Geophysical Journal International

Geophys. J. Int. (2023) **234,** 1771–1785
Advance Access publication 2023 April 18
GJI Seismology

https://doi.org/10.1093/gji/ggad164

Homogenization of magnitudes of the ISC Bulletin

Barbara Lolli [1], Daniele Randazzo,[1] Gianfranco Vannucci [1], Emanuele Biondini [2] and Paolo Gasperini [1,2]

[1] *Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Bologna, Viale Carlo Berti Pichat,* 6/2, 40127 *Bologna BO, Italy. E-mail: barbara.lolli@ingv.it*
[2] *Dipartimento di Fisica e Astronomia, Università di Bologna, Viale Carlo Berti Pichat,* 8, 40127 *Bologna BO, Italy*

Accepted 2023 April 12. Received 2023 March 13; in original form 2022 September 9

SUMMARY

We implemented an automatic procedure to download the hypocentral data of the online Bulletin of the International Seismological Centre (ISC) in order to produce in near real-time a homogeneous catalogue of the Global and EuroMediterranean instrumental seismicity to be used for forecasting experiments and other statistical analyses. For the interval covered by the reviewed ISC Bulletin, we adopt the ISC locations and convert the surface wave magnitude (M_s) and short-period body-wave magnitude (m_b) as computed by the ISC to moment magnitude (M_w), using empirical relations. We merge the so obtained proxies with real M_w provided by global and EuroMediterranean moment tensor catalogues. For the most recent time interval (about 2 yr) for which the reviewed ISC Bulletin is not available, we do the same but using the preferred (prime) location provided by the ISC Bulletin and converting to M_w the M_s and m_b provided by some authoritative agencies. For computing magnitude conversion equations, we use curvilinear relations defined in a previous work and the chi-square regression method that accounts for the uncertainties of both x and y variables.

Key words: Statistical methods; Earthquake source observations; Statistical seismology.

INTRODUCTION

The implementation in near real-time of earthquake operational forecasting methods (Jordan & Jones 2010; Jordan *et al.* 2011; Marzocchi *et al.* 2014) requires the availability of a high-quality seismic catalogue, which needs to be updated timely and reliably. In particular, the homogeneity of magnitude is the most important characteristic when studying properties like the *b*-value of the magnitude–frequency distribution (Gutenberg & Richter 1944) or other statistical parameters of seismic occurrence. Unfortunately, the magnitude is usually computed unevenly even for the same catalogue or bulletin, owing to the availability in different times and in different regions of different types of instrument and computational methods.

For the Italian region, Lolli *et al.* (2020) developed a homogeneous catalogue named HORUS, updated each hour and publicly available at http://horus.bo.ingv.it. HORUS is based on the conversion to M_w (Hanks & Kanamori 1979) of ML (Richter 1935) and M_d (Bisztricsany 1958) magnitudes from the bulletin of the *Istituto Nazionale di Geofisica e Vulcanologia* (INGV) and their merging with true M_w estimates from online moment tensor catalogues. In this work we present a similar homogeneous catalogue both at the global (GBL) and at the EuroMediterranean scales (MED, from latitude 25°N to 60°N and from longitude 20°W to 50°E) based on the hypocentral and magnitude data downloaded from the Bulletin of the International Seismological Centre (ISC 2022; see Data and Resource section).

The ISC hosts the most complete seismic Bulletin at the global scale. It relies on data contributed by seismological agencies from around the world and provides mainly two kinds of Bulletins: reviewed and unreviewed. The reviewed ISC Bulletin, which is manually checked by ISC analysts and relocated (for more details on ISC operations see Appendix in International Seismological Centre 2021), is typically 24 months behind real-time, and in this paper we downloaded data in December 2022 when the ISC reviewed Bulletin was available up to December 2020. The unreviewed ISC Bulletin is typically updated in near real-time and reports hypocentres and magnitudes as provided by contributing agencies. Data are automatically grouped into events and the content remains unreviewed until the time comes when ISC analysts review a batch of one month and relocate a large fraction of reviewed events. For ISC relocated events the ISC may also recompute M_s and m_b as described in Bondár & Storchak (2011).

ISC also hosts the ISC-GEM catalogue (International Seismological Centre 2018), currently covering the period from 1904 to 2018, which reports a unique moment magnitude M_w by merging of real M_w estimates and M_w proxies computed from M_s and m_b according to Storchak *et al.* (2012) and Di Giacomo *et al.* (2015). However, as noted by Lolli *et al.* (2014), the empirical conversion relations computed according to Storchak *et al.* (2012) and Di Giacomo *et al.*

(2015) are biased at low magnitudes by the incompleteness of M_w data and the use of the Ordinary Least Square (OLS) regression method, which neglects the error of the x variable. Di Giacomo *et al.* (2015) warned on the use of their relationships for earthquakes below $\sim M$ 5, in this work we expand the data set to lower magnitudes.

Lolli *et al.* (2014) computed empirical curvilinear relations between reviewed ISC M_s and m_b and moment magnitudes M_w from online databases using the chi-square (CSQ) regression method (Stromeyer *et al.* 2004), which considers the uncertainties of both the y and x variables. Lolli & Gasperini (2012) demonstrated that such method is equivalent to the so-called General Orthogonal Regression (GOR) method (Fuller 1987; Castellaro *et al.* 2006) when the regression model is linear and the ratio between the variances of the y and x variables is constant. The advantages of the CSQ with respect to the GOR are that the former allows to assign different uncertainties to each observation and that can handle with non-linear regression equations (see in the Appendix an overview of the CSQ regression method).

The ISC Bulletin lists locations and magnitudes from contributing agencies as well as ISC recomputed M_s and m_b (if criteria are met) in its reviewed period. Such, magnitudes in the ISC Bulletin are a mixture of local/regional and teleseismic types, often with undocumented procedures. Instead, standards for M_s and m_b (e.g. IASPEI 2013) computation are more widely applied in the seismological community and that makes those more suitable for conversions to M_w, particularly at global scale. In this paper we apply the same kind of regression analysis, made for ISC by Lolli *et al.* (2014), to the data of some of the largest data contributing agencies and derive conversion equations to compute M_w proxies from the M_s and m_b of the unreviewed portion of the ISC Bulletin.

We also redo the regression analysis for ISC recomputed magnitudes because about 10 yr passed since the previous calibration by Lolli *et al.* (2014), which concerned data only up to 2010. Since then, ISC not only added the successive years of data up to 2020, but it even reviewed the data of many earthquakes that occurred before (Storchak *et al.* 2017, 2020).

As reference for M_s and m_b calibrations we use a data set of M_w obtained by combining the estimates provided by various online moment tensor catalogues according to Gasperini *et al.* (2012) and Lolli *et al.* (2020).

REFERENCE DATA SET OF M_w FROM MOMENT TENSOR CATALOGUES

Gasperini *et al.* (2012), using error-in-variables regression methods (Fuller 1987; Stromeyer *et al.* 2004; Castellaro *et al.* 2006), compared five different data sets of direct M_w determined up to 2010 by moment tensors inversion: the Global Centroid Moment Tensor (GCMT) catalogue (Dziewonki *et al.* 1981; Ekström et al 2012), the moment tensor catalogue of the National Earthquake Information Center (NEIC; Sipkin 1994), the Regional Centroid Moment tensor (RCMT) catalogue of the Istituto Nazionale di Geofisica e Vulcanolgia (INGV; Pondrelli *et al.* 2002, 2011), the moment tensor catalogue of the Eidgenössische Technische Hochschule Zürich (ETHZ; Bernardi *et al.* 2004) and the Time Domain Moment Tensor (TDMT) catalogue of INGV (Scognamiglio *et al.* 2009).

Lolli *et al.* (2020) recalibrated such data sets using the data up to 2018 and also calibrated the moment tensor catalogue of the Geo Forschungs Zentrum Potsdam (GFZP; Saul *et al.* 2011) that started in 2011 (see Data and Resource section for details on all data sets). Both works found that in general various M_w data sets scale 1:1 with each other but in some cases differ by average offsets. Following Gasperini *et al.* (2012) and Lolli *et al.* (2020), we build our reference M_w data set merging different sources. In particular, if any other direct source is available, we discard NEIC for earthquakes $M_w > 7$ because they might be underestimated and GCMT for $M_w < 5.4$ because they might be overestimated (see the discussion Lolli *et al.* 2015). In doing so we also apply shifts to the corresponding M_w sources with the factors reported in Table 1. Gasperini *et al.* (2012) and Lolli *et al.* (2020) also estimated the uncertainties of M_w by considering the mean squared deviations σ_d between corrected M_w (after the application of offsets with respect to GCMT) from different MT data sets. They inferred that such M_w uncertainties were 0.10 m.u. up to 1995 for GCMT and NEIC, 0.07 m.u. after 1995 for GCMT, NEIC, ETHZ and RCMT, and for TDMT, 0.13 m.u. up to 2010 and 0.07 after such date (see Table 1).

When more than one M_w estimate is available from different catalogues, Gasperini *et al.* (2012) and Lolli *et al.* (2020) suggested compute the average of available corrected estimates, weighted by the inverse of the respective variance. In such cases, they also suggested to assign to the computed average M_w the minimum uncertainty among those of the averaged data sets.

In this paper, we applied all such prescriptions and suggestions to compile a catalogue (MWREF) of M_w from moment tensors up to 2022 which we use as reference for calibrations of M_s and m_b.

CONVERSION TO M_w OF TELESEISMIC MAGNITUDES M_s AND m_b

It is well known that the conversion relations of M_s and m_b with M_w are not linear. Kanamori & Anderson (1975), based on theoretical considerations on the excitation of surface waves with period of about 20 s (used to compute M_s), inferred that the scaling coefficient between M_s and M_w should be 2/3 at low magnitudes, 1 at intermediate magnitudes and 2 at high magnitudes. The transition between the slopes at intermediate and low magnitudes is estimated to be around $M_s = 6.2$ (Ekström & Dziewonski 1988; Scordilis 2006). The transition at high magnitudes is not precisely known but it probably occurs for $M_s > 8$ where saturation is sometimes observed.

A similar situation should occur for m_b but with transition points shifted down to lower magnitudes, owing to the shorter period of seismic waves sampled (between 1 and 5 s). Gasperini *et al.* (2013b) showed that ISC m_b coincides with M_w at low values ($m_b < 4$–4.5) while it saturates for $m_b > 5$–5.5. Hence, the transitions between slopes 1 and 2 can be assumed around $m_b = 5$–5.5. The transition point between slope 1 and 2/3 is unknown but, according to Gasperini *et al.* (2013b) and Lolli *et al.* (2014), it is not observed above $m_b = 3.5$, which is about the minimum m_b computed by ISC and other agencies.

Ekström & Dziewonski (1988) proposed a curvilinear relation between M_w and M_s in which two straight segments with slope 1 and 2/3 at high and low M_s, respectively are connected by a curved line. Following the same schema, Lolli *et al.* (2014) proposed a circularly connected bi-linear (CBL) model with form

$$M_w = \begin{cases} a M_s + b & M_s < M_{slow} \\ M_{wc} - \sqrt{R^2 - (M_s - M_{sc})^2} & M_{slow} \le M_s < M_{sup}, \\ M_s & M_s > M_{sup} \end{cases} \quad (1)$$

where a and b are empirical parameters and M_{wc}, M_{sc} and R are, respectively, the centre coordinates and the radius of an arc of circle

Table 1. Calibration of various M_w data sets and related uncertainties.

