, oriented along the strike and having a length estimated from the magnitude by applying the Wells and Coppersmith (1994) relationships. We stress that the aim of this computation is not to compute a precise fault-distance measure, but only to gain some insights about possible improvements of the model when the extension of the fault is considered, at least in one direction. In the remainder of this article, we refer to this distance measure as extended distance $R_{ext}$. $R_{ext}$ and $R_{JB}$ would be identical measures of distance in the case of a vertical fault plane with bilateral rupture propagation. The relation between $R_{epi}$ and $R_{ext}$ is shown in **Figure 10**.
The residuals for the models calibrated using $R_{ext}$ (Table 1) are shown in Figure 8 (right panels). Significant improvements are not observed, even for the strongest earthquakes. The standard deviation of the residual is 0.734, similar to the one obtained when using epicentral distance. This result is also confirmed by the comparison between the inter-event errors obtained for the models relevant to the two distance measures shown in **Figure 11**. The two regressions provide very similar inter-event errors over the entire

magnitude range, with values distributed close to the one-to-one relation. Only for the 1911 Kemin earthquake we have a significant increase of the error when $R_{ext}$ is considered, from -0.535 to -0.879 for $R_{epi}$ and $R_{ext}$, respectively. The increase in the overestimation for this earthquake when $R_{ext}$ is considered could be related to the overestimation of the magnitude, which in turn results in an overestimation of the fault length and therefore in a reduction of the distances. This hypothesis is confirmed by the comparison in the right panel of Figure 11, where the results obtained by arbitrarily reducing the magnitude to 7.7 are shown. In this case, the errors reduce to -0.126 and -0.281 for $R_{epi}$ and $R_{ext}$, respectively.

Single earthquake predictions computed considering models for both $R_{epi}$ and $R_{ext}$ are shown in Figure 12, considering the 1889 M=8.3 Chilik, the 1911 M=8.2 Kemin and the 1938 M=6.9 Kemin-Chu earthquakes. As expected, when the rupture extend is taken into account, the shape of the predicted macroseismic field is strongly modified with respect to the isotropic decay obtained for $R_{epi}$. Figure 12 also shows the observation minus prediction residuals for both models. For the 1889 Chilik earthquake, the usage of $R_{ext}$ slightly improves the estimation of intensity for locations close to the upper limit of the fault segment but on average, the spatial distribution of residuals looks very similar. For the 1911 Kemin earthquake, Figure 12 confirms the general overprediction of the models while for the 1938 Kemin-Chu event, only small modifications are obtained when using the extended distance measure.

Finally, in Figure 13 the IPE curves versus epicentral distance are shown for 4 different magnitudes, and compared with data. For magnitude M=5 and 6, the depth $h$ has been fixed to 15 km and observations for magnitude within the range M±0.1 are selected. For the M=8.2 and 8.3 cases, the depth of the 1887 and 1911 earthquakes were used, respectively. In general, the median and standard deviation of the predictions provide a good representation of the average trend and variability of the observations. The overestimation observed for the M 8.2 event is in agreement with the findings discussed previously. Furthermore, the IPE curves proposed in this study seem to better capture the trend in the data than the NS75 model.

## Conclusions

In this study, we derived new Intensity Prediction Equations (IPEs) for Central Asia. Starting from the results obtained by performing a non-parametric regression, a parametric model linear in magnitude and log-distance has been regressed against about 6000 intensity data points, distributed over the magnitude range 4.6-8.3. Different measurements of distance were considered, namely the hypocentral, epicentral and a measurement of distance accounting for the finite length of the fault, refereed to as extended distance. In terms of observed minus predicted values, all the models perform equally well, without any significant dependence of bias on magnitude and distance. In particular, the standard deviation of the residual distribution obtained considering the extended distance ($\sigma$=0.734) is very similar to the value obtained for the epicentral distance ($\sigma$=0.737). The equivalence of the two models in term of average residuals is also confirmed by comparing the inter-event errors obtained for the two regressions, obtaining very similar values for all earthquakes but the 1911 Kemin event. Indeed, the outcomes of several analyses performed in this study suggest a potential overestimation

of the magnitude for this earthquake of the order of about half a magnitude unit. Regarding the variability of residuals, the three components of variance (i.e. the inter-event, inter-location and record-to-record variances) play a similar role, although the largest contribution is related to the record-to-record variability, probably as a consequence of the large uncertainty generally affecting each single intensity data point. Finally, the inter-location errors did not shown any clear spatial pattern that could be directly related to site effects.
The application of the IPEs derived in this study for estimating the seismic hazard and risk in Central Asia is ongoing (e.g., Bindi et al., 2011). Since for such applications a model with a fixed depth is sometimes required, Table 1 also includes the coefficients obtained when regressing equation (4) with the depth *h* fixed to 15 km.

**Acknowledgments**
We thank Gottfried Grünthal and Dietrich Stromeyer for very useful and constructive suggestions which appreciably improved the article. Comments and suggestions from the Editor E. Fukuyama and from two reviewers are also acknowledged. This work has been carried out in the framework of Earthquake Model Central Asia (EMCA) project, partially funded by the Global Earthquake Model (GEM) foundation. Data has been provided by the Central Asia Seismic Risk Initiative (CASRI) project, funded by ISTC (project No KR 1176). The first author benefited from a grant funded by Centre for Disaster Management and Risk Reduction Technology (CEDIM).

