Regression problems for magnitudes

S. Castellaro,1 F. Mulargia1 and Y. Y. Kagan2

1Dipartimento di Fisica, Settore di Geofisica, Università degli Studi di Bologna, v.le C.B. Pichat 8, 40127 Bologna, Italy. E-mail: silvia.castellaro@unibo.it
2Department Earth and Space Sciences, UCLA, Los Angeles, CA, 90095-1567, USA

Accepted 2006 February 7. Received 2006 February 7; in original form 2004 October 15

SUMMARY
Least-squares linear regression is so popular that it is sometimes applied without checking whether its basic requirements are satisfied. In particular, in studying earthquake phenomena, the conditions (a) that the uncertainty on the independent variable is at least one order of magnitude smaller than the one on the dependent variable, (b) that both data and uncertainties are normally distributed and (c) that residuals are constant are at times disregarded. This may easily lead to wrong results. As an alternative to least squares, when the ratio between errors on the independent and the dependent variable can be estimated, orthogonal regression can be applied. We test the performance of orthogonal regression in its general form against Gaussian and non-Gaussian data and error distributions and compare it with standard least-squares regression. General orthogonal regression is found to be superior or equal to the standard least squares in all the cases investigated and its use is recommended. We also compare the performance of orthogonal regression versus standard regression when, as often happens in the literature, the ratio between errors on the independent and the dependent variables cannot be estimated and is arbitrarily set to 1. We apply these results to magnitude scale conversion, which is a common problem in seismology, with important implications in seismic hazard evaluation, and analyse it through specific tests. Our analysis concludes that the commonly used standard regression may induce systematic errors in magnitude conversion as high as 0.3–0.4, and, even more importantly, this can introduce apparent catalogue incompleteness, as well as a heavy bias in estimates of the slope of the frequency–magnitude distributions. All this can be avoided by using the general orthogonal regression in magnitude conversions.

Key words: magnitude conversion, orthogonal regression, seismology, statistical methods.

1 INTRODUCTION
Linear least-squares fitting is the simplest and most commonly applied technique for establishing a linear (or linearized) functional relation between two variables. However, the mathematical and statistical validity of this method is based on stringent constraints, the most important of which is that the independent variable (x) must be known to a much greater accuracy than the dependent variable (y). It follows that this regression can never be inverted, that is, the regression of y against x cannot be inverted to derive the regression of x against y. This suggests that y = α + βx should more properly be written as y ← α + βx.

Linear least squares is applied very frequently, to the point where familiarity may induce users to forget these constraints, and this may lead to wrong results.

Orthogonal regression is a more appropriate technique to deal with least-squares problems in which dependent and independent variables are both affected by uncertainty. This technique is analysed here in its general formulation (general orthogonal regression; GOR) and its performance is compared to that of standard least squares (SR) and inverted standard least-squares regression (ISR), that is, the standard regression in which the roles of dependent (y) and independent (x) variables are reversed. SR and ISR are expected to give very different results not only when errors on X and Y are very different but also when errors are similar and large. The general orthogonal regression technique is then applied to a couple of important seismological problems: magnitude conversion and the way this affects the magnitude distribution.

Explanation of notation. In this paper, we use the upper case letters (for example, X, Y) to indicate the true value of a variable and the lower case letters (for example, x, y) to indicate the value of the same variable affected by (measurement) errors.

Measuring the size of seismic events. The problem of magnitude conversion. Estimating the hazard potential of an earthquake implies measurement of the severity of the shaking in the frequency band to which buildings are sensitive, that is, typically, 0.1–10 Hz, while the ultimate effect will depend also on event duration and on local resonances. Energy magnitude (M_e, Choy & Boatwright 1995) and moment magnitude (M_w), which is based on the scalar seismic moment (M_o), are the most relevant in this respect. The seismic
moment (which is a function of the product of the rupture area and the average displacement) does not tell us much about the relative amount of energy released at the frequencies relevant to the seismic hazard estimates (cf., e.g. NMSOP, Chapter 3).

Ignoring the point that a single indicator of size may be inadequate in seismic hazard estimates, the state of the art is to use $M_L$ on account of its better definition in seismological terms. Unfortunately a moment-tensor solution can be practically worked out only for sizeable events, so, for the bulk of the events in any seismic catalogue, the size estimate is made on the basis of magnitude or, for the historical part of the catalogue, on the basis of intensity. Local magnitude ($M_L$), surface wave magnitude ($M_s$), body wave magnitude ($M_b$) and duration magnitude ($M_d$) are the magnitudes most commonly reported in seismic catalogues (see Table 1 for magnitude acronyms). This variety of indicators of size earthquake reflects the specific instrument used (e.g. the Wood-Anderson seismometer to produce $M_L$ data) as well as the properties of the seismic sequence. For example, amplitude saturation, inhibiting the evaluation of $M_L$ or $M_s$ for strong earthquakes, led to the introduction of the duration magnitude $M_d$, which in turn is not immune from problems, not least, that it lacks a unique definition (see e.g. Lee et al. 1972 for the procedure followed in California and Gasperini 2002, for the procedure adopted in Italy). Additionally, in seismic swarms or aftershock sequences, the beginning of an event may overlap the end of the previous one, making the $M_d$ estimate impossible or highly inaccurate. Finally, a wide variety of events can lead to systematic errors in the reported magnitudes, such as deviations in the instrumental calibration or changes in the seismic equipment (Habermann 1986), changes of the agency operating the earthquake recording network, introduction of new software for the analysis, removal or addition of seismograph stations as well as changes in the magnitude definition (Zúñiga & Wyss 1995). Such systematic errors can be very large, as much as 0.5 magnitude units, as found by Tinti & Mulargia (1985) studied the effect of magnitude errors on the G-R relation and found that they do not practically affect the $b$ but only the $a$ value. If we consider earthquakes with magnitude $m_t$ and larger, and the probability density function for the G-R eq. (1) can be written in the form

$$f(d) = B \exp^{-bd} \quad d \geq 0,$$

where $d = m - m_z$ and $B = b \log e 10 \simeq \log e 10$ (Utsu 1999; Kagan 2005). We now show how the $b$-value of the frequency–magnitude relation is affected when the magnitudes used are converted from other magnitudes through standard or orthogonal regressions. The G-R $b$-value will be calculated through the maximum-likelihood method, since neither standard nor orthogonal regression would apply (see Section 4).