Source	Time coverage	Regions	Shift	σ, yr δ 2010	σ, yr > 2010	Discard
GCMT	1976–present[1]	GBL, MED	0.00	0.07–0.10[2]	0.07	$M_w < 5.4$[3]
NEIC	1980–2010	GBL, MED	0.05	0.07–0.10[2]	-	$M_w > 7.0$[4]
RCMT	1997–present[5]	MED	0.00	0.07	0.07	
ETHZ	1999–2006	MED	− 0.05	0.07	-	
TDMT	2006–present	MED	0.20	0.13	0.07	
GFZ	2011–present	GBL, MED	0.05	-	0.07	

[1]Solutions for some areas available since 1962.
[2]0.10 before 1995.
[3]Data from GCMT are discarded if any other source gives $M_w < 5.4$.
[4]Data from NEIC are discarded if any other source gives $M_w > 7.0$.
[5]Solutions for some areas available since 1964.

in the M_w–M_s plane connecting the two straight lines (see fig. 1 of Lolli *et al.* 2014) and also preserving the continuity of first derivative at the connection points. Lolli *et al.* (2014) showed that M_{slow}, M_{sup}, M_{wc}, M_{sc} and R can all be computed as a function of a, b and of the distance Δ of the two connection points from the intersection of the two straight lines at $M_I = b/(1-a)$ so that the free parameters to regress are only three: a, b and Δ. The main difference with respect to the relation proposed by Ekström & Dziewonski (1988) is that the slope at low magnitudes is left free to vary instead of being fixed to 2/3. Lolli *et al.* (2014) found slopes varying from 0.56 to 0.70, depending on the data set, which are substantially consistent with the theoretical value of 2/3.

Storchak *et al.* (2012) and Di Giacomo *et al.* (2015) proposed instead an exponential regression model (EXP) with the form

$$M_w = \exp(a + b\,m) + c, \tag{2}$$

where m is either M_s or m_b and a, b and c are empirical coefficients. Such expression lacks a physical explanation but is empirically demonstrated to reproduce satisfactorily the relations of M_w as a function of M_s (Lolli *et al.* 2014).

Eq. (2) can be applied even to the relation between M_w and m_b whereas the CBL formulation (1) cannot because, according to Kanamori & Anderson (1975), M_w and m_b are approximately coincident a low m_b values (lower than 5) while at high m_b the slope tends to 2. Lolli *et al.* (2014) in this case proposed a Circularly connected Bi-Linear Reversed (CBL-R) model described by

$$M_w = \begin{cases} m_b & m_b < m_{blow} \\ M_{wc} - \sqrt{R^2 - (m_b - m_{bc})^2} & m_{blow} \le m_b \le m_{bhigh} \\ a m_b + b & m_b > m_{bhigh} \end{cases}, \tag{3}$$

where again a and b are empirical parameters and M_{wc}, m_{bc} and R are the centre coordinates and the radius of an arc of circle in the M_w–m_b plane connecting the two straight lines and preserving the continuity of the first derivative at the connection points. In the CBL-R model M_w and m_b coincide one to the other at low magnitudes and scale with coefficient a at high magnitudes. Even in this case Lolli *et al.* (2014) showed that m_{blow}, m_{bup}, M_{wc}, m_{bc} and R can all be computed as a function of a, b and of the distance Δ between the two connection points and the intersection of the two straight lines at $M_I = b/(1-a)$ so that the free parameters to regress are again only three. Lolli *et al.* (2014) also observed that for CBL-R the parameter Δ cannot be left free to vary because it tends to any upper limit imposed during the numerical minimization and hence, they fixed it to 2 in all cases.

In its supplemental material, Lolli *et al.* (2014) provided the Fortran source codes to compute M_w and its uncertainty as a function of M_s and m_b and their uncertainties, using the CBL, EXP and CBL-R functional models.

We first analyse the M_s and m_b data from the ISC reviewed Bulletin from 1964 to 2020 (we downloaded data in December 2022) and then proceed analogously with the data of agencies as reported by the unreviewed ISC Bulletin. From a preliminary screening, we individuated four agencies providing enough data for reliable calibrations. They are: The National Earthquake Information Centre (NEIC) of the U.S. Geological Survey, the International Data Centre (IDC) of the Comprehensive Nuclear-Test-Ban Treaty Organization in Vienna, the China Earthquake Networks Centre in Bejing (BJI) and the Geophysical Survey of Russian Academy of Sciences in Moskow (MOS). As we will have to use such data to compute M_w proxies in the most recent 2-yr interval not covered by the reviewed ISC Bulletin, we analysed the M_s and m_b data up to 2022 but only since 2005 so that their properties be more like to the most recent data.

We were informed (Di Giacomo, personal communication) that NEIC changed the procedure to compute M_s from 2013. Since such date, NEIC M_s estimates provided by the ISC Bulletin are labelled as M_s_20 (instead of M_s) to indicate that they are computed using surface waves with period of about 20 s according the IASPEI (2013) standard. Hence, we calibrated with M_w only such estimates as they are the only ones provided in the most recent 2 yr when the ISC reviewed Bulletin is not available. As well, we also observed that the earthquakes for which BJI and MOS are the only magnitude providers available are very few (less than 1 per cent). Then they are not so useful for computing proxies. Hence, in the following we will concentrate mainly on the analysis of NEIC and IDC M_s and m_b and will provide the plots for other data sets in the online supplemental material only.

Since recently, even the German Research Centre for Geosciences (GFZ) in Potsdam started to regularly provide locations and m_b magnitudes to ISC at the GBL scale but the data available until now are not sufficient for a reliable calibration with respect to M_w. We then postpone such calibration until there will be enough data available.

To produce the data sets to be regressed, we compared each location computed by the ISC or by other agencies with those reported in the MWREF catalogue and associated them when the time difference was less than 10 s and the spatial distance was less than 20 km. This event matching scheme may lead to mis-associated magnitude entries in our data set, however it is unlikely to bias our results given its size.

According to Lolli *et al.* (2014), for the construction of data sets to regress, we consider only M_s and m_b magnitudes from ISC and agencies with at least 3 observations from different stations because

Table 2. Estimated standard deviations of M_s and m_b proxies due to lateral heterogeneities of seismic wave propagation (see text).

Region	Data set	σ_{M_sG}	σ_{m_bG}
GBL	ISC 1964–2020	0.14	0.23
	ISC 1964–2010	0.15	0.23
	ISC 2011–2020	0.13	0.23
	NEIC	0.13	0.20
	IDC	0.09	0.24
	BJI	0.21	0.21
	MOS	0.17	0.20
	L & al. (2014)	0.14	0.23
MED	ISC 1964–2020	0.18	0.18
	ISC 1964–2010	0.20	0.18
	ISC 2011–2020	0.13	0.18
	NEIC	0.08	0.18
	IDC	0.10	0.16
	BJI	0.23	0.18
	MOS	0.14	0.16
	L & al. (2014)	0.16	0.17

they are more reliable. For M_s we only consider earthquakes with source depth $<$50 km.

The determination of magnitudes uncertainties to be used in error-in-variable regression methods like the CSQ is a challenging problem because usually errors are not provided by the data sources. We already discussed the question for M_w uncertainties in the reference M_w data set section. For M_s and m_b, the ISC provides formal uncertainties but these only represent the dispersion of the magnitudes computed by the different stations rather than the true uncertainties of the average magnitudes. As M_s and m_b are based on global calibration functions, Lolli *et al.* (2014) proposed to compute the uncertainty of M_s and m_b as the combination of two terms: the first one is the standard error of the mean of the magnitudes observed by different stations and the second one represents the dispersion of seismic wave attenuation heterogeneities all over the Earth. In particular, Lolli *et al.* (2014) proposed to compute the error of each M_s and m_b observation as

$$\sigma_{M_s} = \sqrt{\frac{\bar{\sigma}^2_{M_s}}{n} + \sigma^2_{M_sG}}, \tag{4}$$

$$\sigma_{m_b} = \sqrt{\frac{\bar{\sigma}^2_{m_b}}{n} + \sigma^2_{m_bG}}, \tag{5}$$

where σ_{M_s} and σ_{m_b} are the average standard deviations of M_s and m_b observations, respectively, n is the number of stations providing magnitude estimates for the given earthquake and σ_{M_sG} and σ_{m_bG} are the standard deviations of the differences between observed M_w and M_w proxy computed from M_s and m_b, respectively on a tessellation of cells covering the entire Earth surface (see fig. 2 and table 1 of Lolli *et al.* 2014). Based on an analysis of sampled data from 1970 to 2009, Lolli *et al.* (2014) determined $\sigma_{M_s} = 0.33$ and $\sigma_{m_b} = 0.41$. They also assumed as initial values of σ_{M_sG} and σ_{m_bG}, 0.08 and 0.20, respectively, but adjusted them so that the *a priori* and the empirical variances of the regressions become equal (see Lolli *et al.* 2014 for details). In this work we adopted the same approach by assuming $\sigma_{M_s} = 0.33$ and $\sigma_{m_b} = 0.41$, determined by Lolli *et al.* (2014), and inferring σ_{M_sG} and σ_{m_bG} from regressions so that to make the *a priori* (from errors) and empirical (from regressions) variances being approximately equal. In Table 2, we report the obtained values of adjusted errors σ_{M_sG} and σ_{m_bG} for the various data sets.

Such treatment of magnitude errors has been criticized by a reviewer of this paper, who hypothesized that even the error in the equation must be considered by adding some further variance to the y variable. Such opinion is based on the paper by Carrol & Ruppert (1996) that, however, only concerns linear regressions and, in particular, provides a method to evaluate the presence of regression error (the method of moments) which only holds for linear equations. However, as some of (non-linear) equations we use (CBL and CBL-R) are based on physical arguments and the other one (EXP) is found empirically consistent with the previous ones, we believe that there is no evidence of the presence of regression error and then we do not add any additional variance to y variable.

Compared to those obtained by Lolli *et al.* (2014) the values estimated in this work for the reviewed ISC data set are about the same at the GBL scale and slightly larger at the MED scale. For the other agencies we obtained that, at the GBL scale, σ_{M_sG} are smaller than ISC for NEIC and IDC and larger than ISC for BJI and MOS, and at the MED scale they are smaller than ISC for NEIC IDC and MOS and larger than ISC for BJI. These differences are possibly related to the different procedures adopted by different agencies to compute the M_s and by the different attenuation properties in the areas mostly covered by the agencies. For σ_{m_bG}, we obtained values not very different from ISC for the other agencies.

ISC REVIEWED DATA SET

In Fig. 1 and in Table 3 (for EXP models) and Table 4 (for CBL and CBL-R models) we show plots and coefficients of curvilinear regressions of M_w as a function of M_s (top panel) and m_b (bottom panel) for the ISC reviewed Bulletin from 1964 to 2020 at the GBL scale. The red curves indicate the EXP regression model while the black ones the CBL (for M_s) and CBL-R (for m_b) models. In both panels the green dashed curve indicates the EXP model regressed by Lolli *et al.* (2014) using the data up to 2010 and the pink dashed curve the EXP model computed by Di Giacomo *et al.* (2015), using the OLS method on the ISC-GEM catalogue from 1976 to 2009. For M_s [Fig. 1(top panel)] we can note a reasonably good agreement between all the curves indicating that the M_s–M_w relation for ISC does not change much when the new data after 2010 are added. In the supplemental material one can see a more detailed comparison between the data sets up to and after 2010 [Figs S1(top panel) and S2(top panel), respectively], showing an almost perfect coincidence of the regression curves.