**References**
Abdrakhmatov, K., H.-B. Havenith, D. Delvaux, D. Jongmans & P. Trefois 2003. Probabilistic PGA and Arias Intensity maps of Kyrgyzstan (Central Asia), *Journal of Seismology* 7, 203–220

Allen, T.I., & D. J. Wald 2009. Evaluation of ground-motion modeling techniques for use in Global ShakeMap—A critique of instrumental ground-motion prediction equations, peak ground motion to macroseismic intensity conversions, and macroseismic intensity predictions in different tectonic settings: U.S. Geological Survey Open-File Report 2009–1047, 114 p.

Bindi D., Mayfield M., Parolai S., S. Tyagunov, U. T. Begaliev, K. Abdrakhmatov, B. Moldobekov, & J. Zschau 2011. Towards an improved seismic risk scenario for Bishkek, Kyrgyz Republic, *Soil Dyn. Earth. Eng*. **31**, 521-525 doi:10.1016/j.soildyn.2010.08.009

Bormann, P. 2002. Magnitude of seismic events, in IASPEI, New Manual of Seismological Observatory Practice, P. Bormann (Editor), Vol. 1, GeoForschungsZentrum, Potsdam, 16–50.

Colombi, M., Crempien, J., Crowley, H., Erduran, E., Liu, H., Lopez, M., Mayfield, M., & M. Milanesi 2010. Evaluation of seismic risk software for GEM, GEM Technical Report 9. GEM Foundation, Pavia, Italy

Cua G., Wald D. J., Allen T. I., Garcia D., Worden C.B, Gerstenberger M., Lin K., & Marano K. 2010. „Best Practices" for Using Macroseismic Intensity and Ground Motion-Intensity Conversion Equations for Hazard and Loss Models in GEM1, GEM Technical Report 2010-4, GEM Foundation, Pavia, Italy

Delvaux, D., Abdrakhmatov, K.E. & A. L. Strom 2001. Landslides and surface breaks of the Ms 8.2 Kemin earthquake, Kyrgyzstan, *Russian Geology and Geophysics, Geologya i Geophysika* **42**, 1167–1177.

Efron, B. (1979). Bootstrap methods, another look at the jackknife, *Ann. Stat.* **7**, 1–26.

Garcia-Mayordomo, J., E. Faccioli & R. Paolucci 2004. Comparative study of the seismic hazard assessments in European national seismic codes, *Bull Earthquake Engng* **2**, 51 - 73

Gutenberg, B. 1945. Amplitudes of surface waves and magnitudes of shallow earthquakes, *Bull. Seismol. Soc. Am.* **35**, 3–12

Kalmetieva Z.A., Mikolaichuk A.V., Moldobekov B.D., Meleshko A.V., Jantaev M.M. & A. V. Zubovich 2009. Atlas of earthquakes in Kyrgyzstan, ISBN 978-9967-25-829-7

Kanamori, H. 1977. The energy release of great earthquakes, *J. Geophys. Res*. **82**, 2981-2987.

Karnik, V. 1962. Amplitude-distance curves of surface waves at short epicentral distances, *Studia Geophys. Geod.* **6**, 340–346

Kondorskaya, N.V., & Shebalin, N.V., 1982. (Eds.), New Catalog of Strong Earthquakes in the U.S.S.R. from Ancient Times through 1977. National Oceanic and Atmospheric Administration, Boulder, USA, 608 p.

Magri, L., M. Mucciarelli, & D. Albarello 1994. Estimates of site seismicity rates using ill-defined macroseismic data, *Pure Appl. Geophys.* **143**, 618–632

Medvedev, S., W. Sponheuer, & V. Kárník 1964. Neue seismische Skala Intensity scale of earthquakes, 7. Tagung der Europäischen Seismologischen Kommission vom 24.9. bis 30.9.1962. In: Jena, Veröff. Institut für Bodendynamik und Erdbebenforschung in Jena, vol 77. Deutsche Akademie der Wissenschaften zu Berlin, pp 69-76

Mellors, R. J., F. L. Vernon, G. L. Pavlis, G. A. Abers, M. W. Hamburger, S. Ghose, & B. Iliasov, 1997. The Ms=7.3 Suusamyr, Kyrgyzstan earthquake:1: Constraints on fault

geometry and source parameters based on aftershocks and body-wave modeling, *Bull. Seismo. Soc. Amer.*, **87**,11-22.

Musson, R., G. Grünthal, & M. Stucchi 2010. The comparison of macroseismic intensity scales, *Journal of Seismology* **14**, 413-428.

Nazarov A.G. & N.V. Shebalin (Eds) 1975. The seismic scale and methods of measuring seismic intensity, Moscow, (in Russian).

Nurmagambetov A., N. Mikhailova, & W. Iwan 1999. Seismic hazard of the Central Asia region, in Seismic hazard and building vulnerability in post-Sovietic Central Asian republics, S.A. King, V. I. Khalturin & B. E. Tucker (eds), Kluwer Academic Publishers, Nederlands

Oth, A., Bindi, D., Parolai, S. & F. Wenzel 2008. S-wave attenuation characteristics beneath the Vrancea region in Romania: new insights from the inversion of ground-motion spectra, Bull. Seism. Soc. Am., **98**(5), 2482–2497, doi:10.1785/0120080106.