### 2 THE MAGNITUDE CONVERSION PROBLEM

However refined it may be, no analysis can give results of better quality than that of the data it employs. Regarding earthquake size, data are often referred to moment magnitude, $M_w$, which is linked to the seismic scalar moment $M_0$ through the relation

$$M_w = \frac{2}{3} \log_{10} M_0 - C \quad [M_0 \text{ in } N m].$$

Hanks & Kanamori (1979) suggest $C = 6 + \frac{1}{2} \log_{10} M_s$ but for simplicity of expression we use $C = 6.0$, as in Hanks (1992). Moment magnitude, which can be computed when a moment-tensor solution is available, has a number of advantages over other magnitude definitions: (1) it is a physical parameter of the earthquake, which allows the earthquake process to be quantitatively linked to tectonic deformation (cf. Kagan 2002a; Bird & Kagan 2004); (2) it does not saturate for large earthquakes; (3) it allows determination of the moment magnitude with an accuracy two to three times higher than with other magnitudes (cf. Kagan 2002b, 2003). However, we must note that these properties apply only when $M_w$ is estimated directly from the seismic recording and this is generally true for a very small percentage of events (the larger ones). In all other cases $M_w$ is a value converted from other data, so the benefits listed above do not fully apply.

Table 1. Symbols used to indicate different magnitudes: $M_L$ indicates a true or synthetic Wood-Anderson magnitude calculated from broad-band seismic stations; $M_d$ is calculated as in Gasperini (2002); $M_s$ as in Choy & Boatwright (1995).

<table>
<thead>
<tr>
<th>Magnitude definition</th>
<th>$M_L$</th>
<th>$M_s$</th>
<th>$M_d$</th>
<th>$M_w$</th>
<th>$M_0$</th>
<th>$M_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface wave magnitude</td>
<td></td>
<td></td>
<td></td>
<td>$M_s$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Local magnitude</td>
<td>$M_L$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration magnitude</td>
<td></td>
<td></td>
<td>$M_d$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moment magnitude</td>
<td>$M_b$</td>
<td></td>
<td></td>
<td>$M_w$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalar seismic moment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$M_0$</td>
<td></td>
</tr>
<tr>
<td>Energy magnitude</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$M_e$</td>
</tr>
</tbody>
</table>

Table 2. Data sets used to perform the tests on the slope. The left table reports the number of events recorded in Italy between 1981 and 1996 for which both the magnitudes shown in the first two columns are available. The right table reports the standard deviation $\sigma$ of the different magnitudes.

<table>
<thead>
<tr>
<th>Data no.</th>
<th>$M_L$</th>
<th>$M_w$</th>
<th>$M_0$</th>
<th>$M_s$</th>
<th>$M_d$</th>
<th>$M_b$</th>
<th>$M_e$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_s$</td>
<td>109</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.28</td>
</tr>
<tr>
<td>$M_L$</td>
<td>121</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.22</td>
</tr>
<tr>
<td>$m_b$</td>
<td>204</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>$M_w$</td>
<td></td>
<td></td>
<td></td>
<td>204</td>
<td></td>
<td></td>
<td></td>
<td>0.18</td>
</tr>
</tbody>
</table>

Magnitude distribution. The frequency–magnitude distribution of earthquakes (Ishimoto & Iida 1939; Gutenberg & Richter 1944), often called the Gutenberg-Richter or G-R relation, is a basic ingredient of seismic hazard estimates and is also used to test the completeness of seismic catalogues. The G-R relation states that

$$\log N(m) = a - bm,$$

where $N(m)$ is the number of events with magnitude $\geq m$ and $a$, $b$ are constants ($b \approx 1.0$). Tinti & Mulargia (1985) studied the effect of magnitude errors on the G-R relation and found that they do not practically affect the $b$ but only the $a$ value. If we consider
The standard least-square linear regression procedure has so far been the method most commonly used to find the relation between different types of magnitude (see, e.g. Gasperini & Ferrari 2000; Gasperini 2002; Bindi et al. 2005 for the Italian catalogues, Braunmiller et al. 2005 for Switzerland). However, one can argue that this method is misapplied: (1) because both the dependent and the non-dependent variable are affected by similar errors and this contradicts the basic assumption of the standard regression method, (2) because magnitude is not a normally distributed variable. Although other methods exist the most general solution is general orthogonal regression. In this paper we test the performance of this method in the case of magnitude conversion (see Section 3).

We start by summarizing the properties of orthogonal regression in its general formulation (GOR) and then proceed to test its performance through simulations on Gaussian and non-Gaussian data sets and error distributions. Next, we apply it to find the relation in predictors (Madansky 1959; Fuller 1987; Kendall & Stuart 1979, chap. 28; Carroll & Ruppert, 1996). Its general lines can be outlined from the experimental to the ‘true’ line is the same (Fig. 2b and York 1967). In other words the paths from the experimental (x, y) to the ‘true’ (x*, y*) follows lines with slopes which depend on the size of the errors affecting x and y. For constant x and y, the slope of the lines from the experimental to the ‘true’ line is the same (Fig. 2b and York 1967) and, in the case of GOR, is the weighted orthogonal distance. The use of standard regression in the latter case is like projecting x* on the abscissa of the sampled point, which again results in a Gaussian distribution. However, in case of magnitudes, non-linearity in the m1 ÷ m2 relation, saturation of magnitude scales, additive and multiplicative noise and a host of other phenomena may result in a distortion of the Gaussian distribution on y. Since the least-square estimator of a regression coefficient is vulnerable to gross errors and the related confidence interval is sensitive to non-normality of the parent distribution, it is important to test its validity.