For m_b [Fig 1(bottom panel)] we can note a slight discrepancy between the EXP (red) and CBL-R (black) curves at low and high m_b as well as between the EXP curves computed in this work (red) and by Lolli *et al.* (2014) (dashed green). This discrepancy is even more evident for the regression after 2010 [Fig. S2(bottom panel)] where the dispersion of data is lower than for the data set up to 2010 [Fig. S1(bottom panel)]. In particular, there are many more M_w data below 4.5, and the saturation above m_b 5.0 is less pronounced. We can guess that such different distribution, and in particular the increased number of M_w data below 4.5, is due to the contribution of GFZP that started to provide moment tensor solutions just from 2011. In Fig. 1(bottom panel) it is also reported (pink dashed curve) the EXP model computed by Di Giacomo *et al.* (2015), using the OLS method on the ISC-GEM catalogue from 1976 to 2009 (which is somehow similar to the ISC data set from 1964 to 2010 plotted in Fig. S1). We can note that the pink dashed curve at low m_b follows the tail of data distribution around $M_w \approx 4.5$–4.7 whereas our EXP model (red curve) regressed with the CSQ method follows the tail with $M_w \approx m_b$. Such discrepancy was explained by Gasperini *et al.* (2013b) and Lolli *et al.* (2014) as due to the incompleteness of the

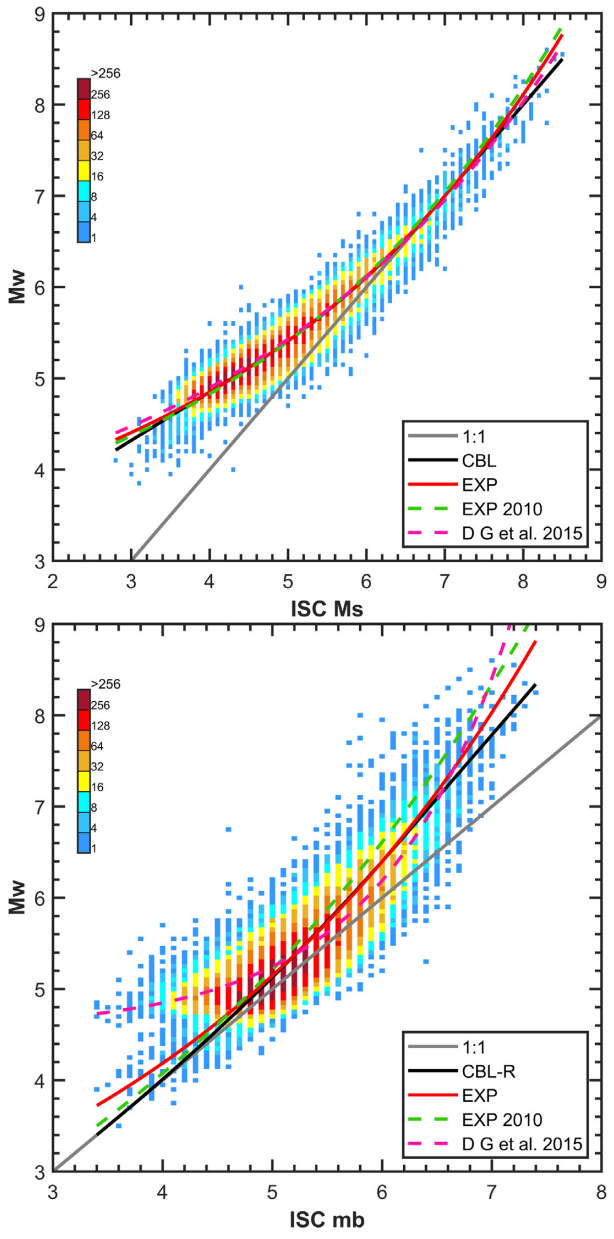


Figure 1. Data frequencies and regression curves for the GBL ISC reviewed data set from 1964 to 2020. Green and pink dashed curves concern EXP regressions by Lolli *et al.* (2014) and by Di Giacomo *et al.* (2015), respectively. Total numbers of data pairs are 34 188 for M_s and 48 498 for m_b.

GBL M_w data set below $M_w = 5.0$ and to the fact that the CSQ regression almost corresponds to the inverse least square regression of m_b as a function of M_w, because the error of M_w (about 0.1 units) is definitely lower than that of m_b (ranging from 0.2 to 0.4 units).

In Fig. 2, we report the plots of the same regressions for a data set limited to the MED region, showing for M_s a marked difference with respect to the GBL one below M_s 5.0. Note in Table 4 that the CBL slope a (at low M_s) for GBL is 0.531 whereas for MED is 0.682 (the latter almost coincident with the theoretical value of 2/3). As already noted by Lolli *et al.* (2014), such difference probably reflects the incompleteness of the M_w GBL data set below 5.0 and/or the overestimation of GCMT magnitudes below 5.4. Also note in Table 4 the values of M_I corresponding to the intersection points of the straight lines in CBL and CBL-R relations of M_w with M_s and m_b (at low and high magnitudes, respectively). For M_s, it is 5.81 (lower than expected) at GBL scale whereas it is 6.17 (about that expected) at the MED scale. We can argue that even this underestimation at the GBL scale is a consequence of the already mentioned incompleteness of the GBL M_w data set below 5.0. For such reasons we confirm here the choice made by Lolli *et al.* (2014) that, for converting to M_w the $M_s < 5.5$ of ISC, the MED regression coefficients are used in place of the GBL ones, even for the reviewed GBL data set.

For m_b, the MED curves (red and black in Fig. 2 bottom) are reasonably similar to the GBL EXP one (dashed pink) with an underestimation of about 0.2 m.u. of MED with respect to GBL at low magnitudes. Figs S3 and S4(bottom panel) show the comparison between the MED data sets up to 2010 and after such date, respectively, indicating a reasonable correspondence between the two and with the GBL data set. This confirms that the correct regression at the GBL scale is that computed by us using the CSQ methods. For m_b the values of M_I are more variable ranging from 4.98 from GBL to 5.50 for MED. In this case we cannot establish which value is more realistic.

NEIC DATA SET

In Fig. 3 and in Table 5 (for EXP models) and Table 6 (for CBL and CBL-R models) we report the plots and the coefficients of the regressions at the GBL scale for NEIC. For M_s [Fig. 3(top panel)] there is a nice correspondence between the EXP regressions (red curve) with the CBL one (black) and even with the GBL ISC (dashed green). For m_b [Fig. 3(bottom panel)] the regressed curves (red and black) are slightly different to those determined for EXP ISC (dashed green). In general, m_b estimated by NEIC appear to be slightly less saturated at high magnitudes than those estimated by ISC and more consistent with the theory ($m_b = M_w$) at low magnitudes even for the EXP model.

In Fig. 4, we show the MED regressions for NEIC where the pink dashed curves reproduce the GBL EXP relations of Fig. 3. We note that for M_s [Fig. 4(top panel)], the GBL curves slightly underestimate the MED ones whereas for m_b [Fig. 4(bottom panel)], the MED curves slightly overestimate GBL ones.

The intersection magnitude M_I for M_s is 6.41 and 6.59 at the GBL and MED scales, respectively). For m_b, M_I is 5.47 and 5.83 at the GBL and MED scales, respectively. This also confirms a lower saturation at high magnitudes for MED with respect to GBL.

IDC DATA SET

In Fig. 5 and Tables 5 and 6, we show the plots and the coefficients of IDC at the GBL scale. For M_s we can see a small M_w overestimation (of about 0.1–0.2 units) of IDC (red and black) with respect to ISC (dashed green) for $3.5 < M_s < 7$. In other words, the deviation with respect to the line of equality between M_w and M_s starts below $M_s \approx 7$ instead of below $M_s \approx 6$–6.5 as in the ISC data set. This discrepancy is confirmed by intersection magnitudes M_I of CBL models, which are well above the values obtained for ISC (6.53 for GBL and 7.14 for MED).

For m_b, the regressed curves clearly overestimate the ISC ones of about 0.5–0.7 m.u. The reason of such large discrepancy can be explained based on the mission of IDC, which is to discriminate anthropogenic events from natural earthquakes. In particular, their filter (usually up to 1.5 s) applied to the first 5 s of *P*-wave signals is much narrower than what ISC, NEIC and many other seismological

Table 3. Regression coefficients of EXP model for ISC data sets.

Mag.	Reg.	Data set	N	a	b	c	σ_r
M_s	GBL	1964–2020	34 188	-0.137 ± 0.032	0.229 ± 0.003	2.673 ± 0.041	0.138
		1964–2010	18 751	-0.018 ± 0.049	0.219 ± 0.005	2.469 ± 0.072	0.149
		2011–2020	15 437	0.124 ± 0.050	0.202 ± 0.005	2.323 ± 0.077	0.130
		L&al 14∗	*19 475*	*−0.109 ± 0.047*	*0.229 ± 0.005*	*2.586 ± 0.065*	*0.146*
		D&a1 15$^+$		*−0.222 ± 0.043*	*0.233 ± 0.004*	*2.863 ± 0.056*	
	MED	1964–2020	2842	3.052 ± 0.399	0.029 ± 0.010	-19.05 ± 8.496	0.171
		1964–2010	1419	2.091 ± 0.340	0.064 ± 0.017	-5.799 ± 2.865	0.182
		2011–2020	1423	4.219 ± 0.691	0.010 ± 0.007	-66.02 ± 47.04	0.149
		L&al 2014	*1200*	*2.133 ± 0.343*	*0.063 ± 0.017*	*−6.205 ± 3.013*	*0.170*
m_b	GBL	1964–2020	48 498	0.082 ± 0.098	0.266 ± 0.011	1.039 ± 0.184	0.293
		1964–2010	27 898	0.238 ± 0.173	0.254 ± 0.018	0.619 ± 0.371	0.312
		2011–2020	20 600	1.512 ± 0.187	0.128 ± 0.014	-3.434 ± 0.997	0.278
		L&al 14∗	*30 326*	*0.741 ± 0.170*	*0.210 ± 0.017*	*−0.785 ± 0.520*	*0.333*
		D&a1 15$^+$		*−4.664 ± 0.085*	*0.859 ± 0.012*	*4.555 ± 0.017*	
	MED	1964–2020	3871	1.111 ± 0.198	0.165 ± 0.018	-1.852 ± 0.742	0.235
		1964–2010	1939	0.424 ± 0.262	0.236 ± 0.029	0.122 ± 0.591	0.241
		2011–2020	1932	2.300 ± 0.443	0.075 ± 0.024	-9.422 ± 4.655	0.228
		L&al 2014	*1673*	*0.719 ± 0.329*	*0.212 ± 0.034*	*−0.737 ± 0.939*	*0.240*

∗Coefficients computed by Lolli *et al.* (2014). $^+$Coefficients computed by Di Giacomo *et al.* (2015).

Table 4. Regression coefficients of CBL (M_s) and CBL-R (m_b) models for ISC data sets.

Mag.	Reg.	Data set	N	a	b	Δ	σ_r	M_I
M_s	GBL	1964–2020	34 188	0.531 ± 0.003	2.726 ± 0.011	1.641 ± 0.015	0.138	5.81
		1964–2010	18 751	0.565 ± 0.004	2.554 ± 0.020	1.648 ± 0.022	0.149	5.88
		2011–2020	15 437	0.545 ± 0.003	2.680 ± 0.013	1.545 ± 0.024	0.130	5.89
		L&al 2014	*19 475*	*0.560 ± 0.005*	*2.582 ± 0.022*	*1.854 ± 0.025*	*0.147*	*5.87*
	MED	1964–2020	2842	0.682 ± 0.005	1.960 ± 0.021	0.963 ± 0.131	0.172	6.17
		1964–2010	1419	0.684 ± 0.008	1.923 ± 0.036	0.996 ± 0.158	0.185	6.09
		2011–2020	1423	0.706 ± 0.001	1.895 ± 0.002	0.210 ± 0.003	0.149	6.45
		L&al 2014	*1200*	*0.701 ± 0.008*	*1.853 ± 0.036*	*1.151 ± 0.170*	*0.171*	*6.20*
m_b	GBL	1964–2020	48 498	1.390 ± 0.011	-1.942 ± 0.063	2.000	0.305	4.98
		1964–2010	27 898	1.639 ± 0.023	-3.419 ± 0136	2.000	0.323	5.35
		2011–2020	20 600	1.211 ± 0.017	-0.926 ± 0.094	2.000	0.289	4.40
		L&al 2014	*30 326*	*1.609 ± 0.018*	*−3.031 ± 0.104*	*2.000*	*0.340*	*4.98*
	MED	1964–2020	3871	1.650 ± 0.068	-3.578 ± 0.404	2.000	0.241	5.50
		1964–2010	1939	1.010 ± 64.21	-10.00 ± 351.4	2.000	0.221	-
		2011–2020	1932	1.170 ± 0.002	-0.833 ± 0.005	2.000	0.231	4.90
		L&al 2014	*1673*	*1.509 ± 0.055*	*−2.475 ± 0.306*	*2.000*	*0.245*	*4.86*

agencies do (up to 3 s). Moreover, the amplitude A is measured on a much longer time window. Hence the A/T by IDC is usually smaller than A/T from ISC, NEIC and others starting from moderate earthquakes (Di Giacomo, personal communication)

The plots displayed in Fig. 6 top for the MED data set show an underestimation with respect to GBL for $M_s < 5$ and a substantial coincidence with MED ISC that seem to be coherent with the mentioned incompleteness of the GBL M_w data set at low M_w. We can infer that even for IDC M_s, the coefficients computed from the MED data set are preferable to the GBL ones for computing M_w proxies for $M_s < 5.5$ even at the GBL scale. For m_b we have a slight underestimation with respect to the GBL but anyway well above the MED ISC curve.