Pasolini, C., P. Gasperini, D. Albarello, B. Lolli, & V. D'Amico 2008a. The attenuation of seismic intensity in Italy, part I: Theoretical and empirical backgrounds. *Bull. Seism. Soc. Am* **98**, 682–691.

Pasolini, C., D. Albarello, P. Gasperini, V. D'Amico, & B. Lolli, 2008b. The attenuation of seismic intensity in Italy, part II: Modeling and validation. *Bull. Seism. Soc. Am* **98**, 682–691.

Rautian, T. G. (1960). Energy of earthquakes. In Methods for the Detailed Study of Seismicity, ed. Y. V. Riznichenko, 75–114. Moscow: Izdatel'stvo Akademii Nauk SSSR (in Russian)

Rautian,T., Khalturin K. Fujita, K. G. Mackey, & A. D. Kendall 2007. Origins and Methodology of the Russian Energy K-Class System and its Relationship to Magnitude Scales, *Seismological Research Letters* **78**, 579-590.

Scherbaum, F., F. Cotton, & P. Smit 2004. On the use of response spectral-reference data for the selection and ranking of ground-motion models for seismic-hazard analysis in regions of moderate seismicity: the case of rock motion, *Bull. Seism. Soc. Am.* **94,** 2164–2185.

Sørensen, M., D. Stromeyer, & G. Grünthal 2009. Attenuation of macroseismic intensity: a new relation for the Marmara Sea region, northwest Turkey, *Bull. Seism. Soc. Am.* **99,** 538-553

Sørensen, M., D. Stromeyer, & G. Grünthal 2010. A macroseismic intensity prediction equation for intermediate depth earthquakes in the Vrancea region, Romania. *Soil Dynamics and Earthquake Engineering* **30**, 1268-1278.

Tyagunov, S., G. Grünthal, R.Wahlström, L. Stempniewski, & J. Zschau 2006. Seismic risk mapping for Germany. *Natural Hazards and Earth System Sciences* **6**, 573–86.

Wells, D., & K. Coppersmith (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. Seism. Soc. Am*. 84, 974–1002.

**Table 1**
Coefficient of the models derived for different definition of distance

| model | a1 | a2 | a3 | a4 | $\sigma$ |
|---|---|---|---|---|---|
| Hypocentral distance, Eq. (4) | 1.071 | 1.003 | 2.621 | $5.567\ 10^{-4}$ | 0.710 |
| Epicentral distance, Eq. (5) | 0.898 | 1.215 | 1.809 | $3.447\ 10^{-3}$ | 0.737 |
| Extended distance, Eq. (5) but for $R=R_{ext}$ | 0.788 | 1.764 | 1.898 | $2.673\ 10^{-3}$ | 0.734 |
| Epicentral distance and fixed depth Eq. (5) with h=15km | 1.049 | 0.686 | 2.706 | $1.811\ 10^{-4}$ | 0.689 |

**Figure Captions**

**Figure 1**: Map depicting the earthquake epicenters

**Figure 2**: Magnitude (left) and intensity (right) versus distance scatter-plots. The circles are color coded according to the K-class value (left) and magnitude range (right).

**Figure 3**: Histograms of the epicentral distance (left), magnitude (middle) and intensity (right) distributions of data.

**Figure 4**. K-class versus magnitude scatter-plot (crosses). The best-fit model obtained in this study (black) as well as the model from Abdrakhmatov et al (2003) (red) are also shown.

**Figure 5**: Residuals versus magnitude (left, top) and epicentral distance (left, bottom) for NS75 and LH distribution (right).

**Figure 6** Top: Bootstrap non-parametric attenuation values (gray circles) compared with attenuation part of the parametric model (blue line). Bottom: Same as top panel considering the source part of the models. The red circles represent the median of the bootstrap results, whereas the histograms of the distribution of the bootstrap standard deviation for all source or attenuation coefficients are shown in the insets.

**Figure 7**. Results of 100 bootstrap regressions performed for the model depending on epicentral distance (see equation 5 in the text). The bottom panels show the internal trade-offs characterizing the source ($a_1$ and $a_2$, left) and attenuation ($a_3$ and $a_4$, right) coefficients.

**Figure 8**. Left: Distribution of the residuals against magnitude (top) and distance (bottom) considering the model using epicentral distance. The vertical bars represent the average ± 1$\sigma$ computed over different distance bins, with the first one being relevant for epicentral distances smaller than 10km. Right: Same as on the left but performing the regression considering the extended distance measure (see text for explanation). The first vertical bar corresponds to the average ± 1$\sigma$ of residuals for extended distances < 1km.

**Figure 9**. Left: Inter-event error against magnitude (top) and K-class (middle) for the model calibrated considering the epicentral distance; inter-location error (bottom) obtained for the considered intensity data points. The red circles indicate the results obtained for localities with more than 5 observations. Right: Spatial distribution of the inter-event (top) and inter-location (bottom) errors shown in the left panels.

**Figure 10** Comparison between epicentral and extended distance measures

**Figure 11** Left: Comparison between inter-event errors obtained considering the epicentral and extended distance measures. Right: The same as on the left, however

lowering the magnitude of the 1911 Kemin earthquake to 7.7. The circles are color-coded according to their magnitude value.