GOR is designed to account for the effects of measurement error in predictors (Madansky 1959; Fuller 1987; Kendall & Stuart 1979, chap. 28; Carroll & Ruppert, 1996). Its general lines can be outlined as follows: let us assume that two variables Y and X are linearly related and that their measurement errors ε and u are independent normal variates with variances σε² and σu² respectively, that is,

\[ y = Y + \epsilon, \]
\[ x = X + u, \]

and the regression-like model

\[ Y = \alpha + \beta X + \hat{\epsilon}, \]

where \( \hat{\epsilon} = \epsilon + u \). Now consider the error variance ratio

\[ \eta = \frac{\sigma_u^2}{\sigma_\epsilon^2}, \]

Figure 1. Standard least-square regression applies when the dependent variable is affected by a Gaussian error much larger than the error affecting the independent variable (A) and it consists in projecting the independent variable along a vertical line (B).

© 2006 The Authors, GJI, 165, 913–930
Journal compilation © 2006 RAS
where \( \sigma_x^2 = \sigma^2 \) and \( \sigma_y^2 = \sigma^2 \), provided that \( \sigma_x^2 \) and \( \sigma_y^2 \) are constants. Orthogonal regression (OR) is often defined as the case in which \( \eta \) in the unknowns, that is \( \eta = \alpha, \beta, X_i \). For \( \eta = 1 \) we have the squared Euclidean distance of the point \((x_i, y_i)\) from the line \((X_i, \alpha + \beta X_i)\). If \( \eta \neq 1 \) then eq. (8) represents a weighted orthogonal distance.

Let us call \( s^2_x, s^2_y \) and \( s_{xy} \) the sample variance of the \( y \), \( x \) and the sample covariance between \( y \) and \( x \). The general orthogonal estimator of slope is then

\[
\hat{\beta} = \frac{s^2_x - \eta s^2_y + \sqrt{(s^2_x - \eta s^2_y)^2 + 4 \eta s^2_{xy}}}{2 s_{xy}},
\]
and the estimator of the intercept is

\[
\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x},
\]
where \( \bar{x} \) and \( \bar{y} \) stand for the average values.

Note that while the standard regression predicts \( Y \) from \( X \), since it assumes that no error is present in measurements of \( X \) (i.e. \( x = X \)), orthogonal regression predicts \( Y \) from \( x \) and \( y \), as shown by eq. (8) above.

As demonstrated in Fuller (1987), in practice, the following formulae can be used to estimate the errors on the regression parameters:

\[
\hat{\sigma}_\beta^2 = \frac{s^2_x s^2_y + \sigma_x^2 s_{xy} - \hat{\sigma}_u^2}{n - 1} \frac{(n - 1) \hat{\sigma}_u^2}{(n - 2) \hat{\sigma}_\beta^2},
\]
and

\[
\hat{\sigma}_\sigma^2 = \frac{s^2_x + \hat{\sigma}_\beta^2 \hat{\sigma}_u^2}{n} \frac{\hat{\sigma}_\beta^2}{\hat{\sigma}_u^2},
\]

where

\[
s_v = \sum_{i=1}^{n} \left| Y_i - \bar{y} - \hat{\beta} (X_i - \bar{x}) \right|^2 = \frac{(n - 1)(\eta + \hat{\beta}^2) \hat{\sigma}_u^2}{(n - 2)},
\]
and

\[
\hat{\sigma}_u^2 = -\hat{\beta} \hat{\sigma}_v,
\]

Under the assumption that the variations of the observations about the line are normal, that is that the errors are all from the same normal distributions \( N(0, \sigma^2) \), it can be shown that the 100(1 - \( \alpha \)) per cent confidence limits for \( \beta \) and \( \sigma \) are

\[
\hat{\beta} \pm t(n - 2, 1 - \frac{1}{2} \alpha) \sqrt{\hat{\sigma}_\beta^2 \hat{\sigma}_u^2 - \frac{s^2_x}{n} \frac{(n - 1) \hat{\sigma}_u^2}{(n - 2) \hat{\sigma}_\beta^2}},
\]

and

\[
\hat{\sigma} \pm t(n - 2, 1 - \frac{1}{2} \alpha) \sqrt{\frac{s^2_x}{n} + \frac{x^2 \hat{\sigma}_\beta^2}{\hat{\sigma}_u^2}},
\]

where \( t(n - 2, 1 - \frac{1}{2} \alpha) \) is the 100(1 - \( \alpha \)) percentage point of a \( t \)-distribution with \( (n - 2) \) degrees of freedom (e.g. Kendall & Stuart 1979).

### 3.1 Simulations

In principle, the theoretical likelihood function can be obtained for the main variable (magnitude) in \( x \) and \( y \) distributed according to

...
Noise is added in four ways: we explore the case in which saturation effects or other problems in the magnitude measurement. Large deviations from this should suggest the presence of a theoretical derivation would be of great value in the magnitude regression problem since other issues such as magnitude threshold variation (Katsura & Ogata 2004) and non-linearity of magnitude relations blur the picture. The most appropriate technique to deal with all these problems seems to be a simulation (see also Stefanski & Cook 1995).

In this section we analyse the performance of GOR compared to standard least-square regression on:

1. normally, log-normally and exponentially distributed variables,
2. normally, log-normally and exponentially distributed errors and
3. different amount of errors.