BJI DATA SET

In Fig. S5 and in Table 5 and Table 6, we show the plots and the coefficients for BJI at the GBL scale. In the top panel, we can note that the coincidence between M_s and M_w goes well below M_s 5.5. The two curves start to deviate from the line of coincidence (grey) only below M_s 5.0. This also indicate that the procedure followed to compute M_s in BJI is different to that applied by ISC. Such procedures are described in a report by Dai & An (2017) where one can read that two kind of M_s are computed named M_s and M_s 7, respectively. In particular, the M_s (which is the only one we considered) is computed according to a formula slightly different from the IASPEI (2013) standard (the constant term is 3.5 instead of 3.3). The MED regression of Fig. S6 (top panel) confirms this finding and suggests an almost exact correspondence between M_w and BJI M_s.

For GBL m_b of Fig. S5 (bottom panel) we see a slightly steeper slope and a wider dispersion at high m_b values with respect to ISC. For the MED data set [Fig. S6 (bottom panel)] the relation between M_w and m_b appears instead almost linear but not very different to the GBL one. In the latter case, however, the relatively small number of data and the high dispersion make such regression not very reliable.

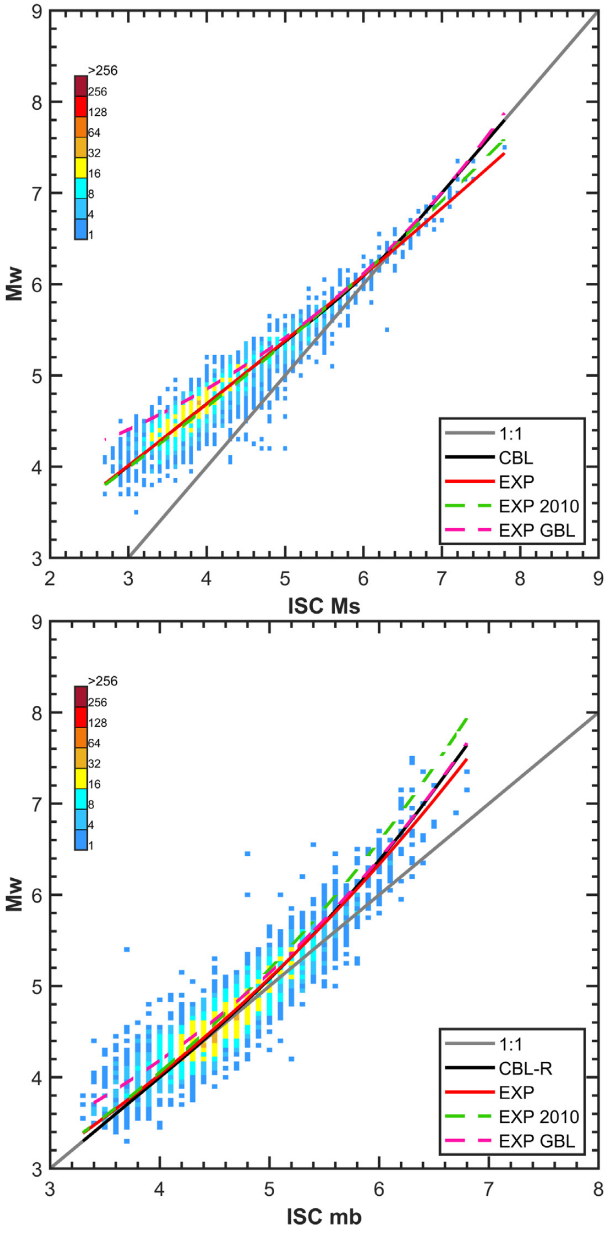


Figure 2. Data frequencies and regression curves for the MED ISC reviewed data set from 1964 to 2020. Green and pink dashed curves concern EXP regressions by Lolli *et al.* (2014) and for the GBL data set (Fig. 1), respectively. Total numbers of data pairs are 2842 for M_s and 3871 for m_b.

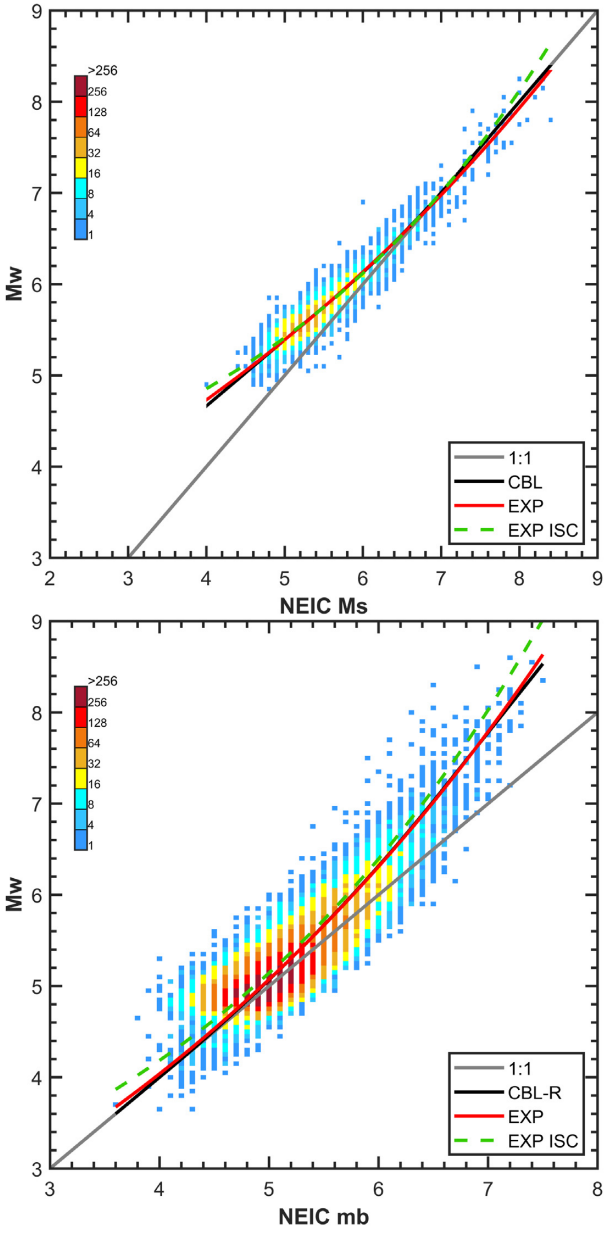


Figure 3. Data frequencies and regression curves for the GBL NEIC data set of M_s (M_s_20) from 2013 to 2022 and of m_b from 2005 to 2022. Green lines concern EXP regressions for the GBL ISC data set (of Fig. 1). Total numbers of data pairs are 3383 for M_s and 33 773 for m_b.

MOS DATA SET

In Fig. S7 top and in Table 5 and Table 6, we see the plots and coefficients for MOS GBL M_s. There is a good correspondence of both the EXP (red) and CBL (black) curves with ISC GBL. In this case the MED plot of Fig. S8 top does not reveal a significant deviation with respect to the GBL curve. For m_b in Figs S7 and S8 (bottom panels), we find a general underestimation of about 0.2–0.3 m.u. with respect to ISC and a fair coincidence between MOS GBL and MED data sets, particularly for $m_b < 5.5$.

BUILDING THE HOMOGENEOUS CATALOGUE IN NEAR REAL-TIME

To build and maintain up to date the catalogue from 1964 to present time, we adopted a procedure like the one used by Lolli *et al.* (2020) for the HORUS catalogue, in which the ISC, GCMT, RCMT and GFZP websites are periodically queried to download their updated versions. We do not download NEIC (sopar) and ETHZ data sets because they are not updated anymore since 2010 and 2006, respectively. NEIC still provide (several) M_w from Moment Tensor inversion after 2010, but they are not collected on a unique data file as done previously. Furthermore, NEIC (sopar) repository was not anymore available after about the end of 2012 (see Data and Resource section).

We access daily all sources for downloading the data of the previous year and monthly the entire database. This is to integrate as soon as possible in our database all data improvements made by various sources but at the same time without loading too much the data providers with too frequent heavy queries.

Table 5. Regression coefficients of EXP model for other agencies.

Mag.	Reg.	Data set	N	a	b	c	σ_r
M_s	GBL	NEIC	3383	1.108 ± 0.180	0.124 ± 0.013	-0.246 ± 0.653	0.141
		IDC	25 995	1.088 ± 0.042	0.124 ± 0.003	0.0128 ± 0.143	0.125
		BJI	13 354	1.913 ± 0.161	0.080 ± 0.009	-4.922 ± 1.180	0.207
		MOS	11 889	1.355 ± 0.093	0.109 ± 0.006	-1.245 ± 0.408	0.172
	MED	NEIC	100	2.212 ± 1.622	0.062 ± 0.073	-7.100 ± 15.61	0.119
		IDC	2442	4.275 ± 0.413	0.010 ± 0.004	-70.03 ± 29.74	0.142
		BJI	806	4.570 ± 0.806	0.010 ± 0.008	-96.49 ± 77.93	0.271
		MOS	731	4.251 ± 1.018	0.010 ± 0.010	-68.33 ± 71.54	0.154
m_b	GBL	NEIC	33 773	0.948 ± 0.117	0.179 ± 0.011	-1.240 ± 0.402	0.262
		IDC	34 584	-0.734 ± 0.166	0.405 ± 0.023	2.136 ± 0.184	0.347
		BJI	16 947	-1.542 ± 0.709	0.500 ± 0.100	2.612 ± 0.543	0.325
		MOS	23 313	1.080 ± 0.168	0.174 ± 0.015	-2.177 ± 0.660	0.296
	MED	NEIC	2321	4.743 ± 0.414	0.010 ± 0.004	-115.6 ± 47.57	0.258
		IDC	2941	3.688 ± 3.512	0.028 ± 0.088	-40.46 ± 141.4	0.263
		BJI	998	5.061 ± 0.465	0.010 ± 0.004	-160.7 ± 73.44	0.338
		MOS	2245	4.719 ± 0.437	0.010 ± 0.004	-112.9 ± 48.99	0.239

Table 6. Regression coefficients of CBL (M_s) and CBL-R (m_b) models for other agencies.

Mag.	Reg.	Data set	N	a	b	Δ	σ_r	M_I
M_s	GBL	NEIC	3383	0.724 ± 0.008	1.769 ± 0.043	1.078 ± 0.069	0.141	6.41
		IDC	25 995	0.646 ± 0.002	2.313 ± 0.009	2.000 ± 0.041	0.126	6.53
		BJI	13 354	0.767 ± 0.010	1.316 ± 0.048	0.985 ± 0.045	0.209	5.64
		MOS	11 889	0.708 ± 0.004	1.897 ± 0.019	2.000 ± 0.042	0.173	6.49
	MED	NEIC	100	0.790 ± 0.031	1.380 ± 0.160	0.200 ± 1.418	0.120	6.59
		IDC	2442	0.750 ± 0.004	1.785 ± 0.017	0.200 ± 3.429	0.142	7.14
		BJI	806	1.278 ± 189.5	4.538 ± 460.6	0.501 ± 124.9	0.267	-
		MOS	731	0.732 ± 0.010	1.808 ± 0.048	0.877 ± 0.407	0.154	6.75
m_b	GBL	NEIC	33 773	1.508 ± 0.018	-2.780 ± 0.106	2.000	0.265	5.47
		IDC	34 584	1.317 ± 0.007	-0.847 ± 0.034	2.000	0.353	2.67
		BJI	16 947	2.104 ± 0.064	-5.783 ± 0.364	2.000	0.331	5.24
		MOS	23 313	1.010 ± 42.79	-10.00 ± 234.2	2.000	0.235	-
	MED	NEIC	2321	1.671 ± 0.109	-3.917 ± 0.654	2.000	0.233	5.83
		IDC	2941	$1.268 \pm 3{\cdot}10^{-4}$	-0.768 ± 0.001	2.000	0.259	2.86
		BJI	998	1.010 ± 163.5	-10.00 ± 894.7	2.000	0.219	-
		MOS	2245	1.010 ± 59.84	-10.00 ± 327.5	2.000	0.210	-

After the data are downloaded from their respective providers (see Data and Resource section), M_s and m_b magnitudes are converted to M_w and merged with real M_w from moment tensor catalogues taken from MWREF catalogue. In this case we consider all magnitudes, even if they are computed by only one station.