**Figure 12** Comparison of predictions with the derived IPE using the epicentral (left) and extended (right) distance measures. Top: 1889 M 8.3 Chilik earthquake; Middle: 1911 M 8.2 Kemin earthquake; Bottom: 1938 M 6.9 Kemin-Chu earthquake. The colored circles indicate the residual values. The segment used to compute the extended distances is shown in green in the right-hand plots.

**Figure 13** Comparison between observations (crosses) and predictions (blue: this study; red: Nazarov and Shebalin, 1975). Top: Predictions for a magnitude M=5 (left) and M=6(right), considering a depth of 15 km. Observations in the range M±0.1 are selected. Bottom: Comparison for the 1911 M 8.2 (left) and 1889 M 8.3 (right) earthquakes

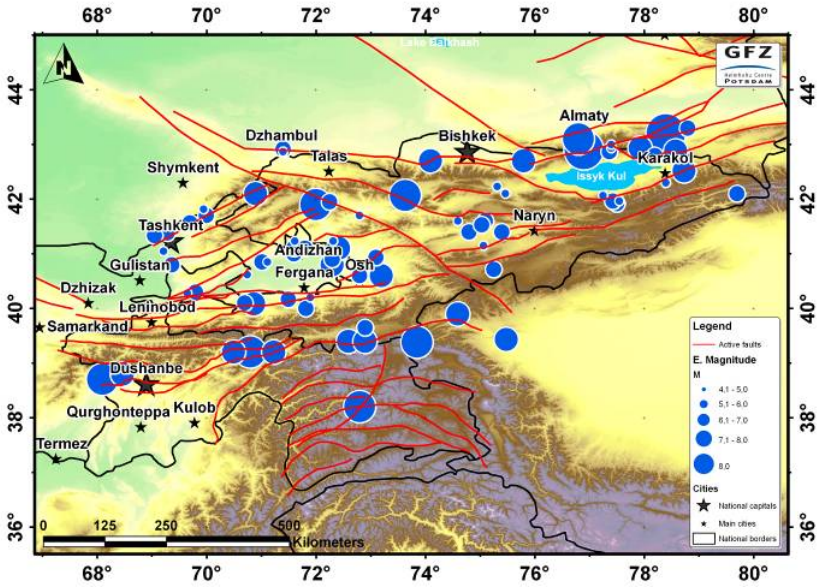


**Figure 1**: Map depicting the earthquake epicenters.

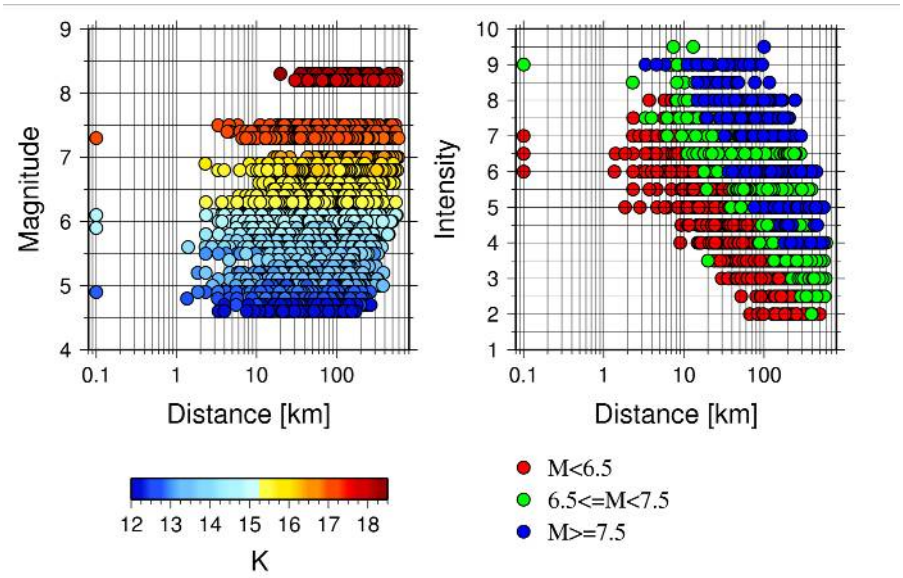


**Figure 2**: Magnitude (left) and intensity (right) versus distance scatter-plots. The circles are color coded according to the K-class value (left) and magnitude range (right).

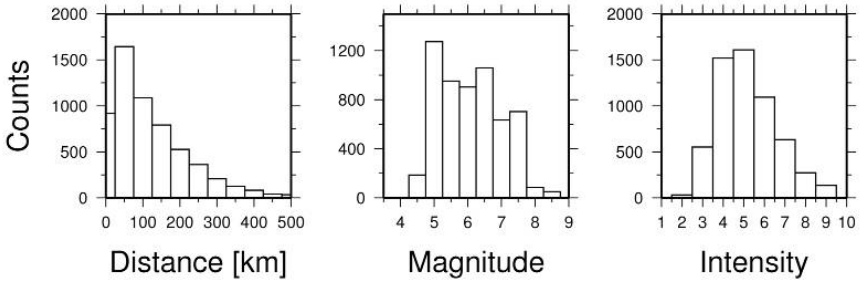


**Figure 3**: Histograms of the epicentral distance (left), magnitude (middle) and intensity (right) distributions of data.