The performance of the GOR method is first analysed versus SR and ISR by studying the distribution curves of the $\beta$ values (that is the slope coefficients) calculated on $10^3$ regessions of 50 different $(x, y)$ sets generated according to some fixed criteria: all the $(x, y)$ sets have been produced starting from $x : y = 1 : 1$ and adding noise to $x$ and/or $y$ according to different criteria. The true result of the regression is then expected to be $\beta = 1$ in all cases. This value is chosen as the most representative of the magnitude regression problem. Large deviations from this should suggest the presence of saturation effects or other problems in the magnitude measurement. Noise is added in four ways: we explore the case in which $\eta = 25$, where $\eta$ is defined as in eq. (7), that is when an error on $Y$ five times larger than the error on $X$ is added. Second, we explore $\eta = 4$, that is the case in which the errors on $Y$ are twice the size of the errors on $X$; third $\eta = 1$, the case in which errors on $Y$ and $X$ are comparable in size. Last, we use $\eta = 0.25$, that is the case in which the errors on $X$ are twice the errors on $Y$. Since the requirements of SR apply strictly only when $\eta \gg 1$ and when variables and their errors are normally distributed, we can expect a priori SR to perform well only in the first case ($\eta = 25$) and with variables and errors normally distributed. In the fourth case ISR is expected to perform as GOR and as SR in the case $\eta = 25$, so this case will be presented only in Section 3.1.1

### 3.1.1 Normal distributions and errors

When data and errors are normally distributed and the error on $Y$ is much larger than the error on $X$, SR and GOR have a practically identical performance, as shown in Fig. 3(a) which shows the case of errors on $Y$ five times larger than the error on $X$. The $\beta$ distribution curve of the ISR, as obvious, performs much worse since error is mostly on the dependent variable.

When the error on $Y$ is twice that on $X$ (Fig. 3b), SR regression performs slightly worse than GOR while ISR gives estimates in error by 5 per cent.

When errors on $Y$ are comparable to errors on $X$ (Fig. 3c), GOR performs better than SR and ISR. The discrepancy between SR and ISR depends on $\sigma_x$ and $\sigma_y$ and increases with data scatter (see Section 3.4 and also Figs 2 and 3 in Bornmann & Wylegalla 1975; and Figs 3 and 4 in Bornmann & Khaturlin 1975).

Last, when errors on $Y$ are half the size of errors on $X$, GOR performs always well while, as expected, SR does not. This case corresponds to Fig. 3(a) where error size on $X$ and $Y$ has been exchanged. ISR is then expected to perform as SR in case $\eta$ above, that is, approximately as GOR.

### 3.1.2 Exponential distributions and Gaussian errors

We now consider sets of data distributed according to exponential functions and with Gaussian errors. This is a case of interest in seismology since frequency–magnitude is an exponential distribution affected by Gaussian errors (see eqs 1 and 2).

As might be anticipated, the performance of the various approaches differs. When errors on $Y$ are five times larger than errors on $X$ (Fig. 4a) and when errors on $Y$ are twice those on $X$ (Fig. 4b), the $\beta$ distributions obtained with GOR and SR are comparable even if the SR distribution appears to be biased towards smaller values by a few percent.

GOR performance remains good while SR decays with decreasing error on $Y$ ratio over error on $X$ (Fig. 4c). We note that ISR also performs poorly in the whole range of $\eta$ considered.

### 3.1.3 Exponential distributions and errors

We now consider sets of data distributed according to exponential functions and with exponential errors. We take into account this case since we consider it as worse than exponentially distributed data with normal errors (Section 3.1.2) and we want to see whether GOR continues to perform well under such unfavourable circumstances.

As in Section 3.1.2, when errors on $Y$ are five times larger than errors on $X$ (Fig. 5a), the $\beta$ distributions obtained with GOR and SR are comparable even if the SR distribution appears to be biased towards smaller values by a few percent.

GOR performance remains good while SR progressively decays with decreasing error on $Y$ ratio over error on $X$ (Figs 5b, c). We note that ISR also performs poorly in the whole range of $\eta$ considered.

### 3.1.4 Lognormal distributions and errors

When used on log-normally distributed data and errors, with errors on $Y$ five times larger than error on $X$, GOR performs much better than SR, which appears to be biased to values smaller than real by a few percent (Fig. 6a). GOR performs much better than SR and ISR in all the other cases investigated (Figs 6b and c).

### 3.2 What happens if $\eta$ is unknown?

It often happens (see Section 3.5) that when the ratio between errors on the $Y$ and $X$ variables is unknown, it is assumed to be equal to unity, that is $\eta = 1$, which formally coincides with the OR assumption. In some cases this assumption is justified and is expected to be not too far from reality. We explore the effects of this assumption when $\eta \neq 1$.

First we considered a set of lognormal data and error (as in Figs 6a–c) with a true $\eta = 0.25$, that is, error on $X$ twice that on $Y$. When $\eta = 1$ is used instead of $\eta = 0.25$, one still obtains a better performance than SR, as it can be seen in Fig. 7(a), although in a less evident way than Figs 6(a)–(c), where the true $\eta$ was applied.

In any case, the use of $\eta = 1$ can lead to wrong results if the true $\eta$ is very large, as shown in Fig. 7(b). Here the $\beta$ coefficients of the regressions have been computed for the OR ($\eta = 1$) instead of the correct value $\eta = 25$ on a set of log-normally distributed data and errors. This case corresponds to that studied in Fig. 6(a) with the correct $\eta$.

In light of this and of the considerations in the previous section, the use of GOR is recommended in all cases, provided that at least an order of magnitude of $\eta$ is available.
Figure 3. Left: examples taken from the $10^3$ generations of 50 couples of points $(x, y)$. Data are normally distributed and variable Gaussian errors have been added on the $x$- and $y$-axis. Right: frequency distribution of the $\beta$ coefficient found with generalized orthogonal regression (GOR), standard least squares (SR) and inverted standard least-squares (ISR) method. (A) $\eta = 25, \sigma_y = 2, \sigma_x = \sqrt{\frac{\sigma_y^2}{\eta}} = 0.4$ as from eq. (7). (B) $\eta = 4, \sigma_y = 2, \sigma_x = 1$. (C) $\eta = 1, \sigma_y = 2, \sigma_x = 2$. (D) $\eta = 0.25, \sigma_y = 2, \sigma_x = 4$. 