From January 1964 to December 2020 we use the reviewed hypocentres and the magnitudes computed by ISC. From January 2020 to December 2022 we use the 'prime' location provided by the unreviewed ISC Bulletin and the M_s and m_b provided by agencies with the following rank of preference: first NEIC then IDC, BJI and finally MOS.

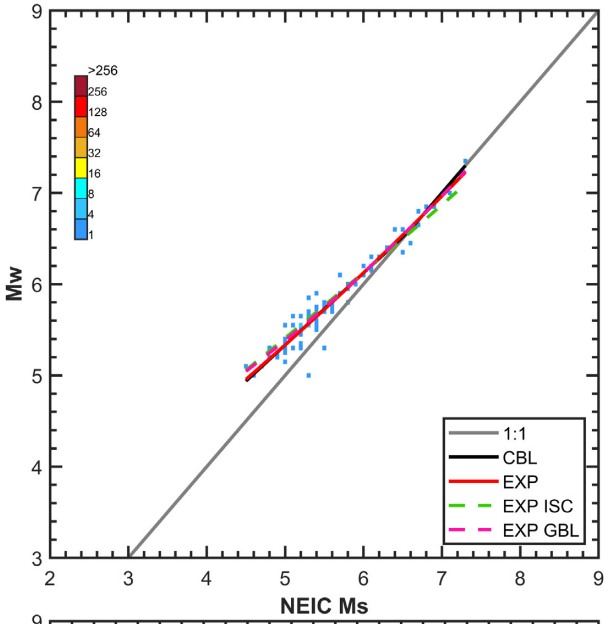
The M_w proxies at GBL and MED scale are computed from M_s and/or m_b according to the EXP relations using the corresponding coefficients reported in Tables 3 for ISC and Table 5 for NEIC, IDC, BJI and MOS. For ISC and IDC GBL we use the respective MED coefficients for $M_s < 5.5$ because, as mentioned above, we suppose that the GBL regressions are biased by the incompleteness of M_w at low magnitudes. For analogous reasons, we also use MED coefficients for m_b for IDC GBL.

The errors of proxies are computed according to Lolli *et al.* (2014) as

$$\sigma^2_{M_w} = \exp(2a+2bx)\left[\sigma_a^2 + x^2\sigma_b^2 + b^2\sigma_x^2 + 2x\,\mathrm{cov}(a,b)\right] + 2\exp(a+bx)\left[\mathrm{cov}(a,c) + x\,\mathrm{cov}(b,c)\right] + \sigma_c^2, \quad (6)$$

where σ_a^2, σ_b^2 and σ_c^2 are the variances, cov(a, b), cov(a, c) and cov(b, c) are the covariances of regression parameters and x and σ_x^2 are the observed magnitude value (M_s or m_b) and its variance, respectively. The variance/covariance matrices of all regressions are provided in the supplemental material. Just to give a few examples of such propagated errors we consider the case of ISC GBL $m_b = 5.5$: with $\sigma_{m_b} = 0.10$ we have $\sigma_{M_w} = 0.13$, with $\sigma_{m_b} = 0.40$ we have $\sigma_{M_w} = 0.50$.

When both M_s and m_b proxies are available for the same earthquake, they are averaged, using as weights the inverses of their squared errors. The error of the weighted average is computed as the square root of the inverse of the sum of weights. In case a true M_w from moment tensor catalogues is available for the same earthquake the latter is used and the proxies are discarded.



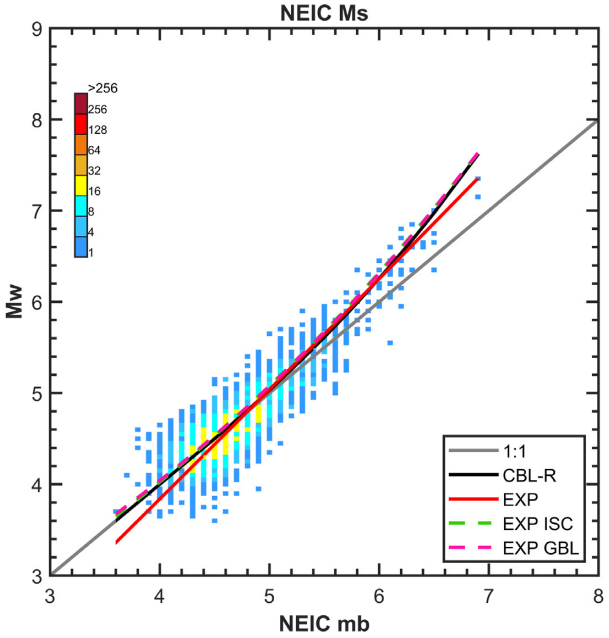
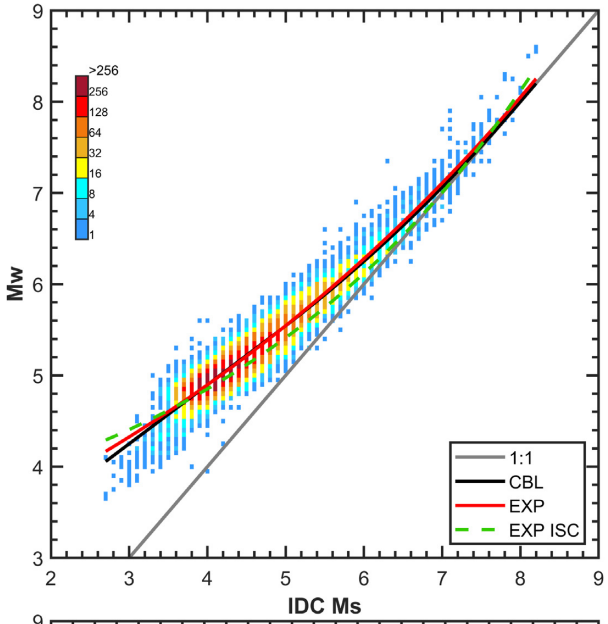


Figure 4. Data frequencies and regression curves for the MED NEIC data set of M_s (M_s_20) from 2013 to 2022 and of m_b from 2005 to 2022. Green and pink dashed curves concern EXP regressions for the MED ISC (Fig. 2) and for GBL NEIC (Fig. 3) data sets, respectively. Total numbers of data pairs are 100 for M_s and 2321 for m_b.



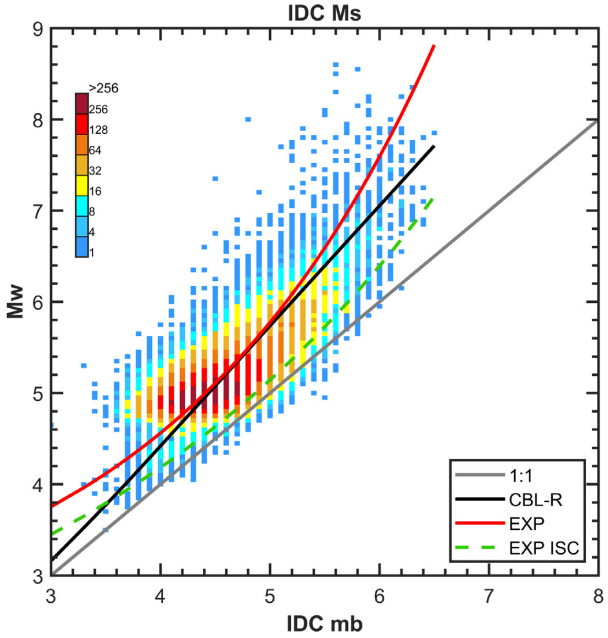


Figure 5. Data frequencies and regression curves or the GBL IDC data set from 2005 to 2022. Green dashed curves concern EXP regressions for the GBL ISC data sets (Fig. 1). Total numbers of data pairs are 25 995 for M_s and 34 584 for m_b.

The resulting catalogue with proxy or true M_w magnitudes computed from reviewed ISC magnitudes from January 1964 to December 2020 includes about 645 000 earthquakes while the unreviewed portion from January 2021 to December 2022 includes about 68 000 earthquakes.

HOMOGENEITY AND COMPLETENESS OF THE RESULTING CATALOGUE

The completeness of a seismic catalogue, over a time interval for which the detection capability of the network is assumed constant, can be evaluated by comparing the frequency magnitude distribution (FMD) with the Gutenberg & Richter (1944) (now on GR) law

$$\log_{10} N = a - b\mathrm{M}, \tag{7}$$

where N is the number of earthquakes above a given magnitude M (cumulative GR) or within magnitude bins centred in M (non-cumulative GR) and a and b are empirical coefficients. Above the completeness magnitude threshold M_c, the observed FMD almost coincides with the GR law, whereas below it, the two functions diverge, and the GR law overestimates the observed FMD. We adopt here an interactive approach based on the visual inspection of plots as proposed by Gasperini *et al.* (2013a) and Lolli *et al.* (2014), which does not differ much from that one based on the b-value stability, proposed by Cao & Gao (2002).

In Fig. 7, concerning the entire catalogue from 1964 to 2022, the

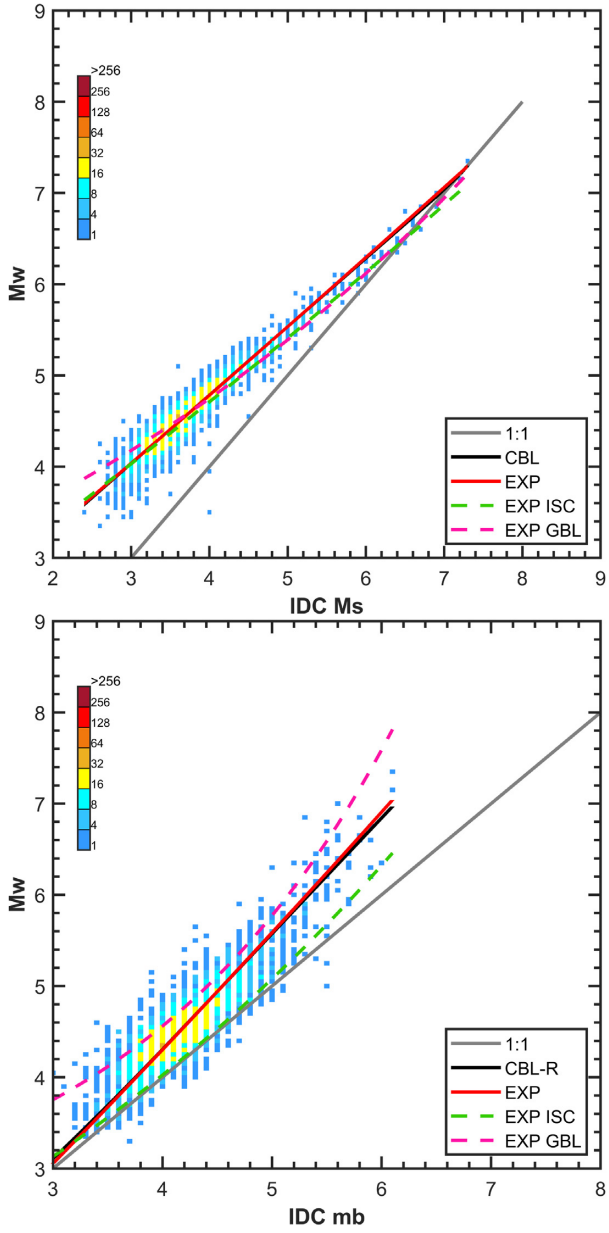


Figure 6. Data frequencies and regression curves for the MED IDC data set from 2005 to 2022. Green and pink dashed curves concern EXP regressions for the MED ISC (Fig. 1) and for GBL IDC (Fig. 5) data sets, respectively. Total numbers of data pairs are 2442 for M_s and 2941 for m_b.

cumulative FMD (solid line) is plotted as the inverse ordering rank of each magnitude and the non-cumulative FMD (black circles) as the number of earthquakes within bins of 0.1 m.u. as a function of the central magnitude of each bin. Both counts are normalized to the total duration (59 yr) of the time interval so that they correspond to annual rates. We also plotted in Fig. 7 the GR lines (black) corresponding to the *b*-value computed according to the maximum likelihood method (Aki 1965) corrected for the data binning (Utsu 1965). The vertical dashed line indicates the assumed completeness magnitude threshold of the catalogue ($M_w = 4.8$). In the upper-right inset we display the behaviour of the completeness rate, defined as the ratio between observed and predicted rates with $M_w \geq M_{min}$. In the lower-left inset we show instead the *b*-value as a function of cut-off magnitude M_{min}.