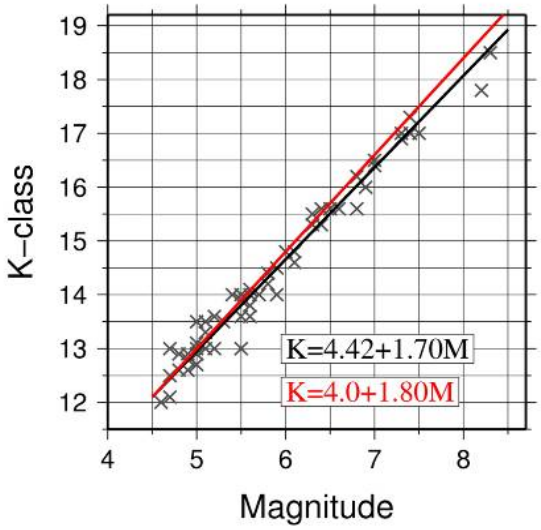


**Figure 4**. K-class versus magnitude scatter-plot (crosses). The best-fit model obtained in this study (black) as well as the model from Abdrakhmatov et al (2003) (red) are also shown.

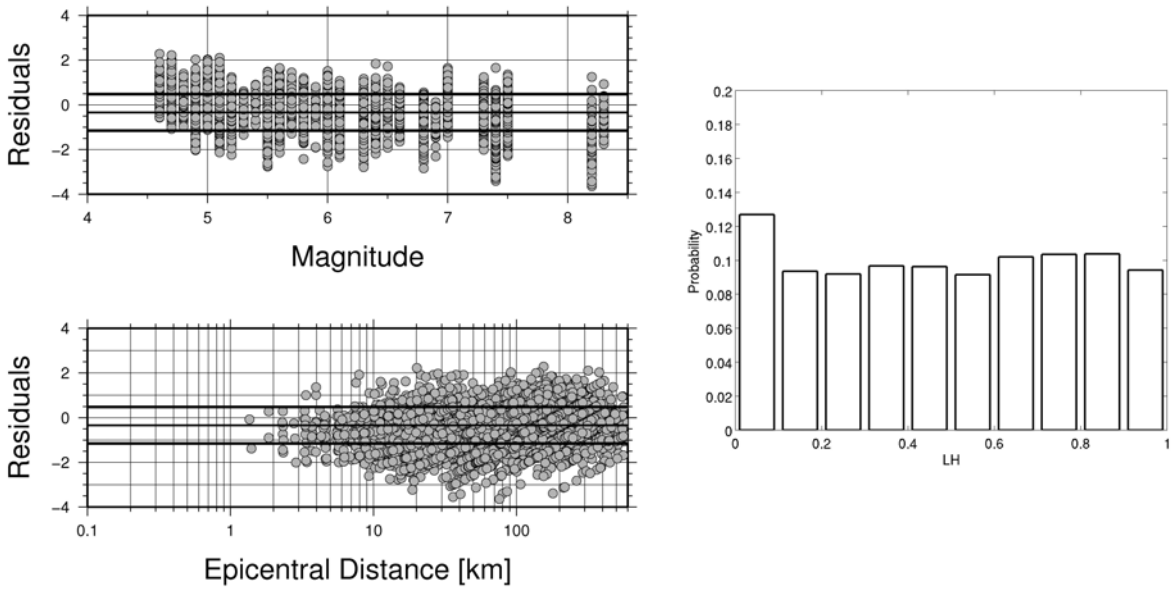


**Figure 5**: Residuals versus magnitude (left, top) and epicentral distance (left, bottom) for NS75 model and LH distribution (right).

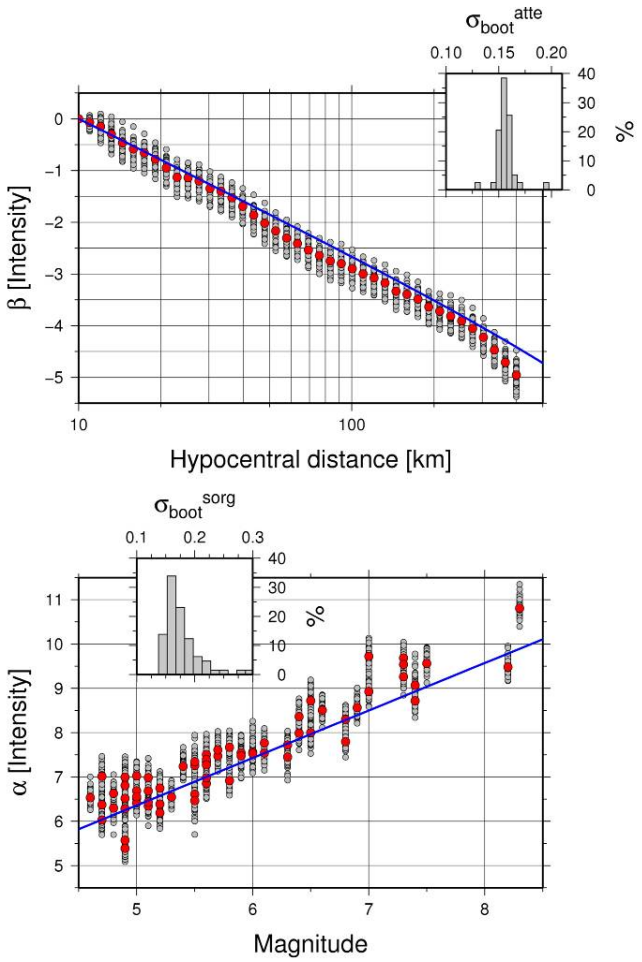


**Figure 6** Top: Bootstrap non-parametric attenuation values (gray circles) compared with attenuation part of the parametric model (blue line). Bottom: Same as top panel considering the source part of the models. The red circles represent the median of the bootstrap results, whereas the histograms of the distribution of the bootstrap standard deviation for all source or attenuation coefficients are shown in the insets.