© 2006 The Authors, GJI, 165, 913–930
Journal compilation © 2006 RAS
Figure 4. Left: examples taken from the $10^3$ generations of 50 couples of points $(x, y)$. Data are exponentially distributed and variable Gaussian errors have been added on the $x$- and $y$-axis. Right: frequency distribution of the $\beta$ coefficient found with the GOR, SR and ISR method. (A) $\eta = 25$, $\sigma_y = 1$, $\sigma_x = 0.2$. (B) $\eta = 4$, $\sigma_y = 1$, $\sigma_x = 0.5$. (C) $\eta = 1$, $\sigma_y = 1$, $\sigma_x = 1$. 

© 2006 The Authors, GJI, 165, 913–930
Journal compilation © 2006 RAS
Figure 5. Left: examples taken from the $10^3$ generations of 50 couples of points $(x, y)$. Data are exponentially distributed and variable exponential errors have been added on the $x$- and $y$-axis. Right: frequency distribution of the $\beta$ coefficient found with the GOR, SR and ISR method. (A) $\eta = 25$, $\sigma_y = 1$, $\sigma_x = 0.2$. (B) $\eta = 4$, $\sigma_y = 1$, $\sigma_x = 0.5$. (C) $\eta = 1$, $\sigma_y = 1$, $\sigma_x = 1$. 
Figure 6. Left: examples taken from the $10^3$ generations of 50 couples of points $(x, y)$. Data are log-normally distributed and variable log-normal errors have been added on the x- and y-axis. Right: frequency distribution of the $\beta$ coefficient found with the GOR, SR and ISR method. (A) $\eta = 25$, $\sigma_y = 2$, $\sigma_x = 0.4$. (B) $\eta = 4$, $\sigma_y = 2$, $\sigma_x = 1$. (C) $\eta = 1$, $\sigma_y = 2$, $\sigma_x = 2$. 
Figure 7. Examples taken from the 10^3 generations of 50 couples of points (x, y). Data are log-normally distributed and variable log-normal errors have been added on the x- and y-axis. Frequency distribution of the $\beta$ coefficient found with OR and the SR method. (A) in the calculation $\eta$ has been set to 1 instead of the true 0.25 value ($\sigma_y = 2$, $\sigma_x = 4$). (B) in the calculation $\eta$ has been set to 1 instead of the true 25 value ($\sigma_y = 2$, $\sigma_x = 0.4$).

3.3 Does magnitude distribution induce a bias in magnitude–magnitude regressions?

Having established the superiority of GOR with respect to SR in general terms, let us now test the GOR on a typical seismological problem. In order to verify whether the exponential distribution of earthquake magnitudes, which has an upper and lower cut-off (cf. Kagan 2002a), induces a bias in magnitude versus magnitude regressions according to different scales, we simulate magnitude–frequency relations, add errors, remove events below the lower threshold and then run standard and orthogonal regressions. To this end, we created a synthetic database of 120 $(m_1, m_2)$ pairs of magnitudes distributed according to an exponential-distribution and affected by a Gaussian error. The parameter $B$ of the exponential distribution is set to $\log_e 10 = 2.3$, corresponding to $b = 1$ in the G-R relation (eqs 1 and 2, Section 4); $\sigma_{m_2}$ is fixed to 0.18 and $\sigma_{m_1}$ is set to 0.22 as in Table 2. We should, therefore, expect that ISR gives more robust results compared to SR. The number of pairs was fixed to such a small value in order to be similar to the true database available (Table 2). We note that $m_1$ and $m_2$ have been generated using the same distribution, hence on average $m_1 = m_2$, that is, in contrast to the real cases which will be discussed later on, there is no systematic bias in

Figure 8. Standard and generalized orthogonal regression values for the 120 couples of simulated events distributed according to an exponential-distribution with parameter $B = 2.30$ and affected by Gaussian error with $\eta = 0.67$ ($\sigma_y = 0.18$, $\sigma_x = 0.22$). Results of the regressions on the whole data set. We define the reduced magnitude $m''_1 = m_1 - 4.5$ and $m''_2 = m_2 - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.
Figure 9. As in Fig. 8 but regressions have been computed on data above the $m = 4.5$ threshold only.

Figure 10. Standard and generalized orthogonal regression values for the 109 couples of events recorded in Italy between 1981 and 1996 gauged with $M_w$ and $M_s$. We define the reduced magnitudes $M'_w = M_w - 4.5$ and $M'_s = M_s - 4.5$. Regression results are shown in inset: GOR $\rightarrow$ general orthogonal regression, SR $\rightarrow$ standard regression, ISR $\rightarrow$ inverted standard regression.

magnitude relation. The results of the standard and orthogonal regression on this data set are reported in Fig. 8, which shows the substantial systematic error (biases) introduced by the use of standard regression. For example, suppose that we measured $m_1 = 6.0$. This value would give $m_2 = 5.70$ according to the SR, $m_2 = 6.24$ according to the ISR but $m_2 = 6.00$ according to the GOR. Systematic errors of opposite sign would occur for $m_1$ approximately lower than 4.5.

Since this set is composed of synthetic data, generated by a simulation that does not include data incompleteness at low magnitude, there is no lower completeness threshold and the cumulative frequency–magnitude distribution is a line with slope equal to $-1$. 

© 2006 The Authors, *GJI*, 165, 913–930 
Journal compilation © 2006 RAS
without any deviation from linearity. However, if by analogy with what is usually done in seismology we consider only events above a magnitude threshold (say 4.5) higher than the simulation threshold (Fig. 9) we still obtain different regression laws and, as expected since $\eta < 1$, the ISR is closer to GOR. For $m_1 = 6$, GOR would estimate $m_2 = 6.18$, SR would estimate $m_2 = 5.79$ and ISR would produce $m_2 = 6.46$. Again, the most correct results are given by GOR.
Table 3. Results of the general orthogonal regression (GOR) applied to our data set. The model is \( y = \alpha + \beta x \), that is \( M_w = \alpha + \beta m_b \). \( \alpha \) applies when magnitudes are reduced so that \( m_w^* = M_w - 4.5, m_b^* = m_b - 4.5 \) and \( m_w^* = \alpha^* + \beta m_b^* \). See Figs 10–12.