Such plots are implemented in a MS Excel worksheet in which one can tentatively vary M_c at wish, with automatic update of counts and plots. The best completeness threshold M_c is assessed as the smallest magnitude from which the plot of *b*-value as a function of cut-off magnitude M_{min} is relatively stable and there is a satisfactorily correspondence between observed rates and those predicted by the GR law as evidenced by a completeness rate close to 100 per cent on a magnitude range as wide as possible.

In Fig. 7, we can also see that both the cumulative and non-cumulative FMD with $M_w \geq 4.8$ are quite well reproduced by the GR straight lines with $b = 1.055 \pm 0.003$. In the in upper right-hand inset, we can note that the completeness rate is close to 100 per cent for M_w ranging from $M_w = 4.8$ to to $M_w = 7.7$ where it falls below 60 per cent. This fall is possibly related to the lack of very large earthquakes in many areas of the Earth owing to the absence of sufficiently large seismogenic sources. We also see in the lower left inset that the *b*-value is rather constant within a range of cut-off magnitude M_{min} from $M_w = 4.8$ to $M_w = 7.0$.

For the same data set, in Fig. S9 we plotted the distribution with time of real and proxy M_w magnitudes dithered by $\pm$ 0.07 units (Agnew 2015). The magnitude scale (vertical) is logarithmic to slightly equalize it.

In Fig. 8, we report the behaviour with time of the numbers of earthquakes in the ISC Bulletin from 1964 to 2022 within different classes 0.5 units wide of M_w magnitude (proxy or real). We can note that a definite increase of numbers (particularly for $M_w \leq 4.0$ but even for larger classes) did occur since about 1995. Such change is also evident in Fig. S9. This date probably represents a milestone at which ISC extended the coverage of data. For this reason, in Fig. 9 we report the same plot of Fig. 7 but limited to the interval from 1995 to 2022. Here the completeness M_c is reached for $M_w \geq 4.3$ for which the FMD is well approximated by the GR line with $b = 1.042 \pm 0.002$. The lower left inset indicates a reasonably constant *b*-value from $M_w = 4.3$ to about $M_w = 7.2$. The distribution with time of real and proxy M_w magnitudes for this data set is shown in Fig. S10.

The M_w proxies, in the last 2 yr (from January 2021 to December 2022) when the reviewed ISC locations and magnitudes are not available, are based on the calibration of contributing agencies. In Table 7, we report the equivalent numbers of magnitudes provided by each agency. When M_s and m_b are provided by two different agencies for the same earthquake we count 1/2 each (for this reason, some of the counts are semi-integers). We can note that, even if the priority of the choice is given to NEIC, the most of proxy magnitudes are computed by IDC. A possible reason is that, In the weeks or 2–3 months before real-time, NEIC reports to the ISC are still preliminary and the magnitude type is not available. In those instances we do not use the magnitude information by NEIC. On the other hand, the other two agencies BJI and MOS are selected as preferred magnitude for less than 0.1 per cent of earthquakes. This means that the near real-time update of the homogenized catalogue is mainly based on IDC magnitudes.

In Fig. 10, we report the same plots of Figs 7 and 9 for the unreviewed portion of the catalogue from January 2021 to December 2022. The FMD is quite well reproduced by the GR line with $b = 1.128 \pm 0.006$ for $M_w \geq 4.2$. Here the completeness rate (upper right inset) is quite constant between 95 and 105 per cent from $M_w = 4.2$ to $M_w = 6.5$ where it starts to increase. Such discrepancy is probably related to a higher *b*-value at low magnitudes (where proxies dominates) with respect to high magnitudes (where true M_w dominates), which might reflect an imperfect calibration of magnitude

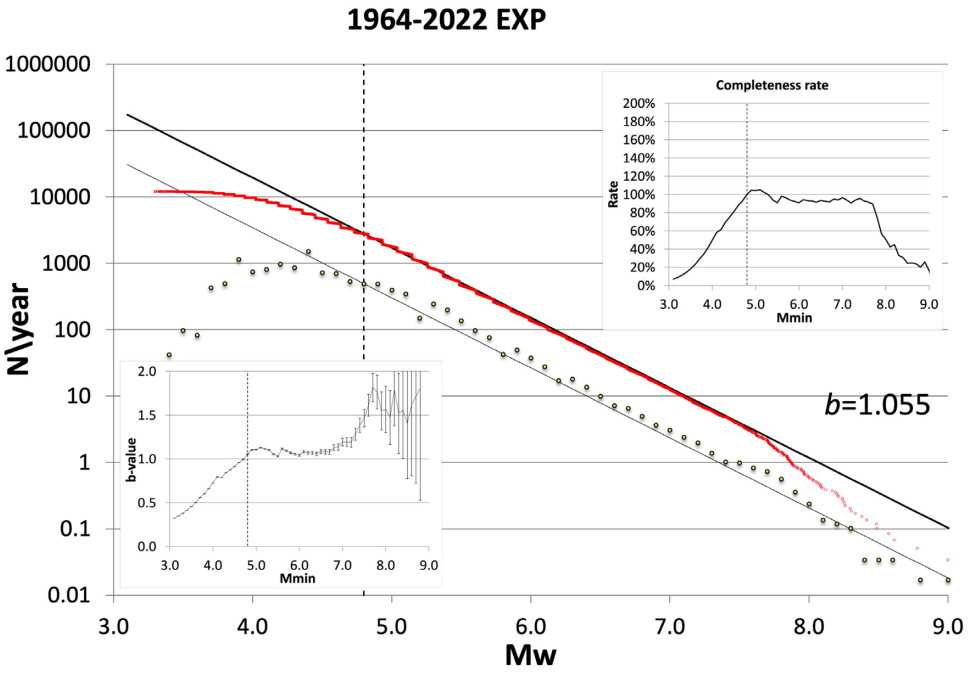


Figure 7. Cumulative (red dots) and differential (black circles) frequency–magnitude distribution of the GBL catalogue from 1964 to 2022 of real or proxy M_w estimates. Plots are normalized to the length of the time interval in years. In the upper-right inset the behaviour of the ratio between observed and predicted numbers of data with M_w proxy equal to or larger than a given M_{min} (completeness rate). In the lower-left insets the *b*-value computed as a function of the assumed cut-off magnitude (M_{min}). The vertical dashed lines indicate the assumed completeness threshold (4.8).

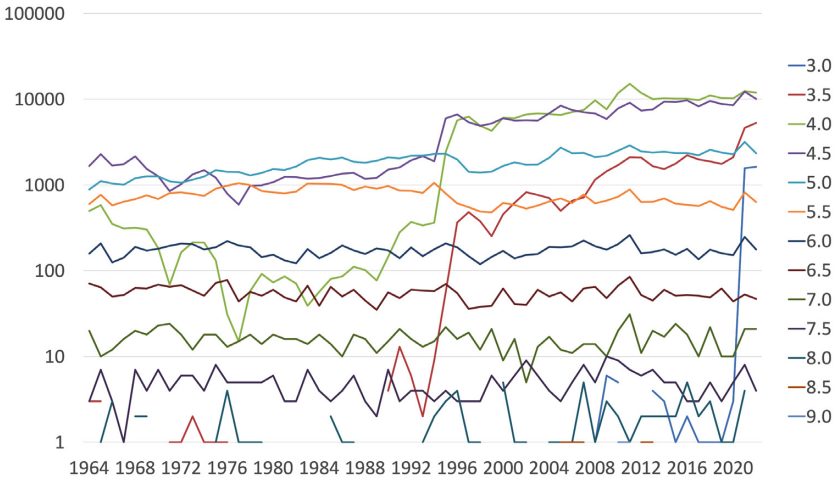


Figure 8. Behaviour with time of the numbers of earthquakes in the ISC Bulletin from 1964 to 2022 within different classes 0.5 units wide of M_w magnitude (proxy or real).

data from agencies (mainly IDC). The distribution with time of real and proxy M_w magnitudes for this data set is shown in Fig. S11.

CONCLUDING REMARKS

We implemented an automatic procedure to build and update in near real-time (with daily updates) a version of the seismic Bulletin of the ISC with homogenized magnitude. For each event we provide a unique M_w magnitude homogeneous with those estimated by the Global CMT project (Dziewonski *et al.* 1981; Ekström et al 2012). The time interval ranges from 1964 to the present time but the accuracy and the completeness vary considerably owing to the progressive improvement of the seismic detection network with time. For the time interval covered by the reviewed ISC Bulletin (usually up to 2 yr before the present time) we convert to M_w the M_s and m_b recomputed by ISC and merge them with available true M_w from moment tensor inversions computed by some global (GBL) and Mediterranean (MED) agencies. For the last 2 yr, when the ISC reviewed Bulletin is not available, we use M_s and m_b provided by some authoritative agencies.

Curvilinear conversion equations (Lolli *et al.* 2014; Di Giacomo *et al.* 2015), that capture well the nature of the relations of M_w as a function of M_s and m_b are computed by the Chi-Square regression method (Stromeyer *et al.* 2004) that considers the uncertainties of both *x* and *y* variables. We found that the coefficients computed for the reviewed ISC Bulletin from 1964 to 2022 are fairly compatible with those determined by Lolli *et al.* (2014) using data up to 2010.

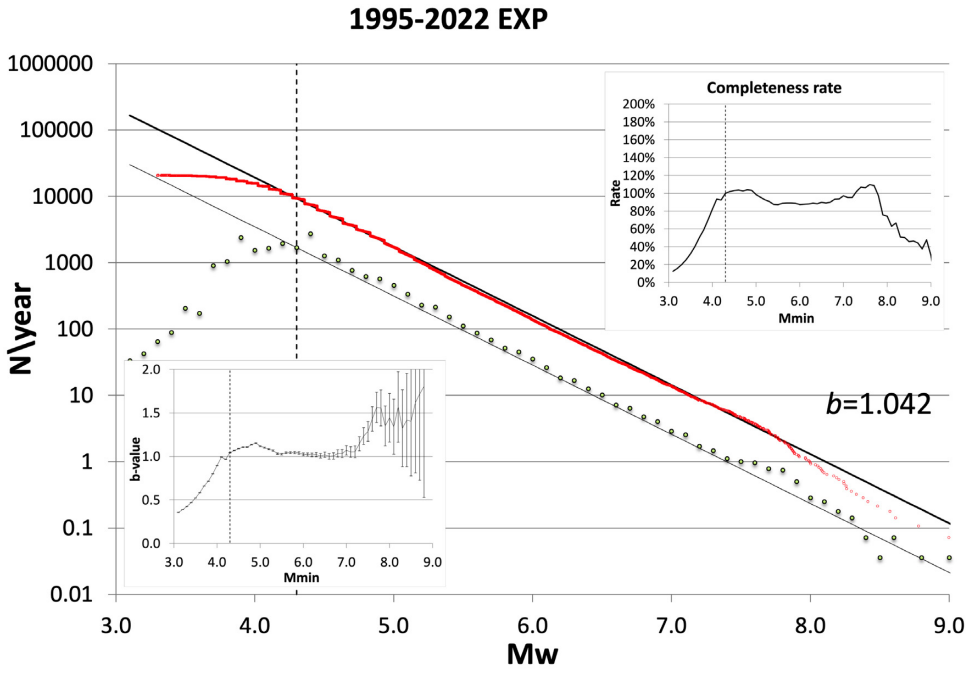


Figure 9. Same of Fig. 7 for the GBL catalogue from 1995 to 2022.