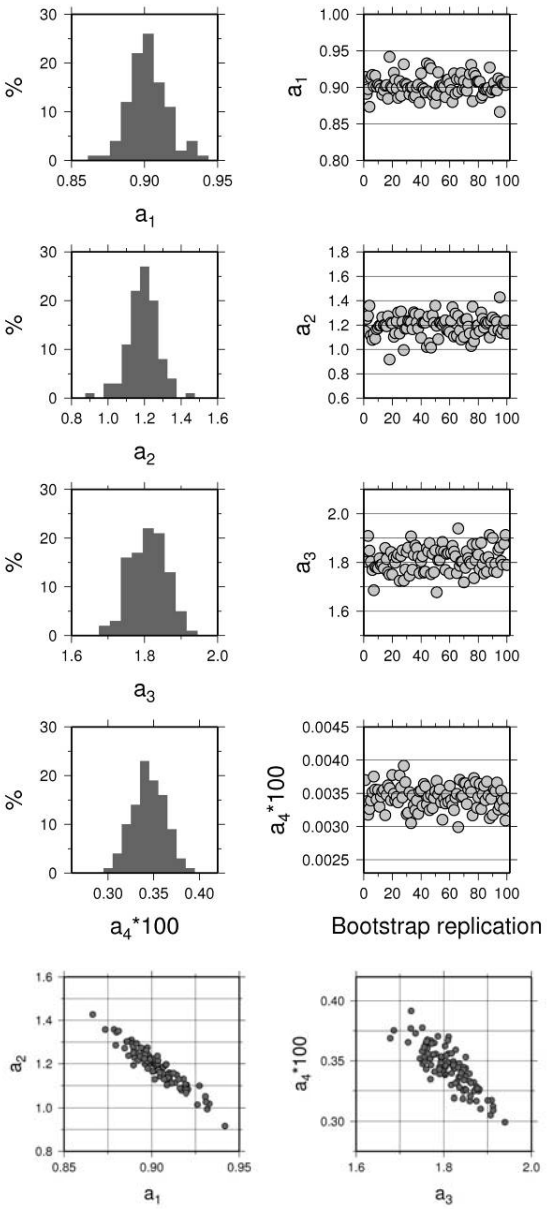


**Figure 7**. Results of 100 bootstrap regressions performed for the model depending on epicentral distance (see equation 5 in the text). The bottom panels show the internal trade-offs characterizing the source ($a_1$ and $a_2$, left) and attenuation ($a_3$ and $a_4$, right) coefficients.

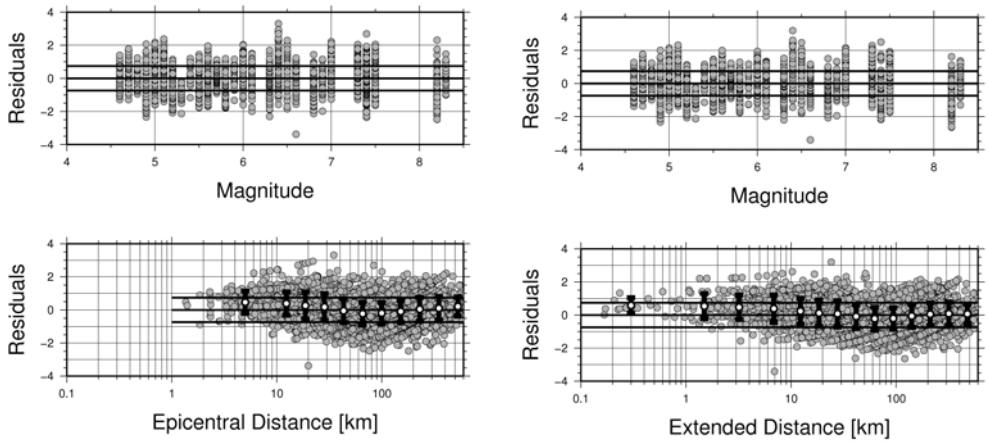


**Figure 8**. Left: Distribution of the residuals against magnitude (top) and distance (bottom) considering the model using epicentral distance. The vertical bars represent the average ± 1σ computed over different distance bins, with the first one being relevant for epicentral distances smaller than 10km. Right: Same as on the left but performing the regression considering the extended distance measure (see text for explanation). The first vertical bar corresponds to the average ± 1σ of residuals for extended distances < 1km.