<table>
<thead>
<tr>
<th>( y )</th>
<th>( x )</th>
<th>( \beta )</th>
<th>( \alpha )</th>
<th>( \alpha' )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( M_w )</td>
<td>( M_s )</td>
<td>0.765 ± 0.035</td>
<td>1.445 ± 0.175</td>
<td>0.426</td>
</tr>
<tr>
<td>( M_w )</td>
<td>( M_L )</td>
<td>0.890 ± 0.033</td>
<td>0.770 ± 0.159</td>
<td>0.283</td>
</tr>
<tr>
<td>( M_w )</td>
<td>( m_b )</td>
<td>1.291 ± 0.046</td>
<td>−1.272 ± 0.271</td>
<td>0.037</td>
</tr>
</tbody>
</table>

3.4 Application

We now apply orthogonal regression to the real data sets of Table 2 and find the regression coefficients and error estimates reported in Table 3. The estimates of standard deviation \( \sigma \) of the different magnitudes, needed to derive the \( \eta \) value, have been computed from the cases in which more than 3 stations recorded and classified the same earthquake. Although they may be rough estimates, the relative values appear reasonable considering how the different magnitudes are derived from earthquake recordings.

As can be seen from Figs 10–12, the GOR estimator has the property of lying between the slope of the standard regression of \( y \) on \( x \) (SR) and the inverse of the slope of the standard regression of \( x \) on \( y \) (ISR). It can also be noted, as already found by Bormann & Wylegalla (1975, Figs 2 and 3) and Bormann & Khalturin (1975, Figs 3 and 4), that the slope difference between SR and ISR increases with data scatter.

Considering the \( M_w = m_b \) regression (Fig. 12) we observe that the GOR slope is much closer to the inverted least-square regression slope when \( m_b \) is taken as the dependent variable, and in fact \( \eta < 1 \).

The above examples show that normal regression may introduce significant errors, (from +0.1 to −0.3 in the example of Fig. 8) during magnitude conversion. Moreover, as Figs 10–12 demonstrate, a substantial systematic bias is present in observational data. Non-linearity in the magnitude relation, which is obvious in the \( M_w = m_b \) plot, adds to conversion errors. In conclusion, this bias, together with improper accounting for magnitudes’ uncertainty, leads to significant distortion of seismicity and of any seismic hazard estimate based on them. Therefore, GOR should always be used, rather than standard regression, in magnitude conversions, provided that at least an order of magnitude of \( \eta \) is available.

In addition, note that if the differences between magnitudes were due to random errors, we should see \( \alpha \)-values of the order of 0.0, and \( \beta \)-values of the order of 1.0 as in Figs 8 and 9. Clearly this is not a case, and the reason for this is not only incorrect use of regression. Magnitudes are empirical quantities, and their inter-relationship was established through various regression relations, so one should not expect that they rigorously characterize earthquake size. Since the \( \beta \)-values often are substantially different from 1.0, this should strongly affect the \( b \)-value (of the G-R relation) bias (see more in Section 4).

3.5 Orthogonal regression in the seismic literature: a critical review

Orthogonal regression has been used several times to study European seismicity (Bormann & Khalturin 1975; Ambraseys 1990; Panza et al. 1993; Cavallini & Rebez 1996; Kaverina et al. 1996; Gutdeutsch et al. 2002, 2005; Grünthal & Wahlström 2003; Stromeier et al. 2004) and global magnitude relations (Bormann & Khalturin 1975). It has also been applied to find the slope of the magnitude–intensity relation (cf. Cavallini & Rebez 1996; Gutdeutsch et al. 2002; Stromeier et al. 2004). All these applications dealt with the special case (OR) of assuming equal errors on the dependent and independent variables, that is, \( \eta = 1 \) in eqs 7–9, whereas we deal with the GOR method of \( \eta \neq 1 \). Note that similar general results would be less comfortably achieved by rescaling the axes as a function of \( \sigma_y \) and \( \sigma_x \) to reach \( \eta = 1 \).

Gutdeutsch et al. (2002) also applied OR to find the \( M_L \sim M_s \) relation for the high-quality Kärnik (1996) earthquake database of Central and Southern Europe, which also includes Italy. It is interesting to note that their \( M_L \sim M_s \) relation, based on about 250 pairs \( (M_L, M_s) \) gives \( \beta = 0.893 \pm 0.163 \), which is in perfect agreement with our \( \beta_{M_L \sim M_s} = 0.859 \) derived by merging the first two equations in Table 3. This probably implies that the error ratio of \( M_L \) and \( M_s \) is close enough to 1 (a more detailed discussion of this problem is left to another manuscript currently in preparation).

As an alternative to OR, Grünthal & Wahlström (2003) and Stromeier et al. (2004) proposed the chi-square maximum likelihood regression. They applied it to find the \( M_w \sim M_s \) relationship for Central Europe earthquakes, including a small part of Northern Italy. An advantage of the chi-square regression is that it is a distribution-free method. It has, however, the same problem as GOR: the ratio of the standard deviations of the data should be known with a reasonable accuracy, for example 30 per cent. Grünthal & Wahlström (2003) and Stromeier et al. (2004), instead, assumed \( \eta = 1 \) as in OR. In any case, we note that Grünthal & Wahlström’s (2003) chi-square regression gave \( \beta_{M_L \sim M_s} = 0.769 \), while our GOR gives \( \beta_{M_L \sim M_s} = 0.769 \), that is, a very similar result in spite of the different method and data set. A comparison between chi-square method and OR can also be found in Gutdeutsch et al. (2005).