Table 7. Numbers of magnitudes from various sources for the unreviewed Bulletin 2021–2022.

Source	All		$M_w \varepsilon$ 5.25		$M_w <$ 5.25	
NEIC	18 534.5	27.28 per cent	205.5	10.09 per cent	18 329.0	27.82 per cent
IDC	43 985.0	64.75 per cent	143.0	7.02 per cent	43 842.0	66.53 per cent
BJI	44.5	0.07 per cent	14.0	0.69 per cent	30.5	0.05 per cent
MOS	66.0	0.10 per cent	7.5	0.37 per cent	58.5	0.09 per cent
MWREF	5303.0	7.81 per cent	1667.0	81.84 per cent	3636.0	5.52 per cent
Total	67 933.0	100.00 per cent	2037.0	100.00 per cent	65 896.0	100.00 per cent

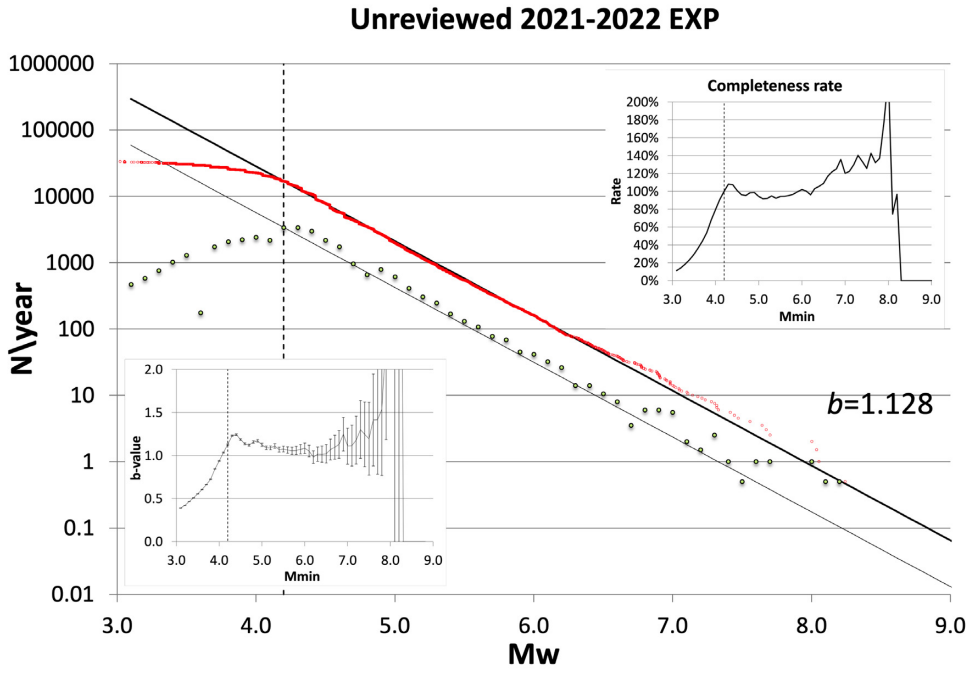


Figure 10. Same of Fig. 7 for the GBL unreviewed catalogue from 2021 to 2022.

Such results well confirm the theory (Kanamori & Anderson 1975) for M_s indicating two different slopes of the relation with M_w: 2/3 below and 1 above a transition magnitude, which we determined as $M_s \approx 6.1$–6.2, in agreement with previous estimates by Ekström & Dziewonski (1988) and Scordilis (2006).

For contributing agencies, we found a fair correspondence of NEIC with ISC for both M_s and m_b. Conversely, IDC M_s and m_b, which are the magnitudes mostly available in the last 2 yr, seem to be very different to those computed by ISC. Namely, IDC M_s deviates from the line of equality with M_w well above $M_s = 6.2$, and IDC m_b overestimates ISC m_b of 0.5 to 0.7 units. Even BJI M_s differs significantly from ISC M_s as the deviation from the line of equality with M_w occurs well below M_s 6.2 (at about M_w 5.0–5.5). BJI m_b and MOS M_s and m_b, more or less, correspond to those computed by ISC. Such discrepancy can be explained based on the mission of IDC, which is to discriminate anthropogenic events from natural earthquakes.

Some discrepancies are also observed between GBL and MED relationships. For ISC and IDC M_s we inferred that they are due to the incompleteness of the GBL M_w data set below 5.4. Hence, in accord with Lolli *et al.* (2014) we decided to adopt MED coefficients for computing ISC and IDC proxies below M_s 5.5 even for the GBL data set.

For BJI and MOS, the MED regressions would seem to suggest that the relationships both for M_s and m_b are linear. However, the scarce number of data available for constraining regression does not allow to decide if such evidence is really true or not.

DATA AND RESOURCE SECTION

Supplemental material for this article includes additional figures and tables useful to better describe methods and results. It also includes variance-covariance matrixes of all regression fitted, the Fortran sources of codes used to compute regressions and to build the MWREF and the final homogenized catalogues, as well as the MWREF and the final homogenized catalogues.

The Bulletin of the ISC from 1964 to present is collected at http://www.isc.ac.uk/iscbulletin/search/bulletin/(last accessed December 2022) https://doi.org/10.31905/D808B830.

The MT catalogue of the Geo Forschungs Zentrum Potsdam (GFZP) from 2011 to present (Saul et al 2011) is collected at http://geofon.gfz-potsdam.de/data/alerts/(last accessed December 2022).

The European-Mediterranean Regional Centroid Moment Tensor (RCMT) catalogue of INGV from 1997 to present (Pondrelli *et al.* 2002, 2011) is collected at http://rcmt2.bo.ingv.it/data/EuroMedCentrMomTensors.csv (last accessed December 2022) for definitive solutions and at http://autorcmt.bo.ingv.it/QRCMT-on-line/(last accessed December 2022) for quick preliminary solutions. Other solutions available for earthquakes before 1997 are collected from webpages linked at http://rcmt2.bo.ingv.it (last accessed December 2022).

The Global Centroid Moment Tensor (GCMT) catalogue from 1976 to present (Dziewonski *et al.* 1981; Ekström et al 2012) is collected at https://www.globalcmt.org (last accessed December 2022). Other solutions available for particular data sets are collected at webpages linked at the same address.

The Time Domain Moment Tensor (TDMT) catalogue of INGV (Dreger *et al.* 2005; Scognamiglio *et al.* 2009) from 2005 to present is collected at http://webservices.ingv.it/(last accessed December 2022).

The MT catalogue of the Eidgenössische Technische Hochschule Zürich (ETHZ) from 1999 to 2006 (Bernardi *et al.* 2004) was collected at http://www.seismo.ethz.ch/prod/tensors/mt_oldcat/index_EN (last accessed December 2012).

The MT catalogue of the National Earthquake Information Center (NEIC) of the U.S. Geological Survey from 1980 to 2010 (Sipkin 1994) was collected using http://earthquake.usgs.gov/earthquakes/eqarchives/sopar/(last accessed December 2012).

DATA AND SOFTWARE AVAILABILITY

All retrieved data and codes are freely available and provided in the supplemental material. The Fortran source of code used to regress magnitude data sets is provided as it is without any guaranty expressed or implied.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

Figure S1. Data frequencies and regression curves for the GBL ISC reviewed data set from 1964 to 2010. Green dashed curves concern EXP regressions by Lolli *et al.* (2014). Total numbers of data pairs are 18 751 for M_s and 27 898 for m_b.

Figure S2. Data frequencies and regression curves for the GBL ISC reviewed data set from 2011 to 2020. Green dashed curves concern EXP regressions by Lolli *et al.* (2014). Total numbers of data pairs are 15 437 for M_s and 20 600 for m_b.

Figure S3. Data frequencies and regression curves for the MED ISC reviewed data set from 1964 to 2010. Green and pink dashed curves concern EXP regressions by Lolli *et al.* (2014) and for the ISC GBL data sets from 1964 to 2010 of the present paper, respectively. Total numbers of data pairs are 1419 for M_s and 1939 for m_b.

Figure S4. Data frequencies and regression curves for the MED ISC reviewed data set from 2011 to 2020. Green and pink dashed curves concern EXP regressions by Lolli *et al.* (2014) and for the ISC GBL data sets from 2011 to 2020 of the present paper, respectively. Total numbers of data pairs are 1423 for M_s and 1932 for m_b.

Figure S5. Data frequencies and regression curves for the GBL BJI data set from 2005 to 2022. Green dashed curves concern EXP regressions for the GBL ISC data sets. Total numbers of data pairs are 13 354 for M_s and 16 947 for m_b.

Figure S6. Data frequencies and regression curves for the MED BJI data set from 2005 to 2022. Green and pink dashed curves concern EXP regressions for the MED ISC and GBL BJI data sets, respectively. Total numbers of data pairs are 806 for M_s and 998 for m_b.

Figure S7. Data frequencies and regression curves for the GBL MOS data set from 2005 to 2022. Green dashed curves concern EXP regressions for the GBL ISC data sets. Total numbers of data pairs are 11 889 for M_s and 23 313 for m_b.

Figure S8. Data frequencies and regression curves for the MED MOS data set from 2005 to 2022. Green and pink dashed curves concern EXP regressions for the MED ISC and GBL MOS data sets, respectively. Total numbers of data pairs are 731 for M_s and 2245 for m_b.

Figure S9. Distribution with time of M_w magnitudes from 1964 to 2022 dithered by ± 0.07 units (Agnew 2015). The magnitude scale is logarithmic to equalize the dot density.

Figure S10. Same of Fig. S9 from 1995 to 2022.

Figure S11. Same as Fig. S9 for the unreviewed catalogue from 2021 to 2022.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

ACKNOWLEDGMENTS

This paper benefited from funding provided by the H2020 EU project RISE contract n. 821115. We thank the useful comments and suggestions provided by the editor Duncan Agnew, by Domenico Di Giacomo and by two anonymous reviewers. In particular, we thank Domenico Di Giacomo for providing very useful information on the ISC Bulletin data.

REFERENCES

Agnew, D.C., 2015. Equalized plot scales for exploring seismicity data, *Seismol. Res. Lett.,* **86**(5), 1412–1423.

Aki, K., 1965. Maximum likelihood estimate of b in the formula logN = a – b M & its confidence limits, *Bull. Earthq. Res. Inst. Univ. Tokyo,* **43,** 237–239.

Bernardi, F., Braunmiller, J., Kradolfer, U. & Giardini, D., 2004. Automatic regional moment tensor inversion in the European-Mediterranean region, *Geophys. J. Int.,* **157,** 703–716.

Bisztricsany, E., 1958. A new method for the determination of the magnitude of earthquakes, *Geofiz. Kozlemen,* **7**(2), 69–76

Bondár, I. & Storchak, D.A., 2011. Improved location procedures at the International Seismological Centre. *Geophys. J. Int.,* **186,** 1220–1244.

Cao, A. & Gao, S.S., 2002. Temporal variation of seismic b-values beneath northeastern Japan island arc, *Geophys. Res. Lett.,* **29**(9), 48–41-48-3.