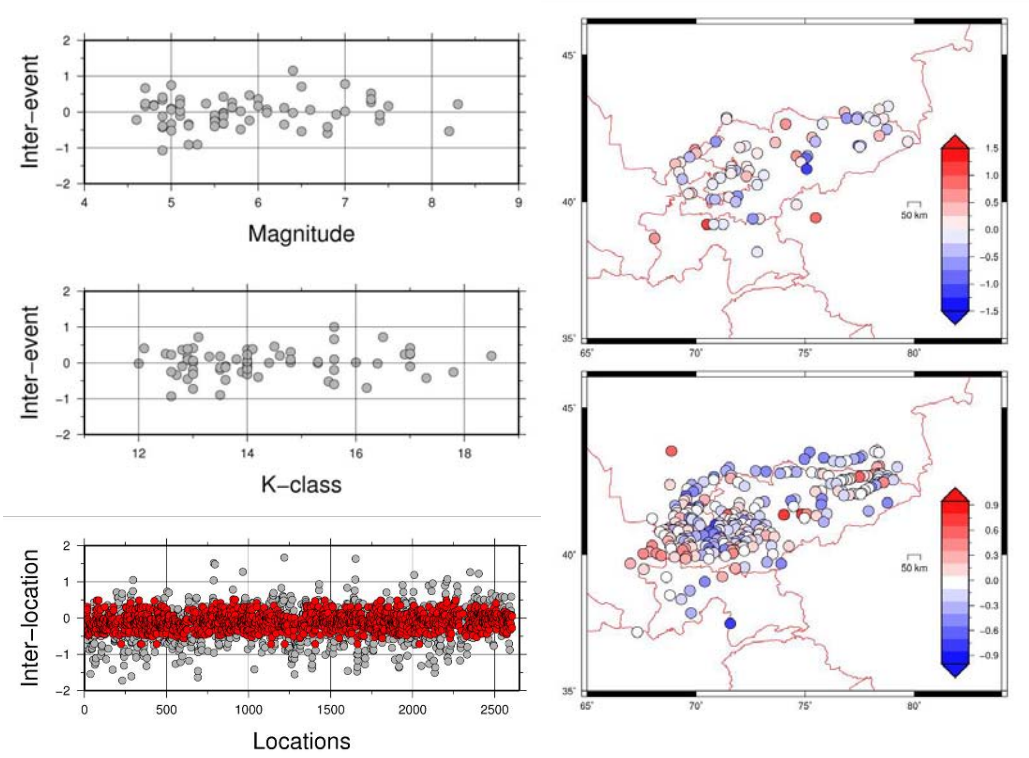


**Figure 9**. Left: Inter-event error against magnitude (top) and K-class (middle) for the model calibrated considering the epicentral distance; inter-location error (bottom) obtained for the considered intensity data points. The red circles indicate the results obtained for localities with more than 5 observations. Right: Spatial distribution of the inter-event (top) and inter-location (bottom) errors shown in the left panels.

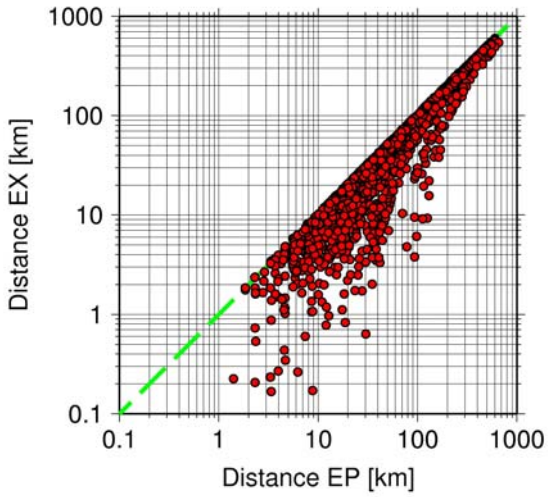


**Figure 10** Comparison between epicentral and extended distance measures.

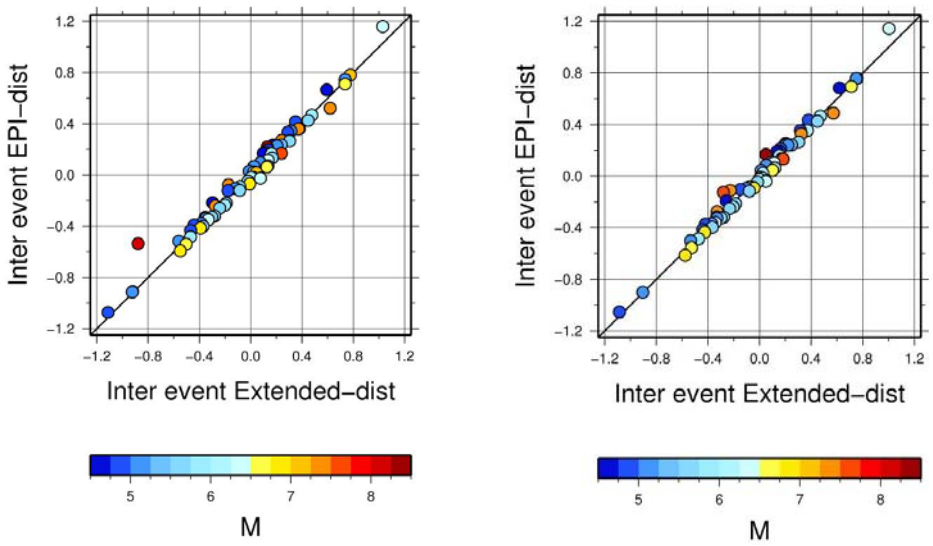


**Figure 11** Left: Comparison between inter-event errors obtained considering the epicentral and extended distance measures. Right: The same as on the left, however lowering the magnitude of the 1911 Kemin earthquake to 7.7. The circles are color-coded according to their magnitude value.