4 FREQUENCY–MAGNITUDE RELATION

Frequency–magnitude relations (G-R relation) are ubiquitously used in seismology, for example, to test the completeness of seismic catalogs, to estimate seismic hazard or in the thermodynamic criticality approach to earthquake modelling (Knopoff 2000). The G-R relation can itself be treated as a regression but it is not amenable to the case treated in this paper since neither the assumptions of SR, nor those of GOR strictly apply. The preferred technique to estimate the \( b \)-value in this case is the maximum likelihood estimator method (MLEM) given by Aki (1965) and discussed by Weichert (1980), Utsu (1999) and Kagan (2005):

\[
b = \frac{\log_{10} e}{M - M_c},
\]

where \( M \) is the average magnitude and \( M_c \) is the lower cut-off magnitude (or the completeness threshold magnitude). The accuracy on the \( b \)-value, derived from the statistical estimation theory (Aki 1965; Utsu 1999), is given by

\[
\sigma_b = \frac{b}{\sqrt{n}},
\]

where \( n \) is the total number of earthquakes in the sample.

Since the G-R relations are evaluated on a single magnitude scale, which is in general derived from converting other magnitudes, it is nevertheless interesting to see how the \( b \) estimations are affected by the use of different regression procedures employed for magnitude conversion. Again, the issue is most effectively treated in simulation.

We start by assuming that we have a set of \( 10^6 \) exponentially distributed events with \( M_s \) magnitude values which follow a G-R relation (eq. 2) without added error and which represent our ‘true’ data. We then add random Gaussian errors with variance \( \sigma^2_{M_s} \) and...
obtain a set of $10^4$ $(m_2)$ data which represent the `noisy' data, that is, the measured data of the same set. These can be seen as the real moment magnitude data (with error). Generally these data are unavailable and are derived by converting estimates measured on other magnitude scales. We consider just one such magnitude scale, which we call $m_1$ and generate a set of $10^4$ $(m_1)$ synthetic data still using an exponential distribution with $B = 2.3$ (eq. 2) but with different realizations and Gaussian errors with variance $\sigma_{m1}^2$. These can be imagined to correspond, say, to $M_2$ data.

We now derive the conversion law for $m_2$ data from the $m_1$ set $(m_2 = \alpha + \beta m_1)$ by using both SR $(m_2)$ and GOR $(\tilde{m}_2)$ and study how the G-R plot of the derived magnitudes $(\tilde{m}_2)$ and $(\tilde{m}_2)$ is consequently affected. The $b$-values in both cases are computed through the MLEM (eqs 19–20).

Figs 13–15 show both the `true' data relative to $M_2$, the corresponding noisy ones with directly added noise $m_2$, and the $\tilde{m}_2$ and $\tilde{m}_2$ ones inferred from $m_1$. We first note that, as predicted by Tinti & Mulargia (1985), a magnitude error affects the $a$, but not the $b$-value in the G-R relation. In particular, in the log-linear part of the plot, $a$ is shifted to $a'$ as a function of the variance $\sigma_{m1}^2$:

$$a' = a + \frac{b^2 \sigma_{m1}^2}{2 \log_{10} e}.$$  

so that the effect of magnitude uncertainties is that the observed number of earthquakes exceeding a given magnitude always appears larger than the true number.

Fig. 13 shows that when $\eta = 25$ ($\sigma_{m2} = 0.5$ and $\sigma_{m1} = 0.1$), that is, when the error on the original variable $m_2$ is five times larger than error on the original variable $m_1$, SR and GOR give overlapping G-R plots, which coincide with the `true' data. As stated in Section 3, SR (when $\sigma_y \gg \sigma_x$) and GOR are capable of retrieving $Y$ from $x$, because $x$ can be considered $\simeq X$. However, $\sigma_y \gg \sigma_x$ hardly occurs in real earthquake catalogues since measurement errors on $m_2$ are unlikely to be five times larger than measurement errors on $m_1$ (see e.g. Table 2 for realistic error estimates on $m_1$ and $m_2$). As mentioned above, in the most common case $m_2 = M_w$ and $m_1 = M_p$. In this example we have used a large $\sigma_{m1}$ in order to emphasize the effect of the magnitude uncertainties (dashed vs. solid line) in Fig. 13. Note that, in this case, the true $a$-value at $m = 5.5$ is 2.51 (corresponding to 326 earthquakes), while the observed $a$-value is 2.78 (corresponding to 640 earthquakes) which fits with the theoretical $a' = 2.79$ derived from eq. (21) (corresponding to 617 earthquakes).

When $\eta$ decreases from 25 to 1 ($\sigma_{m2} = \sigma_{m1} = 0.2$) and 0.25 ($\sigma_{m2} = 0.2$, $\sigma_{m1} = 0.4$), (see Figs 14 and 15) the G-R relation derived using $\tilde{m}_2$ values (converted through $m_1$ through SR) strongly differs from the true G-R, while the G-R obtained using $\tilde{m}_2$ values (converted through the GOR) still replicate the true relation.

An example of average $b$ slope coefficients and their error is shown in Table 4 and is obtained through MLEM on events with $m \geq 4.5$. Due to the large sample size on which we run these simulations, we expect that the $b$-values obtained in this way are less scattered than in the simulations of Figs 3–6.

In summary, Figs 13–15 and Table 4 show that the use of $\tilde{m}_2$ data inferred from $m_1$ through SR produces a strong bias in the slope, which is reflected in a biased estimate of the $b$-value, even when computed through the MLEM. When $\tilde{m}_2$ data inferred from GOR are used, true unbiased $M_2$ data are always produced. This demonstrates the importance of using GOR in converting magnitudes to be used in the G-R relation to obtain correct hazard estimates.

5 Regression problems involving slope: a non-parametric regression test

The performance of a least-square estimator of a regression coefficient can be gauged through a non-parametric test.

© 2006 The Authors, GJI, 165, 913–930
Journal compilation © 2006 RAS
When there are no ties in the data (i.e. all data points are distinct), Theil’s test (p. 416 in Hollander & Wolfe 1999) is suitable. When ties are present (and this is easily the case with magnitudes), an estimate of the regression coefficients can be produced based on Kendall’s tau (Kendall 1962; Sen 1968). This procedure is similar to Theil’s but is based on weaker assumptions and does not require \( x_1, \ldots, x_n \) to be all distinct. We start from the premise that the model is
\[
y_i = \alpha + \beta x_i \quad i = 1, \ldots, n, \tag{22}
\]
where \( x_1 \leq x_2 \leq \ldots \leq x_n \) are the known constants and \( \alpha \) and \( \beta \) the unknown parameters. Calling \( N \) the number of non-zero differences \( x_j - x_i \) (\( 1 \leq i \leq j \leq n \)), Sen (1968) shows that the point estimator is the median of the \( N \) slopes \((y_j - y_i)/(x_j - x_i)\) for which \( x_j \neq x_i \) and this is shown to be unbiased for \( \beta \). The confidence interval for \( \beta \) is also obtained in terms of second-order statistics of this set of \( N \) slopes.

Let us define \( c(u) \) to be 1, 0 or \(-1\) if \( u > 0 \), \( u = 0 \) or \( u > 0 \) respectively. Then the number of positive differences \( x_j - x_i \) is
\[
N = \sum_{1 \leq j < i \leq n} c(x_j - x_i), \tag{23}
\]
and \( N \leq \binom{n}{2} \). For any real \( \beta \), \( Z_i(\beta) = y_i - \beta x_i \). Relying on the Kendall’s tau statistic (Kendall 1962) between \( x_i \) and \( Z_i(\beta) \) we thus have

\[
U_n(\beta) = \left[ N \left( \frac{n}{2} \right) \right]^{1/2} \sum_{i<j \leq n} c(x_j - x_i) c(Z_j(\beta) - Z_i(\beta)).
\]

(24)

Since we deal with quite large sample sizes, we can use the asymptotic formula

\[
U_n(\beta) = \tau_{1/2} \sqrt{V_n},
\]

(25)

where \( \tau_{1/2} \) is the upper 100\% per cent point of a standard normal distribution and the variance \( V_n \) is

\[
V_n = \frac{1}{18} n(n-1)(2n+5) - \frac{\sum_{j=1}^{n} u_j (u_j-1)(2u_j+5)}{n-2}.
\]

(26)

Let \( a_n \) be composed of \( a_n \geq 2 \) distinct sets of elements. In each of the \( i \)th \( (i = 1, \ldots, a_n) \) set there are then \( u_i \) all equal elements which are the values to be inserted in eq. (26). \( U_n(\beta) \) is a strictly distribution-free statistic having a distribution symmetric about 0. This implies that one way of estimating \( \beta \) is to make \( U_n(\beta) \) as close to zero as possible, by varying \( \beta \).

The confidence interval for \( \beta \) having the confidence coefficient \( 1 - \epsilon_n \) is thus

\[
P\{ \beta_n^* \leq \beta < \beta_n^* \} = 1 - \epsilon_n,
\]

(27)

where

\[
\beta_n^* = \text{Sup} \{ \beta : U_n(\beta) \geq -U_n \},
\]

(28)

\[
\beta_n^* = \text{Inf} \{ \beta : U_n(\beta) \leq U_n \}.
\]

(29)

We recall that these formulae apply exclusively when ties are present in the \( x \) variable only. We now consider the set of \( N \) distinct pairs \((i, j)\) for which \( x_j > x_i \) and define

\[
X_{ij} = \frac{y_j - y_i}{x_j - x_i}.
\]

(30)

Thus the \( X_{ij} \) are the slopes of the lines connecting each pair of points \((x_i, y_i)\) and \((x_j, y_j)\) where \( x_i \neq x_j \). After arranging the \( N \) values of \( X_{ij} \) in increasing order, the slope estimator is shown (Sen 1968) to be the median of the \( N \) numbers and the confidence interval is given by the \( M_1 \)th and \( M_2 \)th values

\[
N^* = \sqrt{N \binom{n}{2}} U_n \text{ and } M_i = \frac{1}{2} (N + (-1)^i N^*).
\]

(31)
specifically formulated to handle the case where both variables are affected by errors of non-negligible size, should instead be used to obtain reliable magnitude conversions and trustworthy seismic hazard estimates. This appears clearly in the final application in which we show that the G-R frequency–magnitude relation, which is in general calculated on magnitudes converted from various scales, can be heavily biased if magnitudes are converted through standard least squares. By contrast, it is possible to obtain unbiased estimates of $a$- and $b$-values (eq. 1) by converting magnitudes through generalized orthogonal regression.

The influence of magnitude uncertainties on the $a$- and $b$-values provides a strong indication of serious biases in traditional interpretations of earthquake size distribution. Shifts in the $a$-values due to magnitude errors may have important consequences on seismic data obtained in different time intervals. As a rule, the accuracy of older data is lower, and since the $a$-value shift is always positive and proportional to the square of magnitude errors (Tinti & Mulargia 1985 and eq. 21), this means that the seismic activity for older time intervals may spuriously appear to exceed more recent activity by a significant margin. This effect is likely to be at best in part responsible for the often-claimed discrepancy between earthquake rates in recent and old catalogues. Such disagreement in earthquake rates might even be taken as evidence for the presence of characteristic earthquakes in a region.

Virtually all traditional (non-seismic moment) earthquake catalogues contain several magnitudes, the relationships between which have been established by standard (not orthogonal) regression. Bias and systematic effects in magnitude data are difficult or next to impossible to disentangle. As Figs 13–15 and Table 4 demonstrate, incorrect application of regression methods in the magnitude conversion may induce $b$-value bias of the order of 20–40 per cent. Hence, the variations in $b$-values (claimed by many researchers) may have an additional explanation as an artefact of improper catalogue data processing.

ACKNOWLEDGMENTS

SC and FM acknowledge support from MIUR 60 per cent funds and from the INGV-DPC 2004–2006, Project V3. YYK acknowledges partial support from the National Science Foundation through grants EAR-00-11128, EAR-04-09890, DMS-0306526, and from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. Publication 960, SCEC.

We thank Prof. P. Bormann, Prof. A. Oncel as well as the three anonymous reviewers for their constructive suggestions.

REFERENCES