Carrol, R.J. & Ruppert, D., 1996. The use and misuse of orthogonal regression in linear errors-in-variables models, *Am. Stat.,* **50**(1), 1–6.

Castellaro, S., Mulargia, F. & Kagan, Y.Y., 2006 Regression problems for magnitudes, *Geophys. J. Int.,* **165,** 913–930.

Dai, G. & An, Y., 2017. China earthquake administration: Chinese Seismic Network, *Summ. Bull. Int. Seismol. Centre,* **54**(No II), 28–40.

Di Giacomo, D.,, Bondár, I., Storchak, D.A., Engdahl, E.R., Bormann, P. & Harris, J., 2015. ISC-GEM: global Instrumental Earthquake Catalogue (1900-2009): III. Re-computed MS and mb, proxy MW, final magnitude composition and completeness assessment, *Phys. Earth planet. Inter.,* **239,** 33–47.

Dreger, D.S., Gee, L., Lombard, P., Murray, M.H. & Romanowicz, B., 2005. Rapid finite-source analysis and near-fault strong ground motions: application to the 2003 Mw 6.5 San Simeon and 2004 Mw 6.0 Parkfield earthquakes. *Seismol. Res. Lett.,* **76**(1), 40–48.

Dziewonski, A.M., Chou, T.-A. & Woodhouse, J., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.,* **86,** 2825–2852.

Ekström, G. & Dziewonski, A.M., 1988. Evidence of bias in estimation of earthquake size, *Nature,* **332,** 319–323.

Ekström, G., Nettles, M. & Dziewonski, A.M., 2012. The global CMT project 2004-2010: centroid-moment tensor solutions for 13,017 earthquakes, *Phys. Earth planet. Inter.,* **200-201,** 1–9.

Fuller, W.A., 1987. *Measurement Error Models,* John Wiley, pp440.

Gasperini, P., Lolli, B. & Vannucci, G., 2013a. Empirical calibration of local magnitude data sets versus moment magnitude in Italy, *Bull. seism. Soc. Am.,* **103**(4), 6422–7222.

Gasperini, P., Lolli, B. & Vannucci, G., 2013b. Body-wave magnitude mb is a good proxy of moment magnitude mw for small earthquakes (mb < 4.5–5.0), *Seism. Res. Lett.,* **84**(6), 932–937.

Gasperini, P., Lolli, B., Vannucci, G. & Boschi, E., 2012. A comparison of moment magnitude estimates for the European–Mediterranean and Italian region, *Geophys. J. Int.,* **190,** 1733–1745.

Gutenberg, B. & Richter, C.F., 1944. Frequency of earthquakes in California, *Bull. seism. Soc. Am.,* **34,** 185–188.

Hanks, T. C & Kanamori, H., 1979. A moment magnitude scale, *J. Geophys. Res,* **84,** 2348–2350.

International Association of Seismology and Physics of the Earth's Interior (IASPEI), 2013. Summary of MagnitudeWorking Group recommendations on standard procedures for determining earthquake magnitudes from digital data, Available at: http://www.iaspei.org/commissions/commission-on-seismological-observation-and-interpretation/Summary_WG_recommendations_20130327.pdf, last accessed September 2022.

International Seismological Centre, 2018. ISC-GEM Earthquake Catalogue, http://www.isc.ac.uk/iscgem/ last accessed April 2023.

International Seismological Centre, 2021. Summ. Bull. Internatl. Seismol. Cent, January - June 2019, **56**(I), https://doi.org/10.31905/M8L1R7WI

International Seismological Centre, 2022. On-line Bulletin, at http://www.isc.ac.uk/iscbulletin/, last accessed April 2023.

Jordan, T.H. *et al.*, 2011. Operational earthquake forecasting: state of knowledge and guidelines for implementation, final report of the international commission on earthquake forecasting for civil protection, *Ann. Geophys.,* **54**(4), 315–391.

Jordan, T.H. & Jones, L.M., 2010. Operational earthquake forecasting: some thoughts in why and how, *Seismol. Res. Lett.,* **81**(4), 571–574.

Kanamori, H. & Anderson, D.L., 1975. Theoretical basis of some empirical relations in seismology, *Bull. seism. Soc. Am.,* **65**(5), 1073–1095.

Lolli, B. & Gasperini, P., 2012. A comparison among general orthogonal regression methods applied to earthquake magnitude conversions, *Geophys. J. Int.,* **190,** 1135–1151.

Lolli, B., Gasperini, P. & Vannucci, G., 2014. Empirical conversion between teleseismic magnitudes (mb and Ms) and moment magnitude (Mw) at the Global, Euro-Mediterranean and Italian scale, *Geophys. J. Int.,* **199,** 805–828.

Lolli, B., Gasperini, P. & Vannucci, G., 2015. Erratum: empirical conversion between teleseismic magnitudes (mb and Ms) and moment magnitude (Mw) at the Global, Euro-Mediterranean and Italian scale, *Geophys. J. Int.,* **200,** 199–199.

Lolli, B., Randazzo, D., Vannucci, G. & Gasperini, P., 2020. HOmogenized instRUmental seismic catalog (HORUS) of Italy from 1960 to present, *Seism. Res. Lett.,* **91,** 3208–3222.

Marzocchi, W., Lombardi, A.M. & Casarotti, E., 2014. The establishment of an operational Earthquake forecasting system in Italy, *Seismol. Res. Lett.,* **85**(5), 961–969.

Pondrelli, S., Morelli, A., Ekström, G., Mazza, S., Boschi, E. & Dziewonski, A.M., 2002. European-Mediterranean regional centroid-moment tensors: 1997–2000, *Phys. Earth planet. Inter.,* **130,** 71–101.

Pondrelli, S., Salimbeni, S., Morelli, A., Ekström, G., Postpischl, L., Vannucci, G. & Boschi, E., 2011. European–Mediterranean Regional Centroid Moment Tensor catalog: solutions for 2005–2008, *Phys. Earth planet. Inter.,* **185,** 74–81.

Richter, C.F., 1935. An instrumental earthquake magnitude scale, *Bull. seismol. Soc. Am.,* **25,** 1–31.

Saul, J., Beker, J. & Hanka, H., 2011. Global moment tensor computation at GFZ Potsdam, in *AGU 2011 Fall Meeting.,* San Francisco, CA, abstract id S51A–2202.

Scognamiglio, L., Tinti, E. & Michelini, A., 2009. Real-time determination of seismic moment tensor for the Italian region, *Bull. seism. Soc. Am.,* **99,** 2223–2242.

Scordilis, E.M., 2006. Empirical global relations converting Ms and mb to moment magnitude, *J. Seismol.,* **10,** 225–236.

Sipkin, S.A., 1994. Rapid determination of global moment-tensor solutions, *Geophys. Res. Lett.,* **21**(16), 1667–1670.

Storchak, D.A. *et al.*, 2012. ISC-GEM, Global Instrumental Earthquake Catalogue (1900–2009) GEM Technical Report 2012-01 V1.0.0, Available at https://storage.globalquakemodel.org/media/publication/GEGD-ISC-GEM-Instrumental-Catalogue-Report-201201-V01.pdf, last accessed September 2022.

Storchak, D.A., Harris, J., Brown, L., Lieser, K., Shumba, B. & Di Giacomo, D., 2020. Rebuild of the Bulletin of the International Seismological Centre (ISC)—Part 2: 1980–2010, *Geosci. Lett.,* **7,** 18.

Storchak, D.A., Harris, J., Brown, L., Lieser, K., Shumba, B., Verney, R., Di Giacomo, D. & Korger, E.I.M., 2017. Rebuild of the Bulletin of the International Seismological Centre (ISC), part 1: 1964–1979, *Geosci. Lett.,* **4,**.

Stromeyer, D., Grünthal, G. & Wahlstroöm, R., 2004. Chi-square regression for seismic strength parameter relations, and their uncertainties, with applications to an Mw based earthquake catalogue for central, northern and northwestern Europe, *J. Seismol.,* **8**(1), 143–153.

Utsu, T., 1965. A statistical significance test of the difference in b-value between two earthquake groups, *J. Phys. Earth,* **14**(2), 37–40.

APPENDIX: CHI-SQUARE REGRESSION METHOD

The model that best reproduces a given data set is that maximizing the likelihood function given by

$$l = \prod_{i=1}^{n} p\,(C_i \,|O_i\,), \tag{A1}$$

where p is the probability of observing the values O_i given that the model predicts the values C_i. For Normally distributed continuous variables, p is the probability density function of the standardized Normal distribution, so that l becomes

$$l = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{Z_i^2}{2}\right), \tag{A2}$$

where

$$Z_i = \frac{O_i - C_i}{\sigma\,(\,O_i - C_i)} \tag{A3}$$

are the standardized residuals and σ is the standard deviation of the $O_i - C_i$ differences (having zero mean).

Computing the logarithms, eq. (A2) gives

$$L = -\frac{\mathrm{N}}{2}\ln(2\pi) - \frac{1}{2}\sum_{i\,=\,1}^{N} \frac{(O_i - C_i)^2}{\sigma(O_i - C_i)^2}. \tag{A4}$$

As the first term is constant, maximizing the log-likelihood function is equivalent to minimizing the chi-square statistic (Stromeyer *et al.* 2004)

$$\chi^2 = \sum_{i=1}^{N} \frac{(O_i - C_i)^2}{\sigma(O_i - C_i)^2}. \tag{A5}$$

If the regression model C is for example the EXP function of eq. (2) of main text, between M_w and M_s we have

$$\chi^2 = \sum_{i=1}^{N} \frac{\{M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c]\}^2}{\sigma(M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c])^2}. \tag{A6}$$

If the errors of the two variables are uncorrelated, the observation variance σ^2 can be computed as (see Stromeyer *et al.* 2004; Lolli & Gasperini 2012)

$$\sigma(M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c])^2 = \left(\frac{\partial\,\{M_\mathrm{w} - [\exp(a + bM_\mathrm{s}) + c]\}}{\partial M_\mathrm{w}}\sigma_{M_{\mathrm{w}i}}\right)^2 + \left(\frac{\partial\,\{M_\mathrm{w} - [\exp(a + bM_\mathrm{s}) + c]\}}{\partial M_\mathrm{s}}\sigma_{M_{\mathrm{s}i}}\right)^2 = \sigma^2_{M_{\mathrm{w}i}} + b^2\exp(2a + 2bM_{\mathrm{s}i})\,\sigma^2_{M_{\mathrm{s}i}}, \tag{A7}$$

where $\sigma^2_{M_{\mathrm{s}i}}$ and $\sigma^2_{M_{\mathrm{w}i}}$ are the error variances of M_s and M_w observations, respectively. Then eq. (A6) becomes

$$\chi^2 = \sum_{i=1}^{N} \frac{\{M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c]\}^2}{\sigma^2_{M_{\mathrm{w}i}} + b^2\exp(2a + 2bM_{\mathrm{s}i})\,\sigma^2_{M_{\mathrm{s}i}}}. \tag{A8}$$

Also note that, if all $\sigma^2_{M_{\mathrm{s}i}} = 0$ (as assumed by the direct OLS), we have

$$\chi^2 = \sum_{i=1}^{N} \frac{\{M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c]\}^2}{\sigma^2_{M_{\mathrm{w}i}}}. \tag{A9}$$

In which each observation is weighted with the inverse of the respective M_w error variance.

In the homoscedastic case, all $\sigma^2_{M_{\mathrm{w}i}}$ are equal to the same positive constant, then minimizing eq. (A9) is equivalent to minimize the regression root mean square (r.m.s.)

$$\text{r.m.s.} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\{M_{\mathrm{w}i} - [\exp(a + bM_{\mathrm{s}i}) + c]\}^2}. \tag{A10}$$

In summary if $\sigma^2_{M_{\mathrm{s}i}} \neq 0$, the only regression methods that can be applied is the CSQ of eq. (A8) but in such case it is obvious that the (wrong) OLS regression has a smaller r.m.s. than the (correct) CSQ regression because OLS is computed just minimizing the r.m.s.