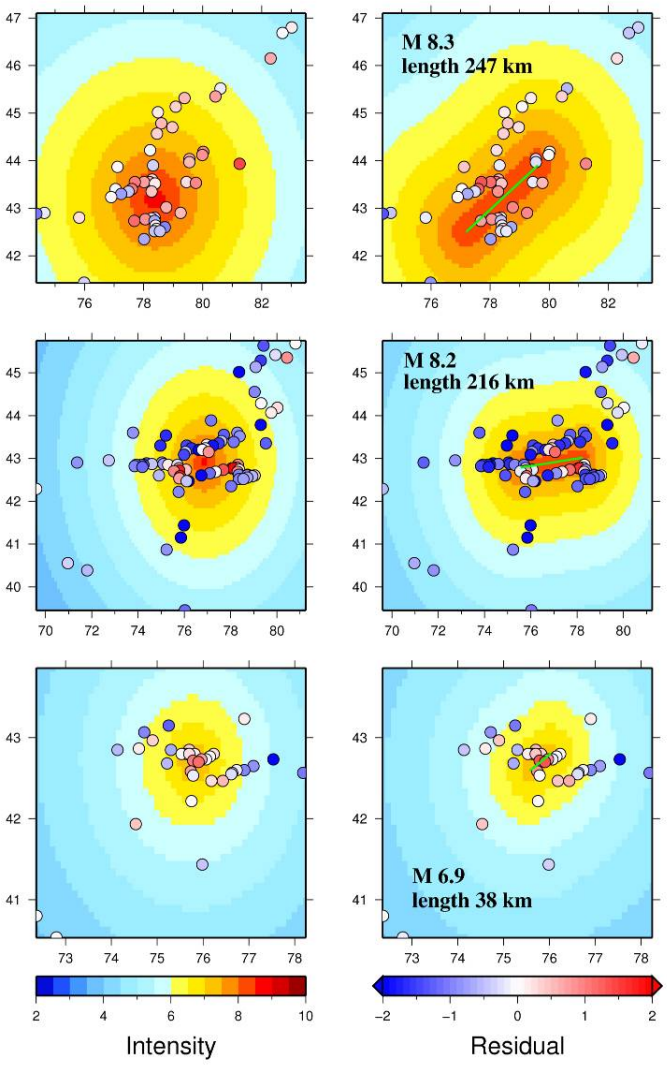


**Figure 12** Comparison of predictions with the derived IPE using the epicentral (left) and extended (right) distance measures. Top: 1889 M 8.3 Chilik earthquake; Middle: 1911 M 8.2 Kemin earthquake; Bottom: 1938 M 6.9 Kemin-Chu earthquake. The colored circles indicate the residual values. The segment used to compute the extended distances is shown in green in the right-hand plots.

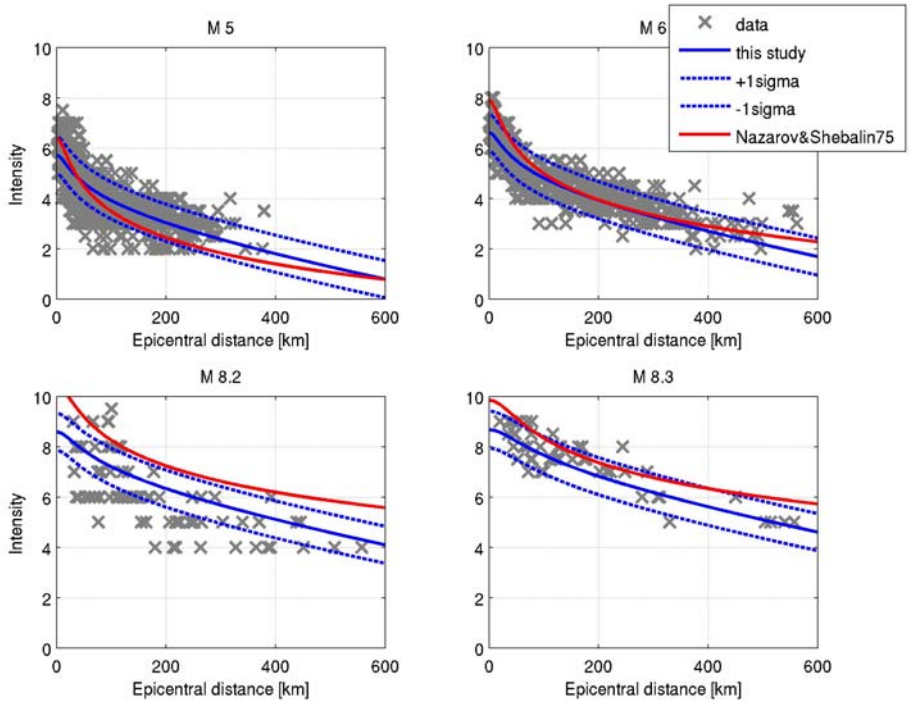


**Figure 13** Comparison between observations (crosses) and predictions (blue: this study; red: Nazarov and Shebalin, 1975). Top: Predictions for a magnitude M=5 (left) and M=6(right), considering a depth of 15 km. Observations in the range M±0.1 are selected. Bottom: Comparison for the 1911 M 8.2 (left) and 1889 M 8.3 (right) earthquakes